Advances in Mathematical Sciences and Applications
Vol. 28, No. 2 (2019), pp. 425–447

GAKKOTOSHO
TOKYO JAPAN

PENALIZATION VIA GLOBAL FUNCTIONALS OF OPTIMAL-CONTROL PROBLEMS FOR DISSIPATIVE EVOLUTION

Lorenzo Portinale
IST Austria,
Am Campus 1, 3400 Klosterneuburg, Austria.
(E-mail: `lorenzo.portinale@ist.ac.at`)

and

Ulisse Stefanelli
Faculty of Mathematics, University of Vienna,
Oskar-Morgenstern-Platz 1, 1090 Vienna, Austria and
Istituto di Matematica Applicata e Tecnologie Informatiche *E. Magenes*,
via Ferrata 1, 27100 Pavia, Italy.
(E-mail: `ulisse.stefanelli@univie.ac.at`)

Abstract. We consider an optimal control problem for an abstract nonlinear dissipative evolution equation. The differential constraint is penalized by augmenting the target functional by a nonnegative global-in-time functional which is null-minimized iff the evolution equation is satisfied. Different variational settings are presented, leading to the convergence of the penalization method for gradient flows, noncyclic and semimonotone flows, doubly nonlinear evolutions, and GENERIC systems.

Communicated by Editors; Received September 27, 2019
This work is supported by Vienna Science and Technology Fund (WWTF) through Project MA14-009 and by the Austrian Science Fund (FWF) projects F 65 and I 2375.
AMS Subject Classification: 49J20, 35K55, 58E30.
Keywords: Optimal control, dissipative evolution, penalization, global variational method.

1 Introduction

We are concerned with the abstract optimal control problem

$$\min\{F(u,y)\ :\ y\in S(u)\}. \tag{1.1}$$

Here, $u:[0,T]\to H$ stands for a time-dependent admissible control, H is a Hilbert space, and $y:[0,T]\to H$ belongs to the set $S(u)$ of a nonlinear evolution equation with datum u to be specified below. The nonnegative *target* functional F is defined on the trajectories u and y.

Relation $y\in S(u)$ corresponds to different models of dissipative evolution. In particular, we will consider the case of u-forced

Gradient flows:	$y'+\partial\phi(y)=u,$
Monotone and pseudomonotone flows:	$y'+A(y)=u,$
Generalized gradient flows:	$\partial_{y'}\psi(y,y')+\partial\phi(y)=u,$
GENERIC flows:	$y'=L(y)\,DE(y)-K(y)(\partial\phi(y)-u).$

The reader is referred to the following sections for all necessary details. In all of these cases, the abstract relation $y\in S(u)$ stands for the variational formulation of a nonlinear partial differential problem of parabolic type, possibly being singular or degenerate.

The differential constraint $y\in S(u)$ will be equivalently reformulated as

$$y\in S(u)\quad\Leftrightarrow\quad G(u,y)=0,$$

where the *constraining functional* G is a nonnegative functional on entire trajectories. This characterization is not new. In the specific case of a gradient flow $y'+\partial\phi(y)=u$, where $\partial\phi$ stands for the subdifferential of the convex energy $\phi:H\to(-\infty,\infty]$, two possible choices of the constraint functional G are given by the *Brezis-Ekeland-Nayroles* functional

$$G_{\rm BEN}(u,y)=\int_0^T\Big(\phi(y)+\phi^*(u-y')-(u,y)\Big)\mathrm{d}t+\frac12\|y(T)\|^2-\frac12\|y_0\|^2$$

and the *De Giorgi* functional

$$G_{\rm DG}(u,y)=\int_0^T\left(\frac12\|y'\|^2+\frac12\|\partial\phi(y)-u\|^2-(u,y')\right)\mathrm{d}t+\phi(y(T))-\phi(y_0).$$

Here, $(\cdot,\cdot)$ and $\|\cdot\|$ denote the scalar product and the norm in H, respectively. The trajectory y is forced to assume the initial value $y(0)=y_0$ by defining $G(u,y)=\infty$ otherwise.

The focus of this note is on the penalization of problem (1.1) by

$$\min E_\varepsilon(u,y)\quad\text{for}\quad E_\varepsilon(u,y):=F(u,y)+\frac1\varepsilon G(u,y). \tag{1.2}$$

This corresponds to approximate the constrained minimization of problem (1.1) by means of a family of unconstrained minimizations.

This approach is indeed classical and has to be traced back to Lions [24], who proposed to penalize the constraint by the residual of the equation. This has already been investigated, both in the stationary and the evolutive case, see [3, 4, 5, 15, 20, 29] among many others. We follow this line by penalizing the minimization by the De Giorgi functional G_{DG}, which corresponds to the residual by nonetheless exploiting the variational structure of the equation in order to simplify the energy. On the other hand, penalization in coordination with the Brezis-Ekeland-Nayroles functional G_{BEN} is not directly related with residual minimization and, to our knowledge, has not been studied yet. Note that the actual choice of the constraining functional G strongly influences the properties of the problem, so that the considering different options for G is a sensible issue.

In the case of the Brezis-Ekeland-Nayroles functional G_{BEN}, problem (1.2) turns out to be a *separately convex* minimization problem. This allows for the implementation of an alternate minimization procedure, where E_ε is alternatively minimized in the state and the control until convergence.

The case of the De Giorgi functional G_{DG} bears its interest in the fact that it is not restricted to convex functionals ϕ. In fact, G_{DG} is suited for nonconvex potentials as well and it can be easily modified to accommodate additional nonlinear features, such as nonlinear dissipative or conservative terms (see Section 4 below).

Our aim is that of checking the solvability of the penalized minimization problem (1.2) and the convergence of its minimizers to minimizers of the constrained problem (1.1) as $\varepsilon \to 0$. This will be achieved by proving the Γ-convergence of the penalized functional E_ε to the limit E_0 defined by

$$E_0(u,y) = F(u,y) \ \text{ if } \ G(u,y) = 0 \ \text{ and } \ E_0(u,y) = \infty \ \text{ otherwise}$$

under different variational settings, corresponding to the above-mentioned different evolution models.

The paper is organized as follows. The abstract functional setup is detailed in Section 2. Then, the application of the abstract theory to the case of the Brezis-Ekeland-Nayroles variational principle for gradient, noncyclic and semimonotone flows, and doubly nonlinear flows is addressed in Section 3. Eventually, Section 4 deals with the applications of De Giorgi principle in the context of gradient, doubly nonlinear, and GENERIC flows.

2 Abstract setup

Let us start by specifying some notation. In the following, H stands for a real separable Hilbert space with scalar product $(\cdot,\cdot)$ and norm $\|\cdot\|$. The norm in the general Banach space E will be denoted by $\|\cdot\|_E$. Given the reference time $T > 0$, we make use of the standard Bochner spaces $L^p(0,T;E)$, $W^{1,p}(0,T;E)$, $C([0,T];E)$ and so on.

A caveat on notation: we will use the same symbol c to indicate positive universal constants, possibly depending on data, and changing from line to line.

Given a topological space (X,τ), we recall that a sequence of functionals $\mathcal{E}_\varepsilon : (X,\tau) \to [0,\infty]$ is said to Γ-converge [11] to the limit $\mathcal{E}_0 : (X,\tau) \to [0,\infty]$ if $\mathcal{E}_0(x) \leq \liminf_{\varepsilon\to 0} \mathcal{E}_\varepsilon(x_\varepsilon)$ for any $x_\varepsilon \to x$ and for all $\hat{x} \in X$ there exists a sequence $\hat{x}_\varepsilon \to \hat{x}$ such that $\mathcal{E}_\varepsilon(x_\varepsilon) \to \mathcal{E}_0(\hat{x})$. The reader is referred to Dal Maso [10] for a thorough presentation.

We record here the following elementary lemma, which serves as basis for proving convergence of the minimizers of problem (1.2) throughout.

Lemma 2.1 (Γ-convergence)**.** *Let (X,τ) be a sequential topological space and the functionals $\mathcal{F}$, $\mathcal{G} : (X,\tau) \to [0,\infty]$ be lower semicontinuous. Assume $\mathcal{E}_\varepsilon := \mathcal{F} + \varepsilon^{-1}\mathcal{G}$ to be proper ($\mathcal{E}_\varepsilon \not\equiv \infty$) and equicoercive for $\varepsilon > 0$ small enough, namely that there exists $\varepsilon_0 > 0$, $\lambda > 0$, and a compact $K \subset X$ such that $\{x \in X \ : \ \mathcal{E}_\varepsilon(x) < \lambda\} \subset K$ for all $\varepsilon < \varepsilon_0$. Then,*

1. *$\mathcal{E}_\varepsilon \xrightarrow{\Gamma} \mathcal{E}_0$ where $\mathcal{E}_0(x) := \mathcal{F}(x)$ if $\mathcal{G}(x) = 0$ and $\mathcal{E}_0 = \infty$ otherwise;*
2. *$\min \mathcal{E}_\varepsilon$ can be solved for all $\varepsilon < \varepsilon_0$. Any sequence x_ε of quasiminimizers, namely $\liminf_{\varepsilon\to 0}(\mathcal{E}_\varepsilon(x_\varepsilon) - \inf \mathcal{E}_\varepsilon) = 0$, admits a subsequence converging to a minimizer of $\mathcal{E}_0$;*
3. *If $\mathcal{E}_0$ admits a unique minimizer x_0, any sequence of quasiminimizers of $\mathcal{E}_\varepsilon$ converges to x_0.*

Proof. Ad 1. Let $x_\varepsilon \to x$ and assume with no loss of generality that $\sup_\varepsilon \mathcal{E}_\varepsilon(x_\varepsilon) \le c < \infty$. In particular, $0 \le \mathcal{G}(x) \le \liminf_{\varepsilon\to 0} \varepsilon\mathcal{E}_\varepsilon(x_\varepsilon) \le \liminf_{\varepsilon\to 0} \varepsilon c = 0$. Then $\mathcal{E}_0(x) = \mathcal{F}(x) \le \liminf_{\varepsilon\to 0} \mathcal{F}(x_\varepsilon) \le \liminf_{\varepsilon\to 0} \mathcal{E}_\varepsilon(x_\varepsilon)$. Fix now any $\hat{x} \in X$. As $\varepsilon^{-1}\mathcal{G}(\hat{x}) \to \infty$ if $\mathcal{G}(\hat{x}) > 0$, one has that $\mathcal{E}_\varepsilon(\hat{x}) \to \mathcal{E}_0(\hat{x})$. This proves the Γ-convergence $\mathcal{E}_\varepsilon \xrightarrow{\Gamma} \mathcal{E}_0$.

Ad 2. The existence of a minimizer x_ε of $\mathcal{E}_\varepsilon$ for $\varepsilon < \varepsilon_0$ follows from the equicoercivity and the lower semicontinuity of the sum $\mathcal{F}+\varepsilon^{-1}\mathcal{G}$. Any sequence x_ε of quasiminimizers belongs to K for ε small enough. As such, it admits a subsequence (not relabeled) converging to x_0 and, for any $x \in X$, we have that $\mathcal{E}_0(x_0) \le \liminf_{\varepsilon\to 0} \mathcal{E}_\varepsilon(x_\varepsilon) = \liminf_{\varepsilon\to 0} \min \mathcal{E}_\varepsilon \le \liminf_{\varepsilon\to 0} \mathcal{E}_\varepsilon(x) = \mathcal{E}_0(x)$. In particular, x_0 minimizes $\mathcal{E}_0$.

Ad 3. This follows from the uniqueness of the minimizer of $\mathcal{E}_0$ and from the fact that the topology is assumed to be sequential. □

3 Brezis-Ekeland-Nayroles principle

In this section, we investigate penalization (1.2) by letting the constraining functional to be of Brezis-Ekeland-Nayroles type. Let us start by presenting a result in the case of the classical gradient flow with forcing u

$$y' + \partial\phi(y) \ni u \ \text{ in } H, \text{ a.e. in } (0,T), \quad y(0) = y_0. \tag{3.1}$$

As usual, the prime denotes here derivation with respect to time. The potential $\phi : H \to (-\infty,\infty]$ is assumed to be convex, proper, and lower semicontinuous, and we denote by $D(\phi) = \{y \in H \ : \ \phi(y) < \infty\}$ its essential domain. The symbol $\partial\phi$ denotes the corresponding subdifferential in the sense of convex analysis. This is defined as

$$\xi \in \partial\phi(y) \quad \Leftrightarrow \quad y \in D(\phi) \ \text{ and } \ (\xi, x-y) \le \phi(x) - \phi(y) \ \ \forall x \in H.$$

The initial datum y_0 is assumed to belong to $D(\phi)$. Given $u \in L^2(0,T;H)$, the solution $y \in H^1(0,T;H)$ of (3.1) exists uniquely [6]. The celebrated result by Brezis & Ekeland

[7, 8] and Nayroles [31, 32] implies that y solves (3.1) iff $G_{\mathrm{BEN}}(u,y) = 0$, where the constraining functional $G_{\mathrm{BEN}}(u,y) : L^2(0,T;H) \times H^1(0,T;H)$ is given by

$$G_{\mathrm{BEN}}(u,y) = \begin{cases} \displaystyle\int_0^T \Big(\phi(y)+\phi^*(u-y')-(u,y)\Big)\,\mathrm{d}t + \frac{1}{2}\|y(T)\|^2 - \frac{1}{2}\|y_0\|^2 & \text{if } y(0) = y_0 \\ \infty & \text{otherwise.} \end{cases} \tag{3.2}$$

Here, ϕ^* denotes the conjugate to ϕ, namely, $\phi^*(y^*) = \sup_y((y^*,y) - \phi(y)$. Note that, for all $(u,y) \in L^2(0,T;H) \times H^1(0,T;H)$ the functions $t \mapsto \phi(y'(t))$ and $y \mapsto \phi^*(u(t)-y'(t))$ are measurable, so that $G_{\mathrm{BEN}}(u,y)$ is well defined. Still, $G_{\mathrm{BEN}}(u,y)$ takes the value ∞ if $t \mapsto \phi(y'(t))$ or $y \mapsto \phi^*(u(t)-y'(t))$ do not belong to $L^1(0,T)$.

Existence results based in the Brezis-Ekeland-Nayroles principle have been obtained by Rios [33], Auchmuty [1], Roubíček [35], and Ghoussoub & Tzou [17] among others. In [17], the authors recast the problem within the far-reaching theory of (anti-)selfdual Lagrangians [18]. A variety of extensions have been proposed, including perturbations [16], long-time dynamics [23], measure data [25], time discretizations [37], second-order [26], doubly-nonlinear [36], monotone [39], pseudomonotone equations and their structural compactness [41], and rate-independent flows [38]. Note however that deriving existence via these extensions may call for more stringent assumptions on the data of the problem.

In the following, we will assume that the set of admissible controls U is a compact subset of $L^2(0,T;H)$. Moreover, we ask the target functional $F : L^2(0,T;H) \times H^1(0,T;H) \to [0,\infty)$ to be lower semicontinuous with respect to the strong $\times$ weak topology of $L^2(0,T;H) \times H^1(0,T;H)$. An example in this class is

$$F(u,y) = \frac{1}{2}\int_0^T \|y - y_{\mathrm{target}}\|^2 \mathrm{d}t + \frac{1}{2}\int_0^T \|y' - y'_{\mathrm{target}}\|^2 \mathrm{d}t + \frac{1}{2}\int_0^T \|u\|^2 \mathrm{d}t$$

for some given $y_{\mathrm{target}} \in H^1(0,T;H)$. The main result of this section is the following.

Theorem 3.1 (Gradient flows, BEN principle). *Let $\phi : H \to (-\infty,\infty]$ be convex, proper, and lower semicontinuous, $y_0 \in D(\phi)$, $\emptyset \neq U \subset\subset L^2(0,T;H)$, $F : L^2(0,T;H) \times H^1(0,T;H) \to [0,\infty]$ lower semicontinuous and coercive w.r.t. the strong $\times$ weak topology τ of $L^2(0,T;H) \times H^1(0,T;H)$, $F(u,y) < \infty$ only if $u \in U$, G_{BEN} defined as in* (3.2)*, and $E_\varepsilon := F + \varepsilon^{-1} G_{\mathrm{BEN}}$ for $\varepsilon > 0$.*

Then, $\min E_\varepsilon$ *admits a solution for all $\varepsilon > 0$. Moreover, $E_\varepsilon \xrightarrow{\Gamma} E_0$ with respect to topology τ where $E_0 = F$ on $\{G_{\mathrm{BEN}} = 0\}$ and $E_0 = \infty$ otherwise, and any sequence of quasiminimizers converges, up to a subsequence, to a solution of* $\min E_0$*. In case* $\min E_0$ *admits a unique minimizer, any sequence of quasiminimizers τ-converges to it.*

Proof. In order to prove the statement we apply Lemma 2.1 with the choices $X = L^2(0,T;H) \times H^1(0,T;H)$ and $\tau =$ strong $\times$ weak topology in X.

We start by checking that E_ε is proper. In fact, by letting $u \in U$ and $y \in H^1(0,T;H)$ be the unique solution of $y' + \partial\phi(y) \ni u$ with $y(0) = y_0$ we have that $E_\varepsilon(u,y) = F(u,y) < \infty$.

In order to prove the lower semicontinuity of G_{BEN}, assume that $(u_n, y_n) \xrightarrow{\tau} (u,y)$. As

$H^1(0,T;H) \subset C([0,T];H)$ and U is compact in $L^2(0,T;H)$ we have that

$$\begin{aligned}
&u_n - y_n' \to u - y' \ \text{ weakly in } L^2(0,T;H),\\
&(u_n, y_n) \to (u,y) \ \text{ in } \ L^1(0,T),\\
&y_n(T) \to y(T) \ \text{ weakly in } H.
\end{aligned}$$

This implies that $G_{\mathrm{BEN}}(u,y) \leq \liminf_{n\to\infty} G_{\mathrm{BEN}}(u_n,y_n)$. The equicoercivity of E_ε follows from that of F. □

A remarkable feature of the penalization of problem (1.1) via the Brezis-Ekeland-Nayroles functional relies in the possibility of exploiting convexity. Indeed, in case F is convex, the penalized $F + \varepsilon^{-1} G_{\mathrm{BEN}}$ turns out to be separately convex, the only non-convexity coming from the bilinear term (u,y). This in turn suggests the possibility of implementing some alternate minimization procedure. Note that, in relation with applications to PDEs, the bilinear term (u,y) is usually of lower order.

In the statement of Theorem 3.1 we have assumed F to be coercive. In fact, the functional G_{BEN} itself cannot be expected to be coercive with respect to topology τ. In particular, this would follow by asking ϕ^* to be superquadratic. This would however induce a quadratic bound to ϕ, a quite restrictive assumption, especially in relation to PDEs.

An alternative possibility is that of augmenting G_{BEN} by a coercive term, which would still vanish on solutions of (3.1). A proposal in this direction is in [36], where the following variant of the Brezis-Ekeland-Nayroles functional is presented

$$\tilde{G}_{\mathrm{BEN}}(u,y) = G_{\mathrm{BEN}}(u,y) + \left(\int_0^T \left(\|y'\|^2 - (u,y') \right) \mathrm{d}t + \phi(y(t)) - \phi(y_0) \right)^+ \tag{3.3}$$

with $r^+ := \max\{r, 0\}$. By letting now $E_\varepsilon = F + \varepsilon^{-1} \tilde{G}_{\mathrm{BEN}}$ one can prove the statement of Theorem 3.1 also for a noncoercive functional F, for coercivity for y with respect to the weak topology of $H^1(0,T;H)$ is provided by $\tilde{G}_{\mathrm{BEN}}$.

Before closing this subsection, let us remark that a time-dependent potential ϕ can be considered as well, namely

$$y'(t) + \partial\phi(t, y(t)) \ni u(t) \ \text{ in } H, \text{ for a.e. } t \in (0,T), \quad y(0) = y_0. \tag{3.4}$$

Here, $\phi : (0,T) \times H \to (-\infty, \infty]$ is asked to be measurable with respect to $\mathcal{L} \otimes \mathcal{B}(H)$, where $\mathcal{L}$ is the Lebesgue σ-algebra in $(0,T)$ and $\mathcal{B}(H)$ is the Borel σ-algebra in H, and such that $y \mapsto \phi(t,y)$ is proper, convex, and lower semicontinuous for a.e. $t \in (0,T)$. Problem (3.4) can be equivalently reformulated as $G_{\mathrm{BEN}}(u,y) = 0$ where

$$G_{\mathrm{BEN}}(u,y) = \begin{cases} \displaystyle\int_0^T \Big(\phi(t, y(t)) + \phi^*(t, u(t) - y(t)') - (u(t), y(t)) \Big) \mathrm{d}t & \\ \qquad \displaystyle + \frac{1}{2}\|y(T)\|^2 - \frac{1}{2}\|y_0\|^2 & \text{if } \ y(0) = y_0 \\ \infty & \text{otherwise.} \end{cases}$$

where of course conjugation in ϕ^* is taken with respect to the second variable only. In order to be sure, however, that pairs (u,y) exist with that $G_{\mathrm{BEN}}(u,y) = 0$, some additional assumptions on the time dependence $t \mapsto \phi(t,y)$ is required. The reader is referred to [21, 22, 30, 42] for a collection of classical results in this direction.

3.1 An example

With the aim of illustrating the statement of Theorem 3.1, we investigate the ODE optimal control problem

$$\min\left\{\frac{1}{2}\int_0^1 (y(t)-\mathrm{e}^{-t})^2\mathrm{d}t+\frac{1}{2}\int_0^1 t^2(u(t)-\mathrm{e}^{-t})^2\mathrm{d}t\ : \tag{3.5}\right.$$

$$\left. y'(t)+y(t)=u(t)\equiv u_0\mathrm{e}^{-t},\ u_0\in[0,1],\ y(0)=1\right\}. \tag{3.6}$$

Here, by taking advantage of the linearity of the constraint one can directly compute $y(t)=S(u_0\mathrm{e}^{-t})(t)=\mathrm{e}^{-t}(1+tu_0)$ and

$$\begin{aligned} u_0\mapsto F(S(u_0\mathrm{e}^{-t}),u_0\mathrm{e}^{-t}) &:= \frac{1}{2}\int_0^1 (S(u_0\mathrm{e}^{-t})(t)-\mathrm{e}^{-t})^2\mathrm{d}t+\frac{1}{2}\int_0^1 t^2(u_0\mathrm{e}^{-t}-\mathrm{e}^{-t})^2\mathrm{d}t \\ &= \left(\frac{1}{2}\int_0^1 t^2\mathrm{e}^{-2t}\mathrm{d}t\right)\left(u_0^2+(u_0-1)^2\right) =: \gamma\left(u_0^2+(u_0-1)^2\right) \end{aligned}$$

In particular, the optimal control corresponds to $u_0=1/2$, the optimal solution is $y(t)=\mathrm{e}^{-t}(1+t/2)$, and the minimum of E_0 is

$$F(\mathrm{e}^{-t}(1+t/2),\mathrm{e}^{-t}/2)=\gamma/2=1/16-5/(16\mathrm{e}^2)\sim 0.0202.$$

The ODE is the gradient flow of the potential $\phi(y)=y^2/2$ under the additional forcing u. Correspondingly, the Brezis-Ekeland-Nayroles functional G_{BEN} is given by

$$G_{\mathrm{BEN}}(u,y)=\begin{cases}\displaystyle\int_0^1\left(\frac{1}{2}y^2+\frac{1}{2}(u-y')^2-uy\right)\mathrm{d}t+\frac{1}{2}y^2(1)-\frac{1}{2} & \text{if } y(0)=1,\\ \infty & \text{otherwise.}\end{cases}$$

The penalized optimal control problem reads then

$$\begin{aligned}\min\Bigg\{&\int_0^1\left(\frac{1}{2}(y(t)-\mathrm{e}^{-t})^2+\frac{t^2}{2}(u(t)-\mathrm{e}^{-t})^2+\frac{1}{2\varepsilon}y^2(t)+\frac{1}{2\varepsilon}(u(t)-y'(t))^2-\frac{1}{\varepsilon}u(t)y(t)\right)\mathrm{d}t\\ &+\frac{1}{2\varepsilon}y^2(1)-\frac{1}{2\varepsilon}\ \ :\ \ u(t)\equiv u_0\mathrm{e}^{-t},\ u_0\in[0,1],\ y(0)=1\Bigg\}.\end{aligned}$$

For all given u, the Euler-Lagrange equation for $E_\varepsilon=F+\varepsilon^{-1}G_{\mathrm{BEN}}$ in terms of y_ε is

$$y''(t)-y(t)-\varepsilon y(t)=-(2u_0+\varepsilon)\mathrm{e}^{-t},\quad y'(1)+y(1)=u_0/\mathrm{e}.$$

Complemented with the initial condition $y(0)=1$, these linear relations uniquely identify a critical point y_ε of E_ε. In fact, this is necessarily the unique minimizer of the convex functional $y\mapsto E_\varepsilon(u,y)$ and can be explicitly determined in terms of u_0 as

$$y_{\varepsilon,u_0}(t)=c_{1\varepsilon}\mathrm{e}^{-\alpha_\varepsilon t}+c_{2\varepsilon}\mathrm{e}^{\alpha_\varepsilon t}+\left(\frac{2u_0}{\varepsilon}+1\right)\mathrm{e}^{-t}$$

where we have used the shorthand notation

$$\begin{aligned}
\alpha_\varepsilon &:= (1+\varepsilon)^{1/2},\\
c_{1\varepsilon} &:= \left(\frac{u_0}{\mathrm{e}} - (1+\alpha_\varepsilon)\left(\frac{2u_0}{\varepsilon}+1\right)\right)\left((1-\alpha_\varepsilon)\mathrm{e}^{-\alpha_\varepsilon} - (1+\alpha_\varepsilon)\mathrm{e}^{\alpha_\varepsilon}\right)^{-1},\\
c_{2\varepsilon} &:= -\frac{2u_0}{\varepsilon} - c_{1\varepsilon}.
\end{aligned}$$

The value of $E_\varepsilon(u_0\mathrm{e}^{-t}, y_{\varepsilon,u_0})$ can be explicitly evaluated. An elementary but tedious computation gives

$$\begin{aligned}
E_\varepsilon(u_0\mathrm{e}^{-t}, y_{\varepsilon,u_0}) &= \left(\frac{c_{1\varepsilon}^2}{2} + \frac{c_{1\varepsilon}^2}{2\varepsilon} + \frac{\alpha_\varepsilon^2 c_{1\varepsilon}^2}{2\varepsilon}\right)\frac{\mathrm{e}^{-2\alpha_\varepsilon}-1}{-2\alpha_\varepsilon} + \left(\frac{c_{2\varepsilon}^2}{2} + \frac{c_{2\varepsilon}^2}{2\varepsilon} + \frac{\alpha_\varepsilon^2 c_{2\varepsilon}^2}{2\varepsilon}\right)\frac{\mathrm{e}^{2\alpha_\varepsilon}-1}{2\alpha_\varepsilon}\\
&+ \left(\frac{2u_0^2}{\varepsilon^2} + \frac{1}{2\varepsilon}\left(\frac{2u_0}{\varepsilon}+1\right)^2 + \frac{1}{2\varepsilon}\left(\frac{2u_0}{\varepsilon}+1+u_0\right)^2 - \frac{u_0}{\varepsilon}\left(\frac{2u_0}{\varepsilon}+1\right)\right)\frac{\mathrm{e}^{-2}-1}{-2}\\
&+ \left(\frac{2c_{1\varepsilon}u_0}{\varepsilon} + \frac{c_{1\varepsilon}}{\varepsilon}\left(\frac{2u_0}{\varepsilon}+1\right) + \frac{\alpha_\varepsilon c_{1\varepsilon}}{\varepsilon}\left(\frac{2u_0}{\varepsilon}+1+u_0\right) - \frac{c_{1\varepsilon}u_0}{\varepsilon}\right)\frac{\mathrm{e}^{-\alpha_\varepsilon-1}-1}{-\alpha_\varepsilon-1}\\
&+ \left(\frac{2c_{2\varepsilon}u_0}{\varepsilon} + \frac{c_{2\varepsilon}}{\varepsilon}\left(\frac{2u_0}{\varepsilon}+1\right) - \frac{\alpha_\varepsilon c_{2\varepsilon}}{\varepsilon}\left(\frac{2u_0}{\varepsilon}+1+u_0\right) - \frac{c_{2\varepsilon}u_0}{\varepsilon}\right)\frac{\mathrm{e}^{\alpha_\varepsilon-1}-1}{\alpha_\varepsilon-1}\\
&+ \left(1+\frac{1}{\varepsilon}-\frac{\alpha_\varepsilon^2}{\varepsilon}\right)c_{1\varepsilon}c_{2\varepsilon} + \frac{1}{2\varepsilon}\left(c_{1\varepsilon}\mathrm{e}^{-\alpha_\varepsilon} + c_{2\varepsilon}\mathrm{e}^{\alpha_\varepsilon} + \left(\frac{2u_0}{\varepsilon}+1\right)\mathrm{e}^{-1}\right)^2 - \frac{1}{2\varepsilon}\\
&+ \gamma(u_0-1)^2.
\end{aligned}\tag{3.7}$$

Different curves $u_0 \mapsto E_\varepsilon(u_0\mathrm{e}^{-t}, y_{\varepsilon,u_0})$ for different choices of ε are depicted in Figure 1. We observe that the minimizer and the minimum approach $1/2$ and 0.0202, respectively, as $\varepsilon \to 0$, as expected.

3.2 Gradient flows in dual space

The statement of Theorem 3.1 can be extended to the case of gradient-flow dynamics in dual spaces. Let us introduce a real reflexive Banach space W, densely and continuously embedded into H, so that $W \subset H \subset W^*$ is a classical Gelfand triplet. We consider the problem

$$y' + \partial\phi(y) \ni u \ \text{ in } W^*\text{, a.e. in } (0,T), \quad y(0) = y_0. \tag{3.8}$$

The potential $\phi : W \to \mathbb{R}$ is assumed to be everywhere defined, convex, proper, and lower semicontinuous. The symbol $\partial\phi$ in (3.8) denotes now the subdifferential between W and W^*. This is defined as

$$\xi \in \partial\phi(y) \quad \Leftrightarrow \quad \langle \xi, x-y\rangle \le \phi(x) - \phi(y) \ \ \forall x \in W$$

where $\langle\cdot,\cdot\rangle$ is the duality pairing between W^* and W. We assume ϕ to be bounded as follows

$$\begin{aligned}
&\phi(y) \ge c\|y\|_W^m - \frac{1}{c} \quad \forall y \in W, \quad \phi^*(y^*) \ge c\|y^*\|_{W^*}^{m'} - \frac{1}{c} \quad \forall y^* \in W^*\\
&\|\xi\|_{W^*}^{m'} \le c(1+\|y\|_W^m) \quad \forall y \in W,\ \xi \in \partial\phi(y)
\end{aligned}$$

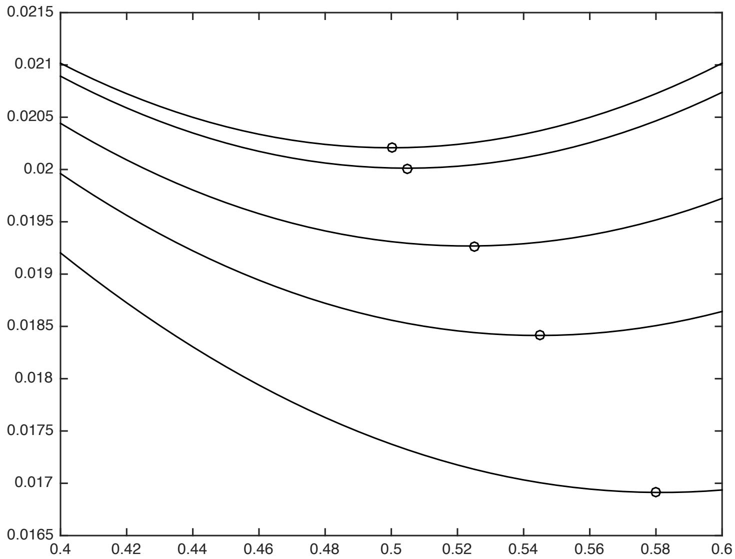

Figure 1: Curves $u_0 \mapsto E_\varepsilon(u_0 \mathrm{e}^{-t}, y_{\varepsilon,u_0})$ from (3.7) for $\varepsilon = 2$, 1, 0.5, 0.1, and 0 (bottom to top). On each curve, the dot indicates the minimizer.

where $m > 1$ and $m' = m/(m-1)$. In particular, the above bounds entail a polynomial control on ϕ of the form

$$\phi(y) \le c\|y\|_W^m + c \quad \forall y \in W, \quad \phi^*(y^*) \le c\|y^*\|_{W^*}^{m'} + c \quad \forall y^* \in W^*$$

which is now compatible with PDE applications.

Given the initial datum $y_0 \in W$ (recall that $D(\phi) = W$), for all $u \in L^{m'}(0,T;W^*)$, the solution $y \in W^{1,m'}(0,T;W^*) \cap L^m(0,T,W)$ of (3.8) exists uniquely. In particular, y solves (3.8) iff $G_{\mathrm{BEN}}(u,y) = 0$ where

$$G_{\mathrm{BEN}}(u,y) = \begin{cases} \displaystyle\int_0^T \Big(\phi(y) + \phi^*(u - y') - \langle u, y\rangle\Big)\,\mathrm{d}t + \frac{1}{2}\|y(T)\|^2 - \frac{1}{2}\|y_0\|^2 & \text{if } \ y(0) = y_0 \\ \infty & \text{otherwise.} \end{cases}$$

The result of Theorem 3.1 can be reformulated in this setting by assuming the set of admissible controls U to be a compact subset of $L^{m'}(0,T;W^*)$ and $F : L^{m'}(0,T;W^*) \times W^{1,m'}(0,T;W^*) \cap L^m(0,T,W) \to [0,\infty]$ to be lower semicontinuous with respect to the strong $\times$ weak topology of $L^{m'}(0,T;W^*) \times W^{1,m'}(0,T;W^*) \cap L^{m'}(0,T,W)$, with $F(u,y) < \infty$ only if $u \in U$. Note that here no coercivity of F is actually needed, for in this case G_{BEN} itself turns out to be coercive, due to the lower bounds on ϕ and ϕ^*.

Once again, G_{BEN} is proper, since it vanishes on solutions to (3.8), which are known to exist. In order to check for the lower semicontinuity of G_{BEN} one would need to recall the embedding $W^{1,m'}(0,T;W^*) \cap L^m(0,T,W) \subset C([0,T];H)$. In particular, the term $\|y(T)\|^2$ turns out to be lower semicontinuous.

3.3 Nonpotential and nonmonotone flows

Originally limited to gradient flows of convex functionals, the Brezis-Ekeland-Nayroles variational approach has been extended to classes of nonpotential monotone flows by

Visintin [39]. By replacing Fenchel duality by the representation theory by Fitzpatrick [14], he noticed that solutions of the nonpotential flow

$$y' + Ay \ni u \quad \text{a.e. in } W^*, \quad y(0) = y_0, \tag{3.9}$$

where $A : W \to 2^{W^*}$ is a maximal monotone, coercive, and *representable* operator and $y_0 \in D(A)$, can be characterized by $G_{\mathrm{BEN}}(u, y) = 0$, where $G_{\mathrm{BEN}} : L^2(0, T; W^*) \times H^1(0, T; W^*) \cap L^2(0, T, W) \to [0, \infty]$ is now given as

$$G_{\mathrm{BEN}}(u, y) = \begin{cases} \displaystyle\int_0^T \Big(f_A(y, u-y')-\langle u, y\rangle\Big)\,\mathrm{d}t + \frac{1}{2}\|y(T)\|^2 - \frac{1}{2}\|y_0\|^2 & \text{if } y(0) = y_0, \\ \infty & \text{otherwise.} \end{cases} \tag{3.10}$$

The function $f_A : W \times W^* \to (-\infty, \infty]$ is convex, lower semicontinuous, with $f_A(y, y^*) \geq \langle y^*, y\rangle$ for all $(y, y^*) \in W \times W^*$, and *represents* the operator A in the following sense

$$y^* \in Ay \quad \Leftrightarrow \quad f_A(y, y^*) = \langle y^*, y\rangle. \tag{3.11}$$

An operator is said to be *representable* when it admits a representing function. All maximal monotone operators are representable, for instance via their *Fitzpatrick function*

$$f_A(y, y^*) := \langle y^*, y\rangle + \sup\{\langle y^* - \tilde{y}^*, \tilde{y} - y\rangle \ : \ \tilde{y} \in W,\ \tilde{y}^* \in A\tilde{y}\}.$$

A monotone operator need however not be cyclic nor maximal to be representable. The reader is referred to [40, 41] for a full account on this theory. By taking advantage of position (3.10), the assertion of Theorem 3.1 can hence be modified to include the case of the differential constraint (3.9) as well.

More generally, the reach of the penalization via the Brezis-Ekeland-Nayroles functional extends even beyond monotone situations. Assume to be given $B : H \times W \to 2^{W^*}$ such that

$$B(h, \cdot) : W \to 2^{W^*} \quad \text{is maximal monotone, } \forall h \in H,$$

$$\begin{aligned} &\forall (h, y) \in H \times W,\ \forall y^* \in B(h, y),\ \forall h_n \to h \text{ in } H \\ &\text{there exists } y_n^* \text{ such that } y_n^* \in B(h_n, y_n) \text{ and } y_n^* \to y^* \text{ in } W^*. \end{aligned}$$

This class of nonmonotone operators $A(y) := B(y, y)$, called *semimonotone* [41], includes the class of *pseudomonotone* operators [9], and it is representable [41, Thm. 4.4] in the sense of (3.11) by means of a weakly lower semicontinuous albeit nonconvex function f_A

$$f_A(y, y^*) := \langle y^*, y\rangle + \sup\{\langle y^* - \tilde{y}^*, \tilde{y} - y\rangle \ : \ \tilde{y} \in W,\ \tilde{y}^* \in B(y, \tilde{y})\}. \tag{3.12}$$

On this basis, the nonmonotone flow

$$y' + A(y) \ni u \quad \text{a.e. in } W^*, \quad y(0) = y_0, \tag{3.13}$$

driven by the semimonotone operator $A(y)$ can be variationally reformulated as $G_{\mathrm{BEN}} = 0$, where G_{BEN} is defined in from (3.10), where however f_A is now defined by (3.12). Note that G_{BEN} is proper and lower semicontinuous with respect to the strong $\times$ weak topology of $L^2(0, T; W^*) \times H^1(0, T; W^*) \cap L^2(0, T, W)$. By letting $E_\varepsilon = F + \varepsilon^{-1} G_{\mathrm{BEN}}$ and assuming again that F is coercive and $F(u, y) < \infty$ only if $u \in U$, the results of Theorem 3.1 can be extended to the case of optimal control problems driven by (3.13) as well.

3.4 Doubly nonlinear flows

A gradient flow can be seen as a particular case of the doubly nonlinear evolution

$$\partial\psi(y') + \partial\phi(y) \ni u \ \text{ in } V^*, \text{ a.e. in } (0,T), \quad y(0) = y_0. \tag{3.14}$$

Here, V is a real reflexive Banach space with $W \subset\subset V$, the symbol ∂ refers to the subdifferential between V and V^*, and $\psi : V \to [0,\infty)$ is a second convex, proper, lower semicontinuous functional defined on the whole V. More precisely, we assume ψ to fulfill $0 \in \partial\psi(0)$ and to be of polynomial growth, namely

$$\begin{aligned}
&c\|y'\|_V^p - \frac{1}{c} \le \langle w, y'\rangle, \quad \|w\|_{V^*}^{p'} \le c(1 + \|y'\|_V^p) \quad \forall y' \in V,\, w \in \partial\psi(y')\\
&\psi^*(w) \ge c\|w\|_{V^*}^{p'} - \frac{1}{c} \quad \forall w \in V^*
\end{aligned}$$

for $p > 1$ and $p' = p/(p-1)$. Additionally, we assume $D(\phi) = W$ and the coercivity

$$\phi(y) \ge c\|y\|_W^m - \frac{1}{c} \quad \forall y \in W$$

for some $m > 1$. In [36] a doubly nonlinear version of the Brezis-Ekeland-Nayroles functional is addressed. In particular, one has that $(u,y,w) \in L^{p'}(0,T;V^*) \times W^{1,p'}(0,T;V^*) \cap L^m(0,T;W) \times L^{p'}(0,T;V^*)$ solve

$$w \in \partial\psi(y'), \quad \partial\phi(y) \ni u - w \ \text{ a.e. in } (0,T), \quad y(0) = y_0$$

iff $G_{\mathrm{BEN}}(u,y,w) = 0$, where $G_{\mathrm{BEN}} : L^{p'}(0,T;V^*) \times W^{1,p'}(0,T;V^*) \cap L^m(0,T;W) \times L^{p'}(0,T;V^*) \to [0,\infty]$ is now defined as

$$G_{\mathrm{BEN}}(u,y,w) = \begin{cases} \left(\displaystyle\int_0^T \Big(\psi(y') + \psi^*(w) - \langle u, y'\rangle\Big)\,\mathrm{d}t + \phi(y(T)) - \phi(y_0)\right)^+ & \\ \quad + \displaystyle\int_0^T \Big(\phi(y) + \phi^*(u-w) - \langle u-w, y\rangle\Big)\,\mathrm{d}t & \text{if } y(0) = y_0 \\ \infty & \text{otherwise.} \end{cases}$$

Indeed, the two nonnegative integrals in the definition of G_{BEN} correspond to the two relations $w \in \partial\psi(y')$ and $\partial\phi(y) \ni u - w$, respectively. At the price of introducing the new variable w, one can penalize the differential constraint (3.14) by minimizing $(u,y,w) \mapsto E_\varepsilon(u,y,w) = F(u,y,w) + \varepsilon^{-1}G_{\mathrm{BEN}}(u,y,w)$. Again, the results of Theorem 3.1 can be extended to this situation. In particular, it can be proved that G_{BEN} is proper and lower semicontinuous with respect to the strong $\times$ weak $\times$ weak topology of $L^{p'}(0,T;V^*) \times W^{1,p'}(0,T;V^*) \cap L^m(0,T;W) \times L^{p'}(0,T;V^*)$. Moreover, it turns out to be coercive as well, as soon as it is restricted to $u \in U$. In particular, no coercivity has to be assumed on F in this case. Indeed, G_{BEN} is here the doubly nonlinear version of the former (3.3), which was in fact introduced to ensure coercivity.

4 De Giorgi principle

Let us now turn out attention to the penalization (1.2) by means of a variational reformulation of dissipative evolution, following the general approach to gradient flows from [12].

Consider again the classical gradient flow in a Hilbert space (3.1) where now the potential $\phi : H \to (-\infty, \infty]$ is asked to be lower semicontinuous and proper, possibly being nonconvex. To keep notation to a minimum, let us assume $\phi = \phi_1 + \phi_2$ with ϕ_1 convex, proper, and lower semicontinuous, and $\phi_2 \in C^{1,1}$. Then, by letting $\partial\phi$ denote the classical *Fréchet subdifferential*, namely

$$\xi \in \partial\phi(y) \quad \Leftrightarrow \quad y \in D(\phi) \ \text{ and } \ \liminf_{w \to y} \frac{\phi(w) - \phi(y) - (\xi, w - y)}{\|y - w\|} \geq 0,$$

(note that the Fréchet subdifferential coincides with the subdifferential of convex analysis on convex functions) we have that $\partial\phi = \partial\phi_1 + D\phi_2$. We will additionally assume $\partial\phi_1$ to be single-valued, whenever nonempty. More general settings are discussed in Subsection 4.2 below.

Solutions to (3.1) correspond to $G_{\mathrm{DG}}(u, y) = 0$, where the functional $G_{\mathrm{DG}} : L^2(0,T;H) \times H^1(0,T;H) \to [0, \infty]$ is defined as

$$G_{\mathrm{DG}}(u,y) = \begin{cases} \displaystyle\int_0^T \left(\frac{1}{2}\|y'\|^2 + \frac{1}{2}\|\partial\phi(y) - u\|^2 - (u, y')\right) \mathrm{d}t + \phi(y(T)) - \phi(y_0) \\ \qquad \text{if } y \in D(\partial\phi) \text{ a.e. and } y(0) = y_0 \\ \infty \quad \text{otherwise.} \end{cases} \tag{4.1}$$

Due to its ties with the variational theory of steepest decent in metric spaces from [12] we call G_{DG} *De Giorgi* functional. In (4.1) we used the notation $D(\partial\phi)$ to indicate the essential domain of $\partial\phi$, namely $D(\partial\phi) = \{y \in H \ : \ \partial\phi(y) \neq \emptyset\}$. Note that, by [34, Lemma 3.4], the map $t \mapsto \partial\phi(y(t))$ is measurable whenever $y \in H^1(0,T;H)$ with $y \in D(\partial\phi)$ a.e. The reformulation of the gradient flow (3.1) via G_{DG} is based on the computation of the squared residual of (3.1), namely,

$$\begin{aligned} \int_0^T \frac{1}{2}\|y' + \partial\phi(y) - u\|^2 \mathrm{d}t &= \int_0^T \left(\frac{1}{2}\|y'\|^2 + \frac{1}{2}\|\partial\phi(y) - u\|^2 + (y', \partial\phi(y) - u)\right) \mathrm{d}t \\ &= G_{\mathrm{DG}}(u,y) \quad \text{if } y(0) = y_0. \end{aligned}$$

The latter computation hinges on the chain rule $(\partial\phi(y), y') = (\phi \circ y)'$, which holds in the case of $\phi = \phi_1 + \phi_2$ in the following precise form [6, Lemme 3.3]

$$\begin{aligned} & y \in H^1(0,T;H), \quad \xi \in L^2(0,T;H), \quad \xi \in \partial\phi(y) \ \text{ a.e. in } (0,T) \\ & \Rightarrow \quad \phi \circ y \in AC(0,T) \ \text{ and } \ (\phi \circ y)' = (\xi, y') \ \text{ a.e. in } (0,T). \end{aligned} \tag{4.2}$$

Indeed, note that $\partial\phi(y) \in L^2(0,T;H)$ if $G_{\mathrm{BEN}}(u,y) < \infty$. The main result of this section is the following.

Theorem 4.1 (Gradient flows, DG principle). *Let $\phi = \phi_1 + \phi_2 : H \to (-\infty, \infty]$ have compact sublevels with ϕ_1 proper, convex, and lower semicontinuous, $\partial\phi_1$ single-valued, and $\phi_2 \in C^{1,1}$. Moreover, let $y_0 \in D(\phi)$, $\emptyset \neq U \subset\subset L^2(0,T;H)$, $F : L^2(0,T;H) \times H^1(0,T;H) \to [0,\infty]$ be lower semicontinuous w.r.t. the strong $\times$ weak topology τ of $L^2(0,T;H) \times H^1(0,T;H)$, $F(u,y) < \infty$ only if $u \in U$, $G_{\rm DG}$ be defined as in (4.12), and $E_\varepsilon := F + \varepsilon^{-1} G_{\rm DG}$ for $\varepsilon > 0$.*

Then, $\min E_\varepsilon$ admits a solution for all $\varepsilon > 0$. Moreover $E_\varepsilon \xrightarrow{\Gamma} E_0$ with respect to topology τ where $E_0 = F$ on $\{G_{\rm DG} = 0\}$ and $E_0 = \infty$ otherwise, and any sequence of quasiminimizers converges, up to a subsequence, to a solution of $\min E_0$. In case $\min E_0$ admits a unique minimizer, any sequence of quasiminimizers converge to it with respect to τ.

Proof. The statement follows by applying Lemma 2.1 in the space $X = L^2(0,T;H) \times H^1(0,T;H)$ endowed with its strong $\times$ weak topology τ.

Let $u \in U$ and let $y \in H^1(0,T;H)$ be the unique solution of $y' + \partial\phi(y) \ni u$ with $y(0) = y_0$. As we have that $E_\varepsilon(u,y) = F(u,y) < \infty$, the functional E_ε is clearly proper.

Functional F is τ-lower semicontinuous by assumption. In order to check the τ-lower semicontinuity of $G_{\rm DG}$ let $(u_\varepsilon, y_\varepsilon) \xrightarrow{\tau} (u,y)$ be given. With no loss of generality, one can assume $\sup_\varepsilon G_{\rm DG}(u_\varepsilon, y_\varepsilon) \le c < \infty$. In particular, we can assume that y'_ε and $\partial\phi(u_\varepsilon)$ are uniformly bounded in $L^2(0,T;H)$. By means of the chain rule (4.2) we obtain that for all $t \in [0,T]$

$$\begin{aligned} \phi(y_\varepsilon(t)) - \phi(y_0) &= \int_0^t (\phi \circ y)' \, \mathrm{d}t = \int_0^t (\partial\phi(y), y') \mathrm{d}t \\ &\le \|\partial\phi(y)\|_{L^2(0,T,H)} \|y'\|_{L^2(0,T,H)} < \infty \end{aligned} \tag{4.3}$$

independently of $t \in [0,T]$ and $\varepsilon > 0$. This implies that $t \mapsto \phi(y_\varepsilon(t))$ is uniformly bounded. As the sublevels of ϕ are compact, this yields that there exists $K \subset\subset H$ such that $y_\varepsilon(t) \in K$ for all $t \in [0,T]$ and $\varepsilon > 0$. The uniform bound on y'_ε gives that y_ε are equicontinuous and the Ascoli-Arzelà Theorem implies that, up to not relabeled subsequences, $y_\varepsilon \to y$ strongly in $C([0,T];H)$. This entails that $\partial\phi(y_\varepsilon) \to \partial\phi(u)$ in $L^2(0,T,H)$ since $\partial\phi$ is strongly $\times$ weakly closed as subset of $L^2(0,T;H) \times L^2(0,T;H)$. Moreover, the strong convergence of y_ε in $C([0,T];H)$ implies that $y_\varepsilon(T) \to y(T)$ strongly in H, so that $\phi(y(T)) \le \liminf_{\varepsilon\to 0} \phi(y_\varepsilon(T))$ as ϕ is lower semicontinuous. Since $(u_\varepsilon, y'_\varepsilon) \to (u, y')$ strongly in $L^1(0,T)$, we can pass to lower limits in all terms in $G_{\rm DG}(u_\varepsilon, y_\varepsilon)$ and thus check that $G_{\rm DG}(u,y) \le \liminf_{\varepsilon\to 0} G_{\rm DG}(u_\varepsilon, y_\varepsilon)$.

The τ-equicoercivity of E_ε follows as U is compact in $L^2(0,T;H)$ and $G_{\rm DG}(u,y)$ controls the $L^2(0,T;H)$ norm of y'. □

Before closing this subsection, let us record that in the former case of (3.6) the two functionals $G_{\rm BEN}$ and $G_{\rm DG}$ coincide. In particular, Figure 1 illustrates the convergence of the penalization via $G_{\rm DG}$ as well. By considering in that same linear ODE example $\phi(y) = \lambda y^2/2$ with $\lambda > 0$ instead of $\phi(y) = y^2/2$ one finds the relation $G_{\rm BEN}(u,y) = \lambda G_{\rm DG}(u,y)$, which implies that the minimizers of $F + \varepsilon^{-1} G_{\rm BEN}$ and $F + (\varepsilon/\lambda)^{-1} G_{\rm DG}$ coincide. Hence, for fixed $\varepsilon > 0$ one has that $G_{\rm BEN}$, respectively $G_{\rm DG}$, delivers the best approximation in terms of minimum and minimizer if $\lambda < 1$, respectively $\lambda > 1$. This in particular proves that, in general, no functional a priori dominates the other in terms of accuracy of the approximation for fixed ε.

4.1 A numerical simulation

In order to present a second illustration of the penalization procedure, let us resort to a nonlinear ODE. We consider the optimal control problem

$$\min\left\{\frac{1}{2}\int_0^1 (y(t)-1)^2\mathrm{d}t + \frac{1}{2}(u-2)^2 \ : \ y'(t)+y^3(t)=u \text{ for } t\in[0,1],\ y(0)=1\right\}. \tag{4.4}$$

with $u\in\mathbb{R}$. By evaluating $u\mapsto F(u,S(u))$ with Matlab, where $y=S(u)$ is the unique solution to $y'+y^3=u$ with $y(0)=1$, one finds a unique optimal $u\sim 1.016$ and, correspondingly, $F(u,S(u))\sim 0.4917$.

The De Giorgi penalized problem for $\varepsilon>0$ reads

$$\begin{aligned}\min\big(F+\varepsilon^{-1}G_{\mathrm{DG}}\big) &= \min\left\{\frac{1}{2}\int_0^1 (y(t)-1)^2\mathrm{d}t + \frac{1}{2}(u-2)^2\right.\\ &\left.+\frac{1}{\varepsilon}\left(\int_0^1\left(\frac{1}{2}(y'(t))^2+\frac{1}{2}(y^3(t)-u)^2-uy'(t)\right)\mathrm{d}t+\frac{1}{4}y^4(1)-\frac{1}{4}\right) \ : \ y(0)=1\right\}.\end{aligned}$$

The corresponding Euler-Lagrange equations, complemented by the initial condition, reads

$$-y''(t)+3(y^3(t)-u)y^2(t)=0 \ \text{ for } t\in(0,1), \quad y'(1)+y^3(1)=u, \quad y(0)=1. \tag{4.5}$$

Given u, by numerically solving the latter boundary-value problem with Matlab, one finds a critical point $y_{\varepsilon,u}$ of E_ε and evaluates $u\mapsto E_\varepsilon(u,y_{\varepsilon,u})$. The results of this simulation are illustrated in Figure 2, showing convergence of minima and minimizers as $\varepsilon\to 0$.

4.2 More general potentials

The proof of Theorem 4.1 can be extended to include some more general classes of potentials. A first generalization of the theory allows to treat the case of $\phi=\phi_1+\phi_2$ with $\partial\phi_1$ not single-valued. In this case, one starts by equivalently rewriting problem (3.1) as

$$y'=\big(u-\partial\phi(y)\big)^\circ \quad \text{in } H, \text{ a.e. in } (0,T), \quad y(0)=y_0. \tag{4.6}$$

Here, $\big(u-\partial\phi(y)\big)^\circ$ denotes the unique element of minimal norm in the convex and closed set $u-\partial\phi(y)=u-\partial\phi_1(y)-D\phi_2(y)$. Let us briefly comment on the equivalence of problems (3.1) and (4.6). On the one hand, a solution to (4.6) clearly solves (3.1) as well. On the other hand, solutions to (3.1) are unique: Let y_1 and y_2 be two solutions, and write

$$y_1'-y_2'+\xi_1-\xi_2=D\phi_2(y_1)-D\phi_2(y_2) \quad \text{in } H, \text{ a.e. in } \ (0,T),$$

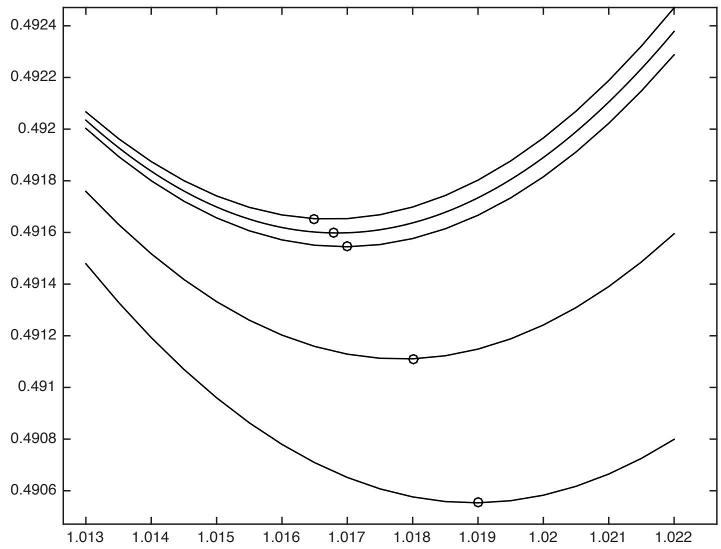

Figure 2: Curves $u \mapsto E_\varepsilon(u, y_{\varepsilon,u})$ for problem (4.4) for $\varepsilon = 1$, 0.5, 0.1, 0.05, and 0 (bottom to top). On each curve, the dot indicates the minimizer.

where $\xi_i \in \partial\phi_1(y_i)$ a.e. in $(0,T)$, for $i = 1, 2$. Test the latter equality by $y_1 - y_2$ and integrate on $(0,t)$. By the monotonicity of $\partial\phi_1$ and the Lipschitz continuity of $D\phi_2$ we obtain

$$\frac{1}{2}\|y_1(t) - y_2(t)\|^2 \leq \|D^2\phi\|_{L^\infty} \int_0^t \|y_1(s) - y_2(s)\|^2 \mathrm{d}s$$

and $y_1 = y_2$ follows by the Gronwall Lemma.

Equality (4.6) can then be equivalently recast as $G_{\mathrm{DG}}(u, y) = 0$ along with the choice

$$G_{\mathrm{DG}}(u,y) = \begin{cases} \displaystyle\int_0^T \left(\frac{1}{2}\|y'\|^2 + \frac{1}{2}\|(\partial\phi(y) - u)^\circ\|^2 - (u, y')\right) \mathrm{d}t + \phi(y(T)) - \phi(y_0) \\ \quad \text{if } y \in D(\partial\phi_1) \text{ a.e. and } y(0) = y_0 \\ \infty \quad \text{otherwise.} \end{cases}$$

Note that G_{DG} is proper, as it vanishes on solutions of the gradient flow. In particular, if $G_{\mathrm{DG}}(u, y) < \infty$ we have $y \in H^1(0, T; H)$ and we can find $\xi \in L^2(0, T; H)$ such that $\xi - u = (\partial\phi(y) - u)^\circ$ and $\xi \in \partial\phi(y)$ a.e. Then, by means of the chain rule (4.2) one computes

$$(\phi \circ y)' = (\xi, y') = (\xi - u, y') + (u, y') \quad \text{a.e. in } (0, T).$$

as well as the chain of equivalences

$$\begin{aligned} y' &= (u - \partial\phi(y))^\circ \ \text{ a.e.} \\ &\Leftrightarrow 0 = \frac{1}{2}\|y' + \xi - u\|^2 = \frac{1}{2}\|y'\|^2 + \frac{1}{2}\|\xi - u\|^2 + (\xi - u, y') \ \text{ a.e.} \\ &\Leftrightarrow 0 = \frac{1}{2}\|y'\|^2 + \frac{1}{2}\|\xi - u\|^2 - (u, y') + (\phi \circ y)' \ \text{ a.e.} \\ &\Leftrightarrow G_{\mathrm{DG}}(u, y) = 0. \end{aligned}$$

In order to extend the results of Theorem 4.1 to this case, one just needs to check that, by replacing the term $\partial\phi(y) - u$ with $(\partial\phi(y) - u)^\circ$ in the functional, coercivity and lower semicontinuity still hold. As for the first, one still has that ϕ is controlled along trajectories as in (4.3), since $(\partial\phi(y) - u)^\circ = \xi - u$ a.e., for some $\xi \in \partial\phi(u)$ a.e. As for lower semicontinuity, one just needs to be able to pass to the lim inf in the term containing $(\partial\phi(y)-u)^\circ$. By letting $y_\varepsilon \to y$ strongly in $C([0,T];H)$ and $\eta_\varepsilon = (u-\partial\phi(y_\varepsilon))^\circ \to \eta$ weakly in $L^2(0,T;H)$ one finds that $\xi_\varepsilon := u-\eta_\varepsilon \in \partial\phi(y_n)$ a.e. are such that $\xi_\varepsilon \to u-\eta =: \xi$ weakly in $L^2(0,T;H)$. Moreover, by the strong $\times$ weak closure of $\partial\phi$ we have that $\xi \in \partial\phi(y)$ a.e. We conclude that

$$\begin{aligned}\frac{1}{2}\int_0^T \|(\partial\phi(y) - u)^\circ\|^2 \mathrm{d}t &\le \frac{1}{2}\int_0^T \|\xi - u\|^2 \mathrm{d}t = \frac{1}{2}\int_0^T \|\eta\|^2 \mathrm{d}t \\ &\le \liminf_{\varepsilon\to 0}\frac{1}{2}\int_0^T \|\eta_\varepsilon\|^2 \mathrm{d}t = \liminf_{\varepsilon\to 0}\frac{1}{2}\int_0^T \|(\partial\phi(y_\varepsilon) - u)^\circ\|^2 \mathrm{d}t\end{aligned}$$

and lower semicontinuity of G_{DG} follows.

Even more generally, the theory could be adapted to potential which are not $C^{1,1}$ perturbations of convex functions. The reader is referred to Rossi & Savaré [34] where a general frame for existence of solutions to gradient flows on nonconvex functionals is addressed. In this context, weaker notions of (sub)differential are introduced and the validity of a corresponding chain rule as in (4.2) is discussed. In particular, examples of operators fulfilling a suitable chain rule are presented, including classes of dominated concave perturbations of convex functions.

Let us mention that the validity of a chain rule *equality*, albeit of a paramount importance in order to relate the minimization of G_{DG} to the solution of (3.1), is actually not needed to prove Theorem 4.1. In fact, the chain rule (4.2) has been used there just to check that the potential ϕ remains uniformly bounded along trajectories. In particular, a suitable chain-rule *inequality* would serve for this purpose as well.

4.3 Generalized gradient flows

The De Giorgi functional approach can be adapted to encompass *generalized gradient flows*, namely relations of the form

$$\partial\psi(y,y') + \partial\phi(y) \ni u \quad \text{for a.e. } t \in (0,T), \quad y(0) = y_0. \tag{4.7}$$

Here, $\psi : H \times H \to [0,\infty)$ and $\partial\psi(y,y')$ denotes partial subdifferentiation with respect to the second variable only. More precisely, we assume that the map $v \in H \mapsto \psi(y,v)$ is convex and lower semicontinuous for all $y \in H$, the map $(y,v,w) \in H \times H \times H \mapsto \psi(y,v) + \psi^*(y,w)$ is weakly lower semicontinuous and

$$\psi(y,v) + \psi^*(y,w) \ge c\|v\|^p + c\|w\|^{p'} \quad \forall y,\ v,\ w \in H \tag{4.8}$$

and some $p > 1$ where $p' = p/(p-1)$ and the Legendre-Fenchel conjugation is taken with respect to the second variable only. An example for ψ satisfying (4.8) is $\psi(y,y) = \beta(y)|y|^p$, where $p > 1$ and β is sufficiently smooth, uniformly positive, and bounded. Note that this includes the case of *doubly nonlinear* flows. As in Theorem 4.1, we assume for simplicity that $\partial\phi = \partial\phi_1 + D\phi_2$ and is single-valued.

Solutions to (4.7) can be characterized via $G_{\rm DG}(u,y) = 0$ where $G_{\rm DG} : L^p(0,T;H) \times W^{1,q}(0,T;H) \to [0,\infty]$ is defined as

$$G_{\rm DG}(u,y) = \begin{cases} \displaystyle\int_0^T (\psi(y,y') + \psi^*(y,u - \partial\phi(y)) - (u,y'))\,{\rm d}t + \phi(y(T)) - \phi(y_0) \\ \qquad \text{if } y \in D(\partial\phi) \text{ a.e and } y(0) = y_0 \\ \infty \quad \text{otherwise.} \end{cases} \tag{4.9}$$

This can be checked by equivalently rewriting

$$\begin{aligned} &\partial\psi(y,y') + \partial\phi(y) \ni u \quad \text{a.e.} \\ &\Leftrightarrow \psi(y,y') + \psi^*(y,u - \partial\phi(y)) - (u - \partial\phi(u),y') = 0 \quad \text{a.e.} \\ &\Leftrightarrow \psi(y,y') + \psi^*(y,u - \partial\phi(y)) - (u,y') + (\phi\circ y)' = 0 \quad \text{a.e.} \\ &\Leftrightarrow G_{\rm DG}(u,y) = 0. \end{aligned}$$

Indeed, the last equivalence follows by integrating the second-last relation in time, in one direction, and by realizing that the integrand is always nonnegative, in the other direction.

By replacing $\|\cdot\|^2/2$ by $\psi(y,\cdot)$ under assumption (4.8), an analogous statement to Theorem 4.1 holds. More precisely, by assuming $F : L^p(0,T;H) \times W^{1,p'}(0,T;H) \to [0,\infty)$ to be lower semicontinuous in $X = L^p(0,T;H) \times W^{1,p'}(0,T;H)$ with respect to the strong$\times$ weak topology and U to be compact in $L^p(0,T;H)$, one can reproduce the former argument. Note however that extra conditions have to be imposed in such a way that pairs with $G_{\rm BEN}(u,y) = 0$ exist.

4.4 GENERIC flows

The applicability of the penalization technique via the De Giorgi functional can be extended to classes of so-called GENERIC flows (General Equations for Non-Equilibrium Reversible-Irreversible Coupling). These are systems of the form

$$y' = L(y)\,DE(y) - K(y)(\partial\phi(y) - u) \quad \text{for a.e. } t \in (0,T), \quad y(0) = y_0. \tag{4.10}$$

Here, $-\phi$ is to be interpreted as the *entropy* and will have the property of being nondecreasing in time. The functional $E : H \to \mathbb{R}$ represents an *energy*, to be conserved along trajectories instead. For the sake of simplicity, we assume E to be Fréchet differentiable, with a linearly bounded, strongly $\times$ weakly closed differential DE. The mapping $K : H \to \mathcal{L}(H)$ (linear and continuous operators) is the so called *Onsager* operator and is asked to be continuous with symmetric and positive semidefinite values. On the other hand, the operator $L : H \to \mathcal{L}(H)$ is required to be continuous with antiselfadjoint values, namely $L^*(y) = -L(y)$.

The GENERIC formalism [19] is a general approach to the variational formulation of physical models and is particularly tailored to the unified treatment of coupled conservative and dissipative dynamics. Potentials and operators are related by the following structural assumptions

$$L^*(y)\partial\phi(y) = K^*(y)DE(y) = 0. \tag{4.11}$$

These guarantee that solutions of (4.10) are such that $(E \circ y)' = 0$ and $(-\phi \circ y)' \geq 0$, namely energy is conserved and entropy increases along trajectories. To date, GENERIC has been successfully applied to a variety of situations ranging from complex fluids [19], to dissipative quantum mechanics [28], to thermomechanics [2, 27], and to the Vlasov-Fokker-Planck equation [13].

By defining the convex potential $\xi \mapsto \psi^*(y,\xi) = (K(y)\xi,\xi)/2$, so that $K(y) = \partial\psi^*(y,\cdot)$ (subdifferential with respect to the second variable only), problem (4.10) can be reformulated as $G_{\rm DG}(u,y) = 0$ where now $G_{\rm DG} : L^2(0,T;H) \times H^1(0,T;H) \to [0,\infty]$ is defined as

$$G_{\rm DG}(u,y) = \begin{cases} \displaystyle\int_0^T \Big(\psi(y, y'{-}L(y)\,DE(y)) + \psi^*(y, u{-}\partial\phi(y))\Big)\,{\rm d}t \\ \quad \displaystyle - \int_0^T (u, y'{-}L(y)\,DE(y))\,{\rm d}t + \phi(y(T)) - \phi(y_0) \\ \quad \text{if } y \in D(\partial\phi) \text{ a.e. and } y(0) = y_0 \\ \infty \quad \text{otherwise.} \end{cases} \tag{4.12}$$

In fact, we have the following chain of equivalencies

$$\begin{aligned} & y' = L(y)\,DE(y) - K(y)(\partial\phi(y) - u) \ \text{ a.e.} \\ \Leftrightarrow\ & \psi(y, y'{-}L(y)\,DE(y)) + \psi^*(y, u{-}\partial\phi(y)) - \big(y'{-}L(y)\,DE(y), u{-}\partial\phi(y)\big) = 0 \ \text{ a.e.} \\ \Leftrightarrow\ & \psi(y, y'{-}L(y)\,DE(y)) + \psi^*(y, u{-}\partial\phi(y)) \\ & \quad - (u, y' - L(y)\,DE(y)) - (DE(y), L^*(y)\partial\phi(y)) + (\phi \circ y)' = 0 \ \text{ a.e.} \\ \Leftrightarrow\ & G_{\rm DG}(u,y) = 0. \end{aligned}$$

Again, the last equivalence follows by integration in time.

The statement of Theorem 4.1 can be extended to cover the case of GENERIC flows as well. Let us assume from the very beginning that for all $u \in H$ there exists y such that $G_{\rm DG}(u,y) = 0$. In applications K and ϕ are often degenerate (see below). Coercivity for the sole $G_{\rm DG}$ is hence not to be expected. In order to state a general result, let us hence assume F itself to be lower semicontinuous and coercive with respect to the strong $\times$ weak topology of $L^2(0,T;H) \times H^1(0,T;H)$. Moreover, let F be coercive with respect to the strong $\times$ strong topology of $L^2(0,T;H) \times C([0,T];H)$ on sublevels of ϕ and to control the $L^2(0,T;H)$ norm of $\partial\phi(y)$ (alternatively, let $\partial\phi(y)$ be linearly bounded). Eventually, we ask ψ^* and ψ to be lower semicontinuous in the following sense

$$\begin{aligned} & \psi(y,\eta) + \psi^*(y,\xi) \leq \liminf_{\varepsilon\to 0} \big(\psi(y_\varepsilon,\eta_\varepsilon) + \psi^*(y_\varepsilon,\xi_\varepsilon)\big) \\ & \quad \forall y_\varepsilon \to y \ \text{ strongly in } \ C([0,T];H) \ \text{ with } \ \sup \phi(y_\varepsilon(t)) < \infty \\ & \quad \text{and } \ (\eta_\varepsilon,\xi_\varepsilon) \to (\eta,\xi) \ \text{ weakly in } \ L^2(0,T;H)^2. \end{aligned} \tag{4.13}$$

Owing to the assumptions on F, in order to reproduce the argument of Theorem 4.1 in this setting, one is left to check the lower semicontinuity of $G_{\rm DG}$. Let $(u_\varepsilon, y_\varepsilon) \to (u,y)$ strongly $\times$ weakly in $L^2(0,T;H) \times H^1(0,T;H)$ and assume with no loss of generality that $\partial\phi(y_\varepsilon)$ is bounded in $L^2(0,T;H)$. By arguing as in (4.3) one can bound $t \mapsto \phi(y_\varepsilon(t))$ so that all trajectories belong to a sublevel of ϕ. From the strong coercivity of F on sublevels of ϕ we deduce strong compactness in $C([0,T];H)$ for y_ε, so that $y_\varepsilon \to y$ uniformly, up to not relabeled subsequences. As DE is assumed to be strongly $\times$ weakly closed and L is continuous, we have that $y'_\varepsilon - L(y_\varepsilon)\,DE(y_\varepsilon) \to y' - L(y)\,DE(y)$ weakly in $L^2(0,T;H)$. On the other hand, the strong $\times$ weak closure of $\partial\phi$ ensures that, again without relabeling, $\partial\phi(y_\varepsilon) \to \partial\phi(y)$ weakly in $L^2(0,T;H)$. We can hence make use of (4.13) and deduce the lower semicontinuity of G_{DG}.

Before closing this discussion, let us give an example of an elementary GENERIC system fitting into this abstract setting. Consider the thermalized oscillator problem

$$q'' + \nu q' + \lambda q + \theta = 0, \tag{4.14}$$
$$\kappa\theta' = \nu(q')^2 + \theta q'. \tag{4.15}$$

Here, $y = (q,p,\theta) \in \mathbb{R}^3 =: H$ where q represents the state of the oscillator, p is its momentum, and $\theta > 0$ is the absolute temperature. The nonnegative constants ν, λ, and κ are the viscosity parameter, the elastic modulus, and the heat capacity, respectively. Relations (4.14) and (4.14) express the conservation of momentum and energy, respectively.

In order to reformulate (4.14)-(4.15) as a GENERIC system, we specify the free energy of the system as

$$\Psi(y) = \frac{\lambda}{2}q^2 + q\theta - \kappa\theta\ln\theta.$$

Moving from this, the entropy $-\phi$ and the total energy E are derived by the classical Helmholtz relations as

$$-\phi(y) = -\partial_\theta\Psi = -q + \kappa\ln\theta + \kappa, \quad E(y) = \frac{1}{2}p^2 + \Psi + \theta\phi = \frac{1}{2}p^2 + \frac{\lambda}{2}q^2 + \kappa\theta.$$

In particular, we have that

$$DE(y) = (\lambda q, p, \kappa), \quad \partial\phi(y) = (-1, 0, \kappa/\theta).$$

By defining the mappings K and L as

$$K(y) = \nu\theta\begin{pmatrix} 0 & 0 & 0 \\ 0 & 1 & -p/\kappa \\ 0 & -p/\kappa & p^2/\kappa^2 \end{pmatrix}, \quad L(y) = \begin{pmatrix} 0 & 1 & 0 \\ -1 & 0 & -\theta/\kappa \\ 0 & \theta/\kappa & 0 \end{pmatrix},$$

we readily check that the compatibility conditions (4.11) hold and that system (4.14)-(4.15) takes the form in (4.10). By computing the conjugate we find

$$\psi^*(y,\xi) = \frac{\nu\theta}{2}(\xi_2 - p\xi_3/\kappa)^2, \quad \psi(y,\eta) = \begin{cases} \dfrac{1}{2\nu\theta}\eta_2^2 & \text{if } \eta_1 = \eta_3 + py_2/\kappa = 0, \\ \infty & \text{otherwise} \end{cases}$$

for all $y = (q,p,\theta) \in \mathbb{R}^3$ with $\theta > 0$ and for all $(\xi,\eta) \in \mathbb{R}^2$. In particular, the lower semicontinuity (4.13) follows as $\sup\phi(y_\varepsilon(t)) < \infty$ implies that $\theta_\varepsilon \geq c > 0$ for some c, hence $1/\theta_\varepsilon \to 1/\theta$ in $C([0,T])$.

In order to give a concrete example of target functional F choose

$$F(u,y) = \frac{1}{2}\int_0^T |y - y_{\rm target}|^2 {\rm d}t + \frac{1}{2}\int_0^T |y' - y'_{\rm target}|^2 {\rm d}t + \int_0^T |1/\theta - 1/\theta_{\rm target}|^2 {\rm d}t + \int_0^T |u|^2\,{\rm d}t + \int_0^T |u'|^2\,{\rm d}t$$

for some given $y_{\rm target} = (q_{\rm target}, p_{\rm target}, \theta_{\rm target}) \in H^1(0,T;H)$ with $1/\theta_{\rm target} \in L^2(0,T)$. The functional F is coercive with respect to the strong $\times$ weak topology of $L^2(0,T;H) \times H^1(0,T;H)$, as well as to the strong$\times$strong topology of $L^2(0,T;H) \times C([0,T];H)$ on sublevels of ϕ. Moreover, it controls the $L^2(0,T;H)$ norm of $\partial\phi(y)$. Hence, the abstract setting described above applies.

References

[1] G. Auchmuty, Saddle-points and existence-uniqueness for evolution equations, Differential Integral Equations, **6** (1993), 1161–1171.

[2] F. Auricchio, E. Boatti, A. Reali, and U. Stefanelli, Gradient structures for the thermomechanics of shape-memory materials, Comput. Methods Appl. Mech. Engrg., **299** (2016), 440–469.

[3] M. Bergounioux, A penalization method for optimal control of elliptic problems with state constraints, SIAM J. Control Optim., **30** (1992), 305–323.

[4] M. Bergounioux, Optimal control of parabolic problems with state constraints: a penalization method for optimality conditions, Appl. Math. Optim., **29** (1994), 285–307.

[5] M. Bergounioux, Optimal control of problems governed by abstract elliptic variational inequalities with state constraints, SIAM J. Control Optim., **36** (1998), 273–289.

[6] H. Brézis, *Operateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, Math Studies, Vol. 5, North-Holland, Amsterdam/New York, 1973.

[7] H. Brézis and I. Ekeland, Un principe variationnel associé à certaines équations paraboliques. Le cas indépendant du temps, C. R. Acad. Sci. Paris Sér. A-B, **282** (1976), A971–A974.

[8] H. Brézis and I. Ekeland, Un principe variationnel associé à certaines équations paraboliques. Le cas dépendant du temps, C. R. Acad. Sci. Paris Sér. A-B, **282** (1976), A1197–A1198.

[9] F. Browder and P. Hess, Nonlinear mappings of monotone type in Banach spaces, J. Funct. Anal., **11** (1972), 251–294.

[10] G. Dal Maso, *An introduction to* Γ*-convergence.* Progress in Nonlinear Differential Equations and their Applications, 8. Birkhäuser Boston Inc., Boston, MA, 1993.

[11] E. De Giorgi and T. Franzoni, On a type of variational convergence, In *Proceedings of the Brescia Mathematical Seminar*, 9, 63–101, Milan, 1979.

[12] E. De Giorgi, A. Marino, and M. Tosques, Problems of evolution in metric spaces and maximal decreasing curve, Atti Accad. Naz. Lincei Rend. Cl. Sci. Fis. Mat. Natur. (8), **68** (1980), 180–187.

[13] M. H. Duong, M. A. Peletier, and J. Zimmer, GENERIC formalism of a Vlasov–Fokker–Planck equation and connection to large-deviation principles, Nonlinearity, **26** (2013), 2951–2971.

[14] S. Fitzpatrick, Representing monotone operators by convex functions, Workshop/Miniconference on Functional Analysis and Optimization (Canberra, 1988), 59–65, *Proc. Centre Math. Anal. Austral. Nat. Univ.* 20, Austral. Nat. Univ., Canberra, 1988.

[15] C. M. Gariboldi and D. A. Tarzia, Convergence of distributed optimal controls in mixed elliptic problems by the penalization method, Math. Notae, **45** (2007/08), 1–19.

[16] N. Ghoussoub and R. J. McCann, A least action principle for steepest descent in a non-convex landscape, In Partial differential equations and inverse problems, vol. 362 of *Contemp. Math.*, Amer. Math. Soc., Providence, RI, 2004, pp. 177–187.

[17] N. Ghoussoub and L. Tzou, A variational principle for gradient flows, Math. Ann., **330** (2004), 519–549.

[18] N. Ghoussoub, *Selfdual partial differential systems and their variational principles*, Universitext. Springer, New-York, 2009.

[19] M. Grmela and H. C. Öttinger, Dynamics and thermodynamics of complex fluids. I. Development of a general formalism, Phys. Rev. E, **56** (1997), 6620–6632.

[20] M. D. Gunzburger and H.-C. Lee, A penalty/least-squares method for optimal control problems for first-order elliptic systems, Appl. Math. Comput., **107** (2000), 57–75.

[21] N. Kenmochi, Nonlinear evolution equations with variable domains in Hilbert spaces, Proc. Japan Acad. Ser. A Math. Sci., **53** (1977), 163–166.

[22] N. Kenmochi and T. Koyama, Nonlinear functional variational inequalities governed by time-dependent subdifferentials, Nonlinear Anal., **17** (1991), 863–883.

[23] B. Lemaire, An asymptotical variational principle associated with the steepest descent method for a convex function, J. Convex Anal. **3** (1996), 63–70.

[24] J.-L. Lions, *Contrôle optimal de systèmes gouvernés par des équations aux dérivées partielles*, Gauthier-Villars, Paris 1968.

[25] M. Mabrouk, A variational approach for a semi-linear parabolic equation with measure data, Ann. Fac. Sci. Toulouse Math., **9** (2000), 91–112.

[26] M. Mabrouk, A variational principle for a nonlinear differential equation of second order, Adv. in Appl. Math. **31** (2003), 388–419.

[27] A. Mielke, Formulation of thermoelastic dissipative material behavior using GENERIC, Contin. Mech. Thermodyn., **23** (2011), 233–256.

[28] A. Mielke, Dissipative quantum mechanics using GENERIC. Proc. of the conference on *Recent Trends in Dynamical Systems*, vol. 35 of Proceedings in Mathematics & Statistics, Springer, 2013, pp. 555–585.

[29] G. Mophou and G. M. N'Guérékata, Optimal control of a fractional diffusion equation with state constraints, Comput. Math. Appl., **62** (2011), 1413–1426.

[30] J.-J. Moreau, Evolution problem associated with a moving convex set in a Hilbert space, J. Differential Equations, **26** (1977), 347–374.

[31] B. Nayroles, Deux théorèmes de minimum pour certains systèmes dissipatifs, C. R. Acad. Sci. Paris Sér. A-B, **282** (1976), A1035–A1038.

[32] B. Nayroles, Un théorème de minimum pour certains systèmes dissipatifs. Variante hilbertienne, Travaux Sém. Anal. Convexe, **6** (1976), 22.

[33] H. Rios, Étude de la question d'existence pour certains problèmes d'évolution par minimisation d'une fonctionnelle convexe, C. R. Acad. Sci. Paris Sér. A-B, **283** (1976), A83–A86.

[34] R. Rossi and G. Savaré, Gradient flows of non convex functionals in Hilbert spaces and applications, ESAIM Control Optim. Calc. Var., **12** (2006), 564–614.

[35] T. Roubíček, Direct method for parabolic problems, Adv. Math. Sci. Appl., **10** (2000), 57–65.

[36] U. Stefanelli, The Brezis-Ekeland principle for doubly nonlinear equations, SIAM J. Control Optim., **47** (2008), 1615–1642

[37] U. Stefanelli, The discrete Brezis-Ekeland principle, J. Convex Anal., **16** (2009), 71–87.

[38] U. Stefanelli, A variational principle for hardening elasto-plasticity, SIAM J. Math. Anal., **40** (2008), 623–652.

[39] A. Visintin, Extension of the Brezis-Ekeland-Nayroles principle to monotone operators, Adv. Math. Sci. Appl., **18** (2008), 633–650.

[40] A. Visintin, On the variational representation of monotone operators, Discrete Contin. Dyn. Syst. Ser. S, **10** (2017), 909-–918.

[41] A. Visintin, Structural compactness and stability of semi-monotone flows, SIAM J. Math. Anal., **50** (2018), 2628–2663.

[42] Y. Yamada, On evolution equations generated by subdifferential operators, J. Fac. Sci. Univ. Tokyo Sect. IA Math., **23** (1976), 491–515.