SIAM J. CONTROL OPTIM.
Vol. 52, No. 2, pp. 1071–1107

STABILITY RESULTS FOR DOUBLY NONLINEAR DIFFERENTIAL INCLUSIONS BY VARIATIONAL CONVERGENCE*

THOMAS ROCHE†, RICCARDA ROSSI‡, AND ULISSE STEFANELLI§

Abstract. We present a stability result for a wide class of doubly nonlinear equations, featuring general maximal monotone operators and (possibly) nonconvex and nonsmooth energy functionals. The limit analysis consists in the reformulation of the differential evolution as a scalar energy-conservation equation with the aid of the so-called Fitzpatrick theory for the representation of monotone operators. In particular, our result applies to the vanishing viscosity approximation of rate-independent systems.

Key words. doubly nonlinear differential inclusions, maximal monotone operators, stability results, graph convergence, self-dual functional, Fitzpatrick functionals

AMS subject classifications. 35A15, 35K50, 35K85, 49Q20, 58E99

DOI. 10.1137/130909391

1. Introduction. This paper is concerned with a convergence result for doubly nonlinear differential inclusions of the type

$$\alpha_n(\dot{u}_n(t)) + \partial\mathcal{E}_t(u_n(t)) \ni 0 \quad \text{in } X^* \qquad \text{for a.a. } t \in (0,T). \tag{1.1}$$

Here, (α_n) is a sequence of maximal monotone (and possibly multivalued) operators $\alpha_n : X \rightrightarrows X^*$, $(X, \|\cdot\|)$ is a (separable) reflexive Banach space, and $\mathcal{E} : [0,T] \times X \to (-\infty, \infty]$ is a (proper) time-dependent *energy* functional. We will prove that for $\alpha_n \to \alpha$ in the graph sense any limit point u of the sequence $(u_n)_{n\in\mathbb{N}}$ is a solution to

$$\alpha(\dot{u}(t)) + \partial\mathcal{E}_t(u(t)) \ni 0 \quad \text{in } X^* \qquad \text{for a.a. } t \in (0,T).$$

Throughout the paper, we write $\mathcal{E}_t(u)$ in place of $\mathcal{E}(t,u)$. We will understand the multivalued operator $\partial\mathcal{E} : (0,T) \times X \rightrightarrows X^*$ to be a suitable notion of subdifferential for the possibly nonsmooth and nonconvex map $u \mapsto \mathcal{E}_t(u)$, namely the so-called *Fréchet subdifferential*, defined at $(t,u) \in \mathrm{dom}(\mathcal{E})$ by

$$\xi \in \partial\mathcal{E}_t(u) \quad \text{ if and only if } \mathcal{E}_t(v) \geq \mathcal{E}_t(u) + \langle \xi, v-u\rangle + \mathrm{o}(\|v-u\|) \quad \text{as } v \to u. \tag{1.2}$$

Observe that, as soon as the mapping $u \mapsto \mathcal{E}_t(u)$ is *convex*, the Fréchet subdifferential $\partial\mathcal{E}_t(u)$ coincides with the subdifferential of $u \mapsto \mathcal{E}_t(u)$ in the sense of convex analysis.

Doubly nonlinear equations such as (1.1) arise in a variety of different applications, ranging from thermomechanics, to phase change, to magnetism. As such, they have

*Received by the editors February 13, 2013; accepted for publication (in revised form) December 17, 2013; published electronically March 25, 2014.
http://www.siam.org/journals/sicon/52-2/90939.html

†Department of Mathematics M6, Technische Universität München, 85748 Garching b. München, Germany (roche@ma.tum.de). This author was supported by the Bavarian Network of Excellence through its graduate program TopMath and the TUM Graduate School through its Thematic Graduate Center TopMath.

‡Dipartimento di Matematica, Università di Brescia, 25133 Brescia, Italy (riccarda.rossi@ing.unibs.it).

§Istituto di Matematica Applicata e Tecnologie Informatiche *E. Magenes* - CNR, 27100 Pavia, Italy, and Faculty of Mathematics, University of Vienna, 1090 Vienna, Austria (ulisse.stefanelli@univie.ac.at). This author was partially supported by the FP7-IDEAS-ERC-StG grant # 200947 *BioSMA* and the CNR-JSPS grant *VarEvol*.

attracted substantial attention in recent years. Correspondingly, the related literature is quite rich. As a comprehensive review is beyond the scope of this paper, we limit ourselves to recording the seminal observations by Moreau [57, 58] and Germain [37], as well as the early existence results by Arai [9], Senba [72], Colli and Visintin [26], and Colli [25]. The reader can find a selection of more recent results in [5, 2, 3, 7, 33, 41, 51, 66, 70]. Without going into details, let us mention that, over the last decade, the convexity requirement on the map $u \mapsto \mathcal{E}_t(u)$ in the pioneering papers [9, 72, 26, 25] has been progressively weakened: in particular, in [51] a quite broad class of nonsmooth and nonconvex energy functionals has been considered. Nonetheless, in all of the aforementioned contributions the operator α is assumed to fulfill some coercivity property, namely to have at least linear growth at infinity. We will refer to this case as *viscous*.

The case of 0-homogeneous operators α has been recently investigated as well, for it connects with the modeling of so-called rate-independent systems. We shall hence refer to this situation as *rate-independent*. Some references in this direction are to be found in the papers [28, 29, 36, 43, 48, 47, 53, 55, 56].

Additionally, relation (1.1) has been considered in connection with the study of the long-time behavior of solutions in [4, 31, 69, 71, 70] and their variational characterization in [6, 8, 73, 76].

The focus of this paper is on the study of the stability of the doubly nonlinear flows (1.1). Namely, we investigate the convergence of solutions u_n to (1.1), under the assumption

$$\alpha_n \to \alpha \text{ in the graph sense in } X \times X^*, \tag{1.3}$$

viz. for all $(\xi, \xi^*) \in \operatorname{graph}(\alpha)$ there exists $(\xi_n, \xi_n^*) \in \operatorname{graph}(\alpha_n)$ such that $\xi_n \to \xi$ in X and $\xi_n^* \to \xi^*$ in X^* as $n \to \infty$. The main result of this paper, Theorem 4.5, states that cluster points u of the curves (u_n) are in fact solutions to the limiting equation

$$\alpha\left(\dot{u}(t)\right) + \partial \mathcal{E}_t(u(t)) \ni 0 \quad \text{in } X^* \qquad \text{for a.a. } t \in (0, T). \tag{1.4}$$

We have to mention that stability results for the doubly nonlinear flows (1.1) are already available in the literature. For *viscous* graphs α_n, a first convergence theorem in the case of convex energies has been obtained by Aizicovici and Yan [1] (see also [73]), whereas stability results for doubly nonlinear equations with nonconvex energies have been proved in [51]. This issue has been recently reconsidered by Visintin [80, 81, 82, 83], who has remarkably extended the theory to treat subdifferential inclusions of the type

$$\beta(\dot{u}(t)) + \gamma(u(t)) \ni 0 \quad \text{in } X^* \qquad \text{for a.a. } t \in (0, T),$$

with $\beta, \gamma : X \rightrightarrows X^*$ maximal monotone operators, β cyclically monotone, and γ noncyclic monotone, by resorting to the so-called Fitzpatrick theory [34].

Let us briefly recall that an operator $\alpha : X \rightrightarrows X^*$ is *cyclically monotone* if α is the *generalized gradient* of some potential. Namely, if $\alpha = \partial\psi$ for some proper, convex, and lower semicontinuous function $\psi : X \to (-\infty, \infty]$, where the symbol ∂ denotes the subdifferential in the sense of convex analysis. In the cyclic-monotone case $\alpha = \partial\psi$, it is well known that the relation $y \in \partial\psi(x)$ can be equivalently reformulated as $\langle y, x \rangle = \psi(x) + \psi^*(y)$, where ψ^* is the Legendre–Fenchel conjugate of ψ and $\langle \cdot, \cdot \rangle$ is the duality pairing between X^* and X. The use of this variational fact for the aim of variationally reformulating evolution equations dates back to Brezis and Ekeland

[18, 17] and Nayroles [60, 61]. Among the many contributions stemming from this idea, the reader is especially referred to the existence proofs by Auchmuty [11] and Roubíček [68] and to the recent monograph by Ghoussoub [39] on self-dual variational principles (see also the references in [73]).

The Fitzpatrick theory allows us to extend this variational view to subdifferential inclusions of the type (1.4), with α possibly *noncyclic* monotone, by introducing *representative* functions $f_\alpha : X \times X^* \to (-\infty, \infty]$ for the operator α. These are convex functions f_α with the property

$$\forall (x,y) \in X \times X^*, \ \langle y, x \rangle \le f_\alpha(x,y), \ \text{and} \\ y \in \alpha(x) \ \text{iff} \ \langle y, x \rangle = f_\alpha(x,y). \tag{1.5}$$

The reader is referred to section 2 below for a selection of relevant results within this theory. In particular, in [81] these tools are used in order to variationally reformulate relations (1.1) for noncyclic monotone operators. This reformulation opens the way to devising a suitable Γ-convergence analysis toward structural stability of the flows.

As for the *rate-independent* case, we mention the stability results for hysteresis operators from the classical monographs [19, 42, 79] (see also [75]) as well as the general approximation theory for energetic solutions from [54]. Another stability result in the rate-independent setting is in [73]. Moreover, we mention Visintin [82, 83], which exploits the Fitzpatrick idea in the rate-independent context but by taking perturbations in $\partial \mathcal{E}_n$ (again, in a possibly noncyclic monotone framework).

Finally, the approximation of rate-independent flows by viscous flows (in the cyclic-monotone case) has recently attracted a great deal of attention. This is especially critical as viscous and rate-independent evolutions usually call for different analytical treatments. The *vanishing viscosity* approach to abstract rate-independent systems has been developed in particular in [32, 49, 50]. More specifically, in the latter two papers it was shown that the vanishing viscosity limit leads to the notion of BV solution to a rate-independent system. In the recent [52], still within the cyclic-monotone framework, the $p_n \to 1$ limit, where p_n is the homogeneity of the potential ψ_n of α_n, was addressed, and it was proved that BV solutions arise in the limit. A stability result with respect to variational convergence for the latter solution concept was also obtained.

Our result. The focus here is that of obtaining a stability result for the differential inclusions (1.1) by allowing for maximal generality on the perturbations α_n and on the functional $\mathcal{E}$. In particular, we shall assume neither superlinear equicoercivity in α_n nor cyclic monotonicity. As for the energy $\mathcal{E}$, we do not require smoothness or convexity with respect to u, but we still ask for lower semicontinuity and some coercivity; see Assumption 3.9 below.

This generality sets our result aside from the available contributions on this topic. In particular, our analysis also encompasses the passage from viscous to rate-independent doubly nonlinear evolution. Indeed, we are able to treat here the $p_n \to 1$ case for noncyclic monotone operators α_n (in this setting, p_n is the coercivity exponent for α_n). In our general context, we prove that the so-called *local solutions* [49, 50] to a rate-independent system arise in the $p_n \to 1$ limit.

The basic idea for handling the noncyclic monotone case is to resort to a variational reformulation of the flows (1.1) which is well suited for discussing limits. By letting f_{α_n} *represent* the monotone operator α_n in the sense of (1.5) and assuming the validity of a suitable chain rule for the energy $\mathcal{E}$, relation (1.1) is proved to be equivalent (see Proposition 3.12) to an *energy conservation* identity, namely

$$\underbrace{\mathcal{E}_t(u_n(t))}_{\text{energy at } t} + \underbrace{\int_0^t f_{\alpha_n}\Big(\dot{u}_n(s), -\partial\mathcal{E}_s(u_n(s))\Big)dt}_{\text{dissipated energy on } [0,t]} = \underbrace{\mathcal{E}_0(u_n(0))}_{\text{initial energy}} + \underbrace{\int_0^t \partial_t\mathcal{E}_s(u_n(s))ds}_{\text{work of ext. actions}} \quad \forall t \in [0,T]. \tag{1.6}$$

The strategy is then to prove that, by passing to the lim inf in (1.6), the structure of the relation is preserved. In particular, we provide sufficient conditions under which the lim inf of the integral of the representative functions f_{α_n} is a representative function of the limit graph α. Care here is given to developing such a lower semicontinuity argument for functions which are only BV in time. This allows us to directly include the case of rate-independent flows.

Let us once more emphasize that we can encompass in our analysis a broad class of time-dependent energies $\mathcal{E} : [0,T] \times X \to (-\infty, +\infty]$, (possibly) nonsmooth and nonconvex with respect to the state variable u, but still satisfying a suitable set of coercivity and regularity type conditions mutuated from [51]. Note that in such a general setting existence of absolutely continuous solutions $u_n : [0,T] \to X$ to (1.1) is presently not known. A possibility in order to overcome this would be to strengthen our assumptions on the energy functional $\mathcal{E}$, for instance, assuming it to be a suitable perturbation of a (λ-)convex functional, as in [67, 66, 51]. However, we refrain from this for the sake of keeping maximal generality for the convergence result. In this respect, our result should be regarded purely as a stability analysis, with a focus on the convergence properties of the operators (α_n). A stability result with respect to suitable convergence of the energy functionals could also be obtained, closely following the lines of [51, Thm. 4.8]. Again, we have chosen not to detail this in order to highlight the usage of the Fitzpatrick theory to deal with the *noncyclic* operators (α_n). This very generality will allow us, for instance, to address in section 4.3 the quasistatic limits of a class of *dynamical* problems, which in fact can be reformulated as doubly nonlinear equations of the form (1.1).

Structure of the paper. Section 2 contains some background material on the Fitzpatrick theory and on the notions of variational convergence for functionals and operators which will be relevant for the subsequent analysis. In section 3.1, some further preliminaries of measure theory and convex analysis are provided, whereas in section 3.2 the basic assumptions on the energy functional $\mathcal{E}$ are stated in detail, and suitable reformulations of (1.1) are discussed. In section 4 we state our main stability result Theorem 4.5 and thoroughly discuss it. We also give two corollaries (i.e., Theorem 4.8 and Proposition 4.9) in two particular cases: specifically, Proposition 4.9 deals with the $p_n \to 1$ vanishing-viscosity limit. We conclude section 4 by discussing classes of energy functionals to which our results apply (cf. section 4.2) and developing applications to rate-independent limits of Hamiltonian systems (in section 4.3). The proof of Theorem 4.5 is developed throughout section 5, also exploiting some results from Young measure theory which are contained in Appendix A.

2. Fitzpatrick theory.

Within this section, we shall systematically use the notation

$$\pi(\xi, \xi^*) := \langle \xi^*, \xi \rangle \quad \forall\, (\xi, \xi^*) \in X \times X^*$$

for the duality pairing between the reflexive space X and X^* and identify possibly multivalued operators $\alpha : X \rightrightarrows X^*$ with the corresponding graphs $\alpha \subset X \times X^*$ without changing notation. We recall that $\alpha : X \rightrightarrows X^*$ is *monotone* if

$$\langle \xi^* - \xi_0^*, \xi - \xi_0 \rangle \geq 0 \quad \forall\, \xi^* \in \alpha(\xi),\ \xi_0^* \in \alpha(\xi_0)$$

(where $\langle \cdot, \cdot \rangle$ denotes the duality pairing between X^* and X) and that it is *maximal monotone* if it is maximal for set inclusion within the class of monotone operators.

We shall provide here a minimal aside on the Fitzpatrick theory, essentially mutuated from [81]. The reader is referred, however, to [20, 21, 22, 23, 38, 39, 45, 46, 63, 64] for additional material and a collection of related results and applications to PDEs.

Representative functionals. We denote by $\mathcal{F}(X)$ the set of functionals $\varphi : X \times X^* \to (-\infty, \infty]$ such that

$$\varphi \ \text{ is convex and lower semicontinuous, and } \ \varphi(\xi, \xi^*) \geq \pi(\xi, \xi^*) \quad \forall (\xi, \xi^*) \in X \times X^*.$$

We associate with $\varphi \in \mathcal{F}(X)$ the set $\alpha \subset X \times X^*$ given by

$$(\xi, \xi^*) \in \alpha \quad \Leftrightarrow \quad \varphi(\xi, \xi^*) = \pi(\xi^*, \xi). \tag{2.1}$$

Whenever (2.1) holds we say that φ *represents* α, that φ is *representative*, and that α is *representable*. A representable operator can be represented by different representative functionals (cf. Example 2.4 below). On the contrary, each representative functional represents only one operator.

Example 2.1. In the cyclically monotone case of $\alpha = \partial\psi$ for some (proper) convex and lower semicontinuous potential $\psi : X \to (-\infty, \infty]$, a representative functional for α is given by the *bipotential* (according to the terminology of [20])

$$\varphi(\xi, \xi^*) = \psi(\xi) + \psi^*(\xi^*). \tag{2.2}$$

We have the following *strict* set inclusions:

$$\{\text{maximal monotone operators}\} \subsetneqq \{\text{representable operators}\} \subsetneqq \{\text{monotone operators}\}.$$

Namely, representable operators are intermediate between monotone and maximal monotone. One may wonder how to translate maximality at the level of representative functionals. The following result provides a useful criterion for the representability of a maximal monotone operator.

Proposition 2.2 (representative of a maximal monotone operator [77]). *A functional $\varphi \in \mathcal{F}(X)$ represents a maximal monotone operator iff $\varphi^* \in \mathcal{F}(X^*)$. In this case, if φ represents α, then φ^* represents $\alpha^{-1} = \{(\xi, \xi^*) : (\xi^*, \xi) \in \alpha\}$.*

The Fitzpatrick and the Burachik–Svaiter functions. Given $\alpha \subset X \times X^*$ with $\alpha \neq \emptyset$, we define the *Fitzpatrick function* (associated with α) $f_\alpha : X \times X^* \to (-\infty, \infty]$ by

$$\begin{aligned} f_\alpha(\xi, \xi^*) &:= \pi(\xi, \xi^*) + \sup\{\pi(\xi_0 - \xi, \xi^* - \xi_0^*) \ : \ (\xi_0, \xi_0^*) \in \alpha\} \\ &= \sup\{\pi(\xi_0, \xi^*) - \pi(\xi_0 - \xi, \xi_0^*) \ : \ (\xi_0, \xi_0^*) \in \alpha\} \quad \forall\, (\xi, \xi^*) \in X \times X^* \end{aligned} \tag{2.3}$$

and the *Burachik–Svaiter function* [22] (associated with α) $\rho_\alpha : X \times X^* \to (-\infty, \infty]$ by

$$\rho_\alpha := (\pi + I_\alpha)^{**}, \tag{2.4}$$

where I_α stands for the indicator function of the graph α, namely $I_\alpha(\xi, \xi^*) = 0$ if $\xi^* \in \alpha(\xi)$ and $I_\alpha = \infty$ elsewhere. Both f_α and ρ_α represent α. Moreover, f_α and

ρ_α are, respectively, the minimal and maximal elements (with respect to pointwise ordering) of the functional *interval*

$$\mathcal{I}(\alpha) = \{\varphi \in \mathcal{F}(X) \ : \ \varphi \text{ represents } \alpha\}. \tag{2.5}$$

In particular, in the cyclically monotone case of $\alpha = \partial\psi$, there holds

$$f_\alpha(\xi, \xi^*) \leq \psi(\xi) + \psi^*(\xi^*) \leq \rho_\alpha(\xi, \xi^*) \quad \forall\, (\xi, \xi^*) \in X \times X^*. \tag{2.6}$$

Let us also point out that, in view of the definitions (2.3) and (2.4) of f_α and ρ_α, the following formulae hold:

$$f_\alpha(\xi, \xi^*) = \sup\{\pi(\xi_0, \xi^*) + \pi(\xi, \xi_0^*) - \rho_{\alpha^{-1}}(\xi_0^*, \xi_0) \ : \ (\xi_0, \xi_0^*) \in X{\times}X^*\} \quad \forall\, (\xi, \xi^*) \in X{\times}X^*, \tag{2.7}$$

$$\rho_\alpha(\xi, \xi^*) = \sup\{\pi(\xi_0, \xi^*) + \pi(\xi, \xi_0^*) - f_{\alpha^{-1}}(\xi_0^*, \xi_0) \ : \ (\xi_0, \xi_0^*) \in X{\times}X^*\} \quad \forall\, (\xi, \xi^*) \in X{\times}X^*, \tag{2.8}$$

namely f_α (ρ_α, resp.) is the convex conjugate of the Burachik–Svaiter (Fitzpatrick, resp.) function of the inverse operator α^{-1}.

Finally, for later use we observe that

$$0 \in \alpha(0) \ \Rightarrow \ f_\alpha(\xi, \xi^*) \geq 0 \quad \forall\, (\xi, \xi^*) \in X \times X^*. \tag{2.9}$$

Self-dual representatives. Recall that a function $\varphi : X \times X^* \to (-\infty, \infty]$ is called *self-dual* iff

$$\varphi(\xi, \xi^*) = \varphi^*(\xi^*, \xi) \quad \forall\, (\xi, \xi^*) \in X \times X^*.$$

The interval $\mathcal{I}(\alpha)$ from (2.5) includes *self-dual* representative functions [64, Thm. 3.3]. In the cyclically monotone case of $\alpha = \partial\psi$, an example in this direction is given by the *bipotential* (2.2). Out of the cyclically monotone realm, an example of a self-dual representative in the case $\alpha = \gamma + \partial\psi$ with γ skew adjoint is $(\xi, \xi^*) \mapsto \psi(\xi) + \psi^*(-\gamma\xi + \xi^*)$ [38].

In the general case, the indirect proof of the existence of self-dual representative functions is due to Penot [62, 63] and Svaiter [77], whereas direct constructions were first provided by Penot and Zălinescu [65] under some restriction on α. An explicit self-dual representative function in the general maximal monotone case was recently obtained by Bauschke and Wang [15] and reads

$$(\xi, \xi^*) \mapsto \frac{1}{2} \inf_{(\zeta, \zeta^*) \in X \times X^*} \left\{ f_\alpha(\xi + \zeta, \xi^* + \zeta^*) + f_\alpha(\xi - \zeta, \xi^* - \zeta^*) + \|\zeta\|^2 + \|\zeta^*\|_*^2 \right\}. \tag{2.10}$$

Note that neither the Fitzpatrick function f_α nor the Burachik–Svaiter function ρ_α is self-dual in general.

Let us now recast the characterization of maximal monotonicity of Proposition 2.2 in the following.

PROPOSITION 2.3 (self-dual representatives = maximality). *An operator $\alpha : X \rightrightarrows X^*$ is maximal monotone iff it is represented by a self-dual functional φ.*

Proof. By [15], if α is maximal monotone, then it admits the self-dual representative (2.10).

As for the converse implication, note that by self-duality of φ and π we get

$$\varphi^*(\xi^*, \xi) = \varphi(\xi, \xi^*) \geq \pi(\xi, \xi^*) = \pi^*(\xi^*, \xi).$$

Thus, $\varphi^* \in \mathcal{F}(X^*)$ and Proposition 2.2 applies. $\square$

Self-dual representatives versus Fitzpatrick and Burachik–Svaiter functions. Let $\alpha \subset X \times X^*$ be a cyclically monotone operator with $\alpha = \partial\psi$ for some convex and lower semicontinuous potential $\psi : X \to (-\infty, \infty]$. As already observed, a self-dual representative of α is the sum of ψ and its convex conjugate. However, in general, the Fitzpatrick functional f_α may differ from $\psi + \psi^*$, as shown by the following.

Example 2.4. Consider $X = \mathbb{R} = X^*$, and set $\alpha =$ identity, namely $\alpha = \partial\psi$ with $\psi(\xi) = \frac{1}{2}\xi^2$. The Fitzpatrick function of $\partial\psi$ is $f_{\partial\psi}(\xi, \xi^*) = \xi^2/4 + (\xi^*)^2/4 + \xi \cdot \xi^*/2$, which is not self-dual.

Fitzpatrick and Burachik–Svaiter functions in the case of 1-positively homogeneous potentials. Our next result reveals that, when $\alpha = \partial\psi$ and $\psi : X \to (-\infty, +\infty]$ is positively homogeneous of degree 1, then also the Fitzpatrick functional f_α coincides with the bipotential (2.2).

PROPOSITION 2.5. *Let $\psi : X \to (-\infty, \infty]$ be convex, lower semicontinuous and positively homogeneous of degree* 1, *viz. $\psi(\lambda\xi) = \lambda\psi(\xi)$ for all $\xi \in X$ and $\lambda \geq 0$.*

Then, the Fitzpatrick function of the subdifferential of ψ coincides with the sum of ψ and its convex conjugate, i.e.,

$$f_{\partial\psi}(\xi, \xi^*) = \psi(\xi) + \psi^*(\xi^*) \qquad \forall\, (\xi, \xi^*) \in X \times X^*. \tag{2.11}$$

Before developing the *proof*, we recall that, for all ψ convex, lower semicontinuous, and 1-homogeneous, there exists a closed convex set $0 \in K \subset X$ such that ψ coincides with the *Minkowski functional* of K, viz.

$$\psi(\xi) = M_K(\xi) = \inf\left\{\sigma > 0 \,:\, \frac{\xi}{\sigma} \in K\right\}. \tag{2.12a}$$

Furthermore,

$$\begin{aligned} &\psi^*(\xi^*) = I_{K^*}(\xi^*), \text{ where } K^* \subset X^* \text{ is the } \textit{polar} \text{ set of } K, \text{ i.e.,} \\ &\qquad K^* = \{\xi^* \in X^* \,:\, \pi(\xi, \xi^*) \leq 1 \;\forall\, \xi \in K\}, \end{aligned} \tag{2.12b}$$

so that

$$\forall\, (\xi, \xi^*) \in \partial\psi \qquad M_K(\xi) + I_{K^*}(\xi^*) = \pi(\xi, \xi^*)\,. \tag{2.12c}$$

For later convenience, we also recall that

$$\psi(\xi) = \sup_{\xi^* \in K^*} \pi(\xi, \xi^*) \quad \forall\, \xi \in X\,. \tag{2.12d}$$

Proof of Proposition 2.5. It follows from the definition of the Fitzpatrick function (2.3) and from (2.12c) that

$$f_{\partial\psi}(\xi, \xi^*) = \sup\{\underbrace{\pi(\xi_0, \xi^*) - M_K(\xi_0)}_{\leq I_{K^*}(\xi^*)} + \underbrace{\pi(\xi, \xi_0^*) - I_{K^*}(\xi_0^*)}_{\leq M_K(\xi)} :\; (\xi_0, \xi_0^*) \in \partial\psi\}.$$

Hence, we obtain that $f_{\partial\psi}(\xi, \xi^*) \leq M_K(\xi) + I_{K^*}(\xi^*)$ for all $(\xi, \xi^*) \in X \times X^*$.

For the opposite inequality assume first that $\xi^* \notin K^*$. Then there exists $\xi_0 \in K$ such that $\pi(\xi_0, \xi^*) > 1$. Choose an arbitrary $\xi_0^* \in \partial M_K(\xi_0) = \partial M_K(\lambda\xi_0)$ for any positive $\lambda > 0$. Then, taking into account that $M_K(\lambda\xi_0) + I_{K^*}(\xi_0^*) = \lambda\pi(\xi_0, \xi_0^*) \leq \lambda$, we get

$$\begin{aligned} &\pi(\lambda\xi_0, \xi^*) - M_K(\lambda\xi_0) + \pi(\xi, \xi_0^*) - I_{K^*}(\xi_0^*) \\ &\geq \lambda\,(\pi(\xi_0, \xi^*) - 1) + \pi(\xi, \xi_0^*) \to +\infty \qquad \text{as } \lambda \to \infty. \end{aligned}$$

Therefore, $f_{\partial\psi}(\xi,\xi^*) \geq I_{K^*}(\xi^*)$. On the other hand, taking into account that $\partial M_K(0) = K^*$, we deduce that

$$f_{\partial\psi}(\xi,\xi^*) \geq \sup\{\pi(\xi,\xi_0^*) : \xi_0^* \in K^*\} = M_K(\xi)\,.$$

Eventually, we get that $f_{\partial\psi}(\xi,\xi^*) \geq M_K(\xi) + I_{K^*}(\xi^*)$ for all $(\xi,\xi^*) \in X \times X^*$, which concludes the proof. □

COROLLARY 2.6. *Let $\psi : X \to (-\infty,\infty]$ be convex, lower semicontinuous, and positively homogeneous of degree* 1*. Then*

$$f_{\partial\psi}(\xi,\xi^*) = \psi(\xi) + \psi^*(\xi^*) = \rho_{\partial\psi^*}(\xi,\xi^*) \qquad \forall\,(\xi,\xi^*) \in X \times X^*. \tag{2.13}$$

Proof. To prove (2.13) we observe that

$$f_{\partial\psi} = \psi + \psi^* = (\psi+\psi^*)^* = f^*_{\partial\psi} = \rho_{\partial\psi^*},$$

where the first identity is due to Proposition 2.5, the second to the fact that $\psi + \psi^*$ is self-dual, and the last to (2.8). □

2.1. Approximation of maximal monotone operators. In the following lines, the symbol X may stand for the space X, for X^*, or for $X \times X^*$. Let f_n, $\mathsf{f} : \mathsf{X} \to (-\infty,\infty]$ be convex, proper, and lower semicontinuous functionals, and let α_n, $\alpha \subset X \times X^*$ be maximal monotone operators. We introduce the notation

$$\begin{aligned}\Gamma\text{–}\liminf_{n\to\infty} \mathsf{f}_n(\mathsf{x}) &:= \min\{\liminf_{n\to\infty} \mathsf{f}_n(\mathsf{x}_n),\ \mathsf{x}_n \rightharpoonup \mathsf{x} \text{ in } \mathsf{X}\},\\ \Gamma\text{–}\limsup_{n\to\infty} \mathsf{f}_n(\mathsf{x}) &:= \min\{\limsup_{n\to\infty} \mathsf{f}_n(\mathsf{x}_n),\ \mathsf{x}_n \to \mathsf{x} \text{ in } \mathsf{X}\}.\end{aligned}$$

These correspond to classical *sequential* Γ–liminf and Γ–limsup constructions (cf., e.g., [27]). Note that Γ–liminf is defined in relation with the weak topology of X, whereas Γ–limsup refers to the strong topology instead. We will use the following convergence notions, for which the reader is referred to [10]:

$$\begin{aligned}\mathsf{f}_n \xrightarrow{\text{M}} \mathsf{f} \quad &\Leftrightarrow \quad \Gamma\text{–}\limsup_{n\to\infty} \mathsf{f}_n \leq \mathsf{f} \leq \Gamma\text{–}\liminf_{n\to\infty} \mathsf{f}_n,\\ \alpha_n \xrightarrow{\text{g}} \alpha \quad &\Leftrightarrow \quad \forall(\xi,\xi^*) \in \alpha\ \ \exists(\xi_n,\xi_n^*) \in \alpha_n\ :\ \xi_n \to \xi,\ \xi_n^* \to \xi^*,\end{aligned}$$

where the symbol $\xrightarrow{\text{M}}$ stands for *Mosco convergence* in X and $\xrightarrow{\text{g}}$ is usually referred to as *graph convergence*. In particular, Mosco convergence corresponds to Γ-convergence with respect to both the strong and the weak topologies of X and can be made more explicit by

$$\mathsf{f}_n \xrightarrow{\text{M}} \mathsf{f} \quad \Leftrightarrow \quad \begin{cases} \forall \mathsf{x}_n \rightharpoonup \mathsf{x}, \quad \mathsf{f}(\mathsf{x}) \leq \liminf_{n\to\infty} \mathsf{f}_n(\mathsf{x}_n),\\ \forall \mathsf{x} \in \mathsf{X},\ \exists \mathsf{x}_n \to \mathsf{x}\ :\ \mathsf{f}_n(\mathsf{x}_n) \to \mathsf{f}(\mathsf{x}).\end{cases}$$

In the case of cyclically monotone operators, graph convergence is known to be equivalent to the Mosco convergence of the respective potentials (up to some normalization condition), viz. we have the following result.

THEOREM 2.7 (see [10, Thm. 3.66, p. 373]). *Let $\phi_n,\ \phi : X \to (-\infty,+\infty]$ be proper, convex, and lower semicontinuous functionals. The following are equivalent:*

(i) *$\partial\phi_n \xrightarrow{\text{g}} \partial\phi$ in $X \times X^*$, and there exist $(\xi_n,\xi_n^*) \in \partial\phi_n$ and $(\xi,\xi^*) \in \partial\phi$ such that $\xi_n \to \xi$ in X, $\xi_n^* \to \xi^*$ in X^*, and $\phi_n(\xi_n) \to \phi(\xi)$;*

(ii) $\phi_n \xrightarrow{\text{M}} \phi$ *in* X.

The importance of graph convergence is revealed by the following identification lemma, basically consisting of the *approximation version* of [16, Prop. 2.5, p. 27].

LEMMA 2.8. *Let* $\alpha_n \xrightarrow{\text{g}} \alpha$, $(\xi_n, \xi_n^*) \in \alpha_n$, $\xi_n \rightharpoonup \xi$, $\xi_n^* \rightharpoonup \xi$, *and* $\liminf_{n\to\infty} \pi(\xi_n, \xi_n^*) \leq \pi(\xi, \xi^*)$. *Then* $(\xi, \xi^*) \in \alpha$.

In order to prove this quite classical approximation lemma in the particular case of cyclically monotone operators, what is actually needed is just the implication (i) $\Rightarrow$ (ii) in Theorem 2.7 above. Indeed, let $\partial\phi_n \xrightarrow{\text{g}} \partial\phi$, $(\xi_n, \xi_n^*) \in \partial\phi_n$, with $\xi_n \rightharpoonup \xi$, $\xi_n^* \rightharpoonup \xi^*$, and $\liminf_{n\to\infty} \pi(\xi_n, \xi_n^*) \leq \pi(\xi, \xi^*)$ be given. We readily have that

$$0 \leq \phi(\xi) + \phi^*(\xi^*) - \pi(\xi, \xi^*) \leq \liminf_{n\to\infty} \big(\phi_n(\xi_n) + \phi_n^*(\xi_n^*) - \pi(\xi_n, \xi_n^*)\big) = 0,$$

where we have used the fact that $\phi_n^* \xrightarrow{\text{M}} \phi^*$ iff $\phi_n \xrightarrow{\text{M}} \phi$ [10, Thm. 3.18, p. 295], i.e., the bicontinuity of the Legendre–Fenchel transformation with respect to the topology induced by the Mosco convergence. Therefore, we conclude that $(\xi, \xi^*) \in \alpha$.

Remark 2.9. As indeed one just needs

$$\partial\phi_n \xrightarrow{\text{g}} \partial\phi \quad \Rightarrow \quad \phi \leq \Gamma\text{–}\liminf_{n\to\infty} \phi_n \ \text{ in } \ V \ \text{ and } \ \phi^* \leq \Gamma\text{–}\liminf_{n\to\infty} \phi_n^* \ \text{ in } \ V^*$$

in order to check for Lemma 2.8, one may wonder if directly asking the two Γ–liminf conditions above would weaken the convergence requirements on the functionals. This, however, is not the case. Indeed, under some very general condition of *equiproperness* type, we have that the two separate Γ–liminf conditions are indeed equivalent to $\phi_n \xrightarrow{\text{M}} \phi$ [73, Lemma 4.1] and hence entail $\partial\phi_n \xrightarrow{\text{g}} \partial\phi$.

Our next aim is that of extending the above arguments to the case of noncyclically maximal monotone operators. In particular, we present an extension of Theorem 2.7 in terms of representative functions, and in particular of the Fitzpatrick and Burachik–Svaiter functions.

THEOREM 2.10. *Let* (α_n), α *be maximal monotone operators* $\alpha_n, \alpha : X \rightrightarrows X^*$. *The following are equivalent:*

(i) $\alpha_n \xrightarrow{\text{g}} \alpha$ *in* $X \times X^*$,

(ii) $f_\alpha \leq \Gamma\text{–}\liminf_{n\to\infty} f_{\alpha_n}$ *in* $X \times X^*$,

(iii) $\Gamma\text{–}\limsup_{n\to\infty} \rho_{\alpha_n} \leq \rho_\alpha$ *in* $X \times X^*$.

Exactly as above, the proof of Lemma 2.8 follows just from implication (i) $\Rightarrow$ (ii). We shall, however, give a full equivalence proof for the sake of completeness and comparison with Theorem 2.7. In particular, note that condition (ii) above is weaker than $f_{\alpha_n} \xrightarrow{\text{M}} f_\alpha$. That is to say that the former Theorem 2.7 does not follow directly from Theorem 2.10.

Proof. Claim 1: (i) $\Rightarrow$ (ii). Fix $(\xi_0, \xi_0^*) \in \alpha$, and let $(\xi_{0n}, \xi_{0n}^*) \in \alpha_n$ be such that $\xi_{0n} \to \xi_0$ in X and $\xi_{0n}^* \to \xi_0^*$ in X^*. Moreover, let $\xi_n \rightharpoonup \xi$ in X and $\xi_n^* \rightharpoonup \xi^*$ in X^*. We have that

$$\begin{aligned}\liminf_{n\to\infty} f_{\alpha_n}(\xi_n, \xi_n^*) &\geq \liminf_{n\to\infty} \left(\pi(\xi_{0n}, \xi_n^*) - \pi(\xi_{0n} - \xi_n, \xi_{0n}^*)\right) \\ &= \pi(\xi_0, \xi^*) - \pi(\xi_0 - \xi, \xi_0^*).\end{aligned}$$

In particular, by passing to the supremum with respect to $(\xi_0, \xi_0^*) \in \alpha$, we conclude that $f_\alpha \leq \Gamma\text{–}\liminf_{n\to\infty} f_{\alpha_n}$.

Claim 2: (i) $\Rightarrow$ (iii). Observe that $\alpha_n \xrightarrow{\text{g}} \alpha$ implies $\alpha_n^{-1} \xrightarrow{\text{g}} \alpha^{-1}$; hence Claim 1 yields $f_{\alpha^{-1}} \leq \Gamma\text{–}\liminf_{n\to\infty} f_{\alpha_n^{-1}}$.By convex conjugation and taking into account (2.8) and [10, Thm. III.3.7, p. 271], we then have

$$\rho_\alpha = (f_{\alpha^{-1}})^* \geq \left(\Gamma\text{–}\liminf_{n\to\infty} f_{\alpha_n^{-1}}\right)^* = \Gamma\text{–}\limsup_{n\to\infty} \rho_{\alpha_n}.$$

Claim 3: (iii) $\Rightarrow$ (i). Fix $(\xi,\xi^*) \in \alpha$, and let (ξ_n,ξ_n^*) fulfill $\xi_n \to \xi$ in X, $\xi_n^* \to \xi^*$ in X^*, and

$$\limsup_{n\to\infty} \rho_{\alpha_n}(\xi_n,\xi_n^*) \leq \rho_\alpha(\xi,\xi^*) = \pi(\xi,\xi^*)$$

(such sequences exist as $\Gamma\text{–}\limsup_{n\to\infty} \rho_{\alpha_n} \leq \rho_\alpha$). In particular, we have that

$$\rho_{\alpha_n}(\xi_n,\xi_n^*) < \pi(\xi,\xi^*) + \varepsilon_n$$

for some sequence $\varepsilon_n \to 0$. By exploiting the extension of the *Brønsted–Rockafellar* approximation lemma from [44, Thm. 3.4], we have that there exist $(\tilde{\xi}_n,\tilde{\xi}_n^*) \in \alpha_n$ such that for all $n \in \mathbb{N}$

$$\|\xi_n - \tilde{\xi}_n\|^2 \leq \varepsilon_n, \quad \|\xi_n^* - \tilde{\xi}_n^*\|_*^2 \leq \varepsilon_n.$$

Then, a classical diagonal-extraction argument yields $\tilde{\xi}_n \to \xi$ in X and $\tilde{\xi}_n^* \to \xi^*$ in X^*.

Claim 4: (ii) $\Rightarrow$ (i). Again, by convex conjugation and (2.8), we deduce from (ii) that

$$\rho_{\alpha^{-1}} \geq \Gamma\text{–}\limsup_{n\to\infty} \rho_{\alpha_n^{-1}}.$$

Therefore, in view of Claim 3 we have that $\alpha_n^{-1} \xrightarrow{\text{g}} \alpha^{-1}$, whence $\alpha_n \xrightarrow{\text{g}} \alpha$. ∎

3. Setup and preliminary results. Before stating our working assumptions in section 3.2, in section 3.1 we recall all the basic definitions and tools of measure theory and convex analysis, which we will use in the following.

3.1. Preliminaries of measure theory, *BV* functions, and convex analysis. We start with the notion of measure with values in a Banach space X, which later will coincide either with the reflexive space X or with $\mathbb{R}$.

Definition 3.1 (vector measure). *Let (Ω,Σ) be a measurable space. A function $\mu : \Sigma \to \mathsf{X}$ is called a (Banach space valued) vector measure if*

$$\forall\, (A_i)_{i\in\mathbb{N}}, A_i \in \Sigma \text{ with } (i \neq j \Rightarrow A_i \cap A_j = \emptyset) \quad \text{it holds that} \quad \mu\left(\bigcup_{i\in\mathbb{N}} A_i\right) = \sum_{i=1}^{\infty} \mu(A_i). \tag{3.1}$$

Here the convergence of the series on the right-hand side has to be understood in terms of the norm of X.

Definition 3.2 (variation of a measure). *Let (Ω,Σ) be a measurable space and $\mu : \Sigma \to \mathsf{X}$ a vector measure. Then the variation of μ, denoted by $\|\mu\| : \Sigma \to [0,\infty]$, is given by*

$$\|\mu\|(A) := \sup\left\{\sum_{i=1}^{\infty} \|\mu(A_i)\| \,:\, (A_i)_{i\in\mathbb{N}} \subset \Sigma,\ \bigcup_{i\in\mathbb{N}} A_i = A \ \ \forall i \neq j : A_i \cap A_j = \emptyset\right\} \tag{3.2}$$

for all $A \in \Sigma$. If $\|\mu\|(\Omega) < \infty$, then we say μ is of bounded variation.

Indeed, $\|\mu\|$ itself is a (positive) measure on (Ω, Σ); see [30, Prop. 9, p. 3].

DEFINITION 3.3 (absolute continuity and singularity of measures). *Let (Ω, Σ) be a measurable space, $\mu : \Sigma \to \mathsf{X}$ be a vector measure, and $\nu : \Sigma \to [0, \infty]$ be a (real-valued, positive) measure. We say that μ is absolutely continuous w.r.t. to ν, and write $\mu \ll \nu$, if*

$$\forall A \in \Sigma : \Big[\ \nu(A) = 0 \quad \Longrightarrow \quad \mu(A) = 0\ \Big]. \tag{3.3}$$

Moreover, we say that two real-valued, positive measures μ and ν are singular, and write $\mu \perp \nu$, if there exist $B_1, B_2 \in \Sigma$ with $B_1 \cup B_2 = \Omega$ and $B_1 \cap B_2 = \emptyset$ such that

$$\forall A \in \Sigma : \qquad \mu(A) = \mu\left(A \cap B_1\right) \quad \text{and} \quad \nu(A) = \nu\left(A \cap B_2\right). \tag{3.4}$$

We recall the following generalization of the Lebesgue decomposition theorem; see, e.g., [30, Thm. 9, p. 31].

THEOREM 3.4 (Lebesgue decomposition theorem). *Let (Ω, Σ) be a measure space, σ be a Banach space valued measure of bounded variation, and λ be a real valued, positive measure. Then there exist two unique vector measures σ_{ac}, σ_{sin} on (Ω, Σ), which are of bounded variation, such that*

$$\|\sigma_{\mathrm{ac}}\| \ll \lambda\,, \ \|\sigma_{\mathrm{sin}}\| \perp \lambda, \quad \text{and} \quad \sigma = \sigma_{\mathrm{ac}} + \sigma_{\mathrm{sin}}\,. \tag{3.5}$$

***BV* functions.** We fix here some definitions and notation concerning BV functions on $[0, T]$ with values in a Banach space X, referring the reader, e.g., to [59] for a comprehensive introduction to the topic. We denote by $BV([0, T]; \mathsf{X})$ the space of the measurable, pointwise defined at every time $t \in [0, T]$, functions $v : [0, T] \to \mathsf{X}$ such that their *pointwise* total variation on $[0, T]$ is finite, i.e.,

$$\mathrm{Var}(v; [0, T]) = \sup\left\{\sum_{m=1}^{M} \|v(t_m) - v(t_{m-1})\| \ : \ 0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = T\right\} < \infty\,.$$

We endow $BV([0, T]; \mathsf{X})$ with the total variation norm. More generally, given a convex, lower semicontinuous, 1-positively homogeneous functional $\psi : X \to [0, +\infty)$, we denote by Var_ψ the induced total variation, i.e.,

$$\mathrm{Var}_\psi(v; [0, T]) = \sup\left\{\sum_{m=1}^{M} \psi(v(t_m) - v(t_{m-1})) \ : \ 0 = t_0 < t_1 < \cdots < t_{M-1} < t_M = T\right\} < \infty\,. \tag{3.6}$$

It is well known that the distributional derivative $\mathrm{d}v$ of a curve $v \in BV(0, T; X)$ is a vector measure in $\mathcal{M}(0, T; \mathsf{X})$, where

$$\mathcal{M}(0, T; \mathsf{X}) = \{\text{Radon vector measures } \mu : \mathcal{B}(0, T) \to \mathsf{X} \text{ with bounded variation}\},$$

where $\mathcal{B}(0, T)$ denotes the Borel sets of $(0, T)$. We endow it with the weak*-topology.

Notation 3.5. Let $u \in BV([0,T];X)$ with X reflexive. Applying Thm. 3.4 with the choices $\sigma = \mathrm{d}u$ and $\lambda = \mathcal{L}$ (where $\mathcal{L}$ denotes the one-dimensional Lebesgue measure on $[0,T]$), we find that there exist vector measures $(\mathrm{d}u)_{\mathrm{ac}}$, $(\mathrm{d}u)_{\mathrm{sin}} \in \mathcal{M}(0,T;X)$ such that

$$\|(\mathrm{d}u)_{\mathrm{ac}}\| \ll \mathcal{L}, \quad \|(\mathrm{d}u)_{\mathrm{sin}}\| \perp \mathcal{L}, \quad \text{and} \quad \mathrm{d}u = (\mathrm{d}u)_{\mathrm{ac}} + (\mathrm{d}u)_{\mathrm{sin}}. \tag{3.7}$$

Thanks to the Radon–Nikodým property of the *reflexive* space X, the Radon–Nikodým derivatives

$$\dot{u}_{\mathrm{ac}}(t) := \frac{(\mathrm{d}u)_{\mathrm{ac}}}{\mathrm{d}\mathcal{L}}\,, \ \dot{u}_{\mathrm{sin}}(t) := \frac{(\mathrm{d}u)_{\mathrm{sin}}}{\|(\mathrm{d}u)_{\mathrm{sin}}\|} \qquad \text{exist for a.a. } t \in (0,T). \tag{3.8}$$

For later use, we remark that for any convex, lower semicontinuous, 1-positively homogeneous $\psi : X \to [0,+\infty)$ there holds

$$\mathrm{Var}_{\psi}(u;[0,T]) = \int_0^T \psi(\dot{u}_{\mathrm{ac}}(t))\,\mathrm{d}t + \int_0^T \psi(\dot{u}_{\mathrm{sin}}(t))\|(\mathrm{d}u)_{\mathrm{sin}}\|(t). \tag{3.9}$$

The recession function. Finally, we recall the concept of *recession function* (see [35, Chap. 4]). Note that the following definitions and results, which are stated in [35] for convex functions on $\mathbb{R}^m$, in fact extend to an infinite-dimensional setting, as it can be easily checked.

DEFINITION 3.6 (recession function). *Let X be a vector space and $g : \mathsf{X} \to (-\infty,\infty]$ be a convex functional. Its recession function g^∞ is defined as*

$$g^\infty(z) := \sup\left\{g(y+z) - g(y) : y \in D(g)\right\}. \tag{3.10}$$

Trivially adapting the argument from [35, Thm. 4.70, p. 290], it can be shown that that g^∞ is positively homogeneous of degree 1 and convex. Moreover, if g is lower semicontinuous, so is g^∞. Furthermore, there holds

$$g^\infty(z) = \lim_{t\to\infty} \frac{g(y+tz) - g(y)}{t} = \sup_{t>0} \frac{g(y+tz)-g(y)}{t} \quad \text{for every } y \in D(g). \tag{3.11}$$

In what follows, we will denote by f_α^∞ the recession function of the Fitzpatrick function f_α, viz.

$$f_\alpha^\infty(\xi,\xi^*) = \sup\left\{f_\alpha(\xi+x,\xi^*+x^*) - f_\alpha(x,x^*) : (x,x^*) \in D(f_\alpha)\right\}. \tag{3.12}$$

We now prove a useful representation formula for f_α^∞; cf. [35, Prop. 4.77, p. 294].

LEMMA 3.7. *There holds*

$$f_\alpha^\infty(\xi,\xi^*) = \sup\{\langle\xi^*,\xi_0\rangle + \langle\xi_0^*,\xi\rangle \ : \ (\xi_0,\xi_0^*) \in D(\rho_{\alpha^{-1}})\} \quad \forall\,(\xi,\xi^*) \in X\times X^*. \tag{3.13}$$

Proof. Following the proof of [35, Prop. 4.77], from (3.11) and (2.7) we infer

$$
\begin{aligned}
f_\alpha^\infty(\xi,\xi^*) &= \sup_{t>0} \frac{f_\alpha(x+t\xi, x^*+t\xi^*) - f_\alpha(x,x^*)}{t} \\
&\geq \sup_{t>0} \frac{1}{t}\Big(t\,\langle \xi^*, \xi_0\rangle + t\,\langle \xi_0^*, \xi\rangle \\
&\qquad\quad + \langle \xi_0^*, x\rangle + \langle x^*, \xi_0\rangle - \rho_{\alpha^{-1}}(\xi_0,\xi_0^*) - f_\alpha(x,x^*)\Big) \\
&\geq \langle \xi^*, \xi_0\rangle + \langle \xi_0^*, \xi\rangle \\
&\quad + \frac{1}{t}\left(\langle \xi_0^*, x\rangle + \langle x^*, \xi_0\rangle - \rho_{\alpha^{-1}}(\xi_0,\xi_0^*) - f_\alpha(x,x^*)\right) \\
&\quad \forall\,(\xi_0,\xi_0^*) \in X\times X^*,\ t>0.
\end{aligned}
$$

In view of (2.7), we thus conclude that

$$
f_\alpha^\infty(\xi,\xi^*) \geq \sup\{\langle \xi^*, \xi_0\rangle + \langle \xi_0^*, \xi\rangle \ :\ (\xi_0,\xi_0^*) \in D(\rho_{\alpha^{-1}})\}.
$$

The converse inequality may be proved arguing along the very same lines; cf. also the proof of [35, Prop. 4.77]. ∎

As a direct consequence of Lemma 3.7, we have the following representation formula for the recession function of f_α in the case that α is the subdifferential of a 1-positively homogeneous potential.

Corollary 3.8. *Let $\psi : X \to \mathbb{R}$ be convex, lower semicontinuous, and positively homogeneous of degree* 1, *and let $K^* \subset X^*$ be the associated polar set; cf.* (2.12). *Then,*

(3.14)
$$
f_{\partial\psi}^\infty(\xi,\xi^*) = \sup\{\langle \xi_0^*, \xi\rangle + \langle \xi^*, \xi_0\rangle \ :\ (\xi_0,\xi_0^*) \in X\times K^*\} \ \text{ for all}\,(\xi,\xi^*) \in X\times X^*.
$$

Proof. Formula (3.14) follows from (3.13), taking into account that

$$
\rho_{\alpha^{-1}} = \rho_{\partial\psi^*} = \psi + \psi^* = \psi + I_{K^*}
$$

and that $D(\psi) = X$ by assumption. ∎

3.2. Basic assumptions. In what follows, we will suppose that

$$
X \text{ is a reflexive Banach space} \tag{3.15}
$$

and that

$$
\alpha : X \rightrightarrows X^* \text{ is a maximal monotone operator with } 0 \in \alpha(0). \tag{3.α_0}
$$

As for the energy functional $\mathcal{E}$, along the lines of [51] we require the following *coercivity* and *regularity* type conditions. Recall that $\partial\mathcal{E}$ denotes the *Fréchet subdifferential* of the map $u \mapsto \mathcal{E}_t(u)$; cf. (1.2).

Assumption 3.9 (assumptions on the energy). *We assume that the pair* $(\mathcal{E}, \partial\mathcal{E})$ *has the following properties.*

Lower semicontinuity. *The domain of $\mathcal{E}$ is of the form $D(\mathcal{E}) = [0,T]\times D$ for some $D \subset X$, and $\partial\mathcal{E} : [0,T]\times D \rightrightarrows X^*$. Furthermore, we ask that*

(3.$\mathcal{E}_0$)
$$
\begin{gathered}
u \mapsto \mathcal{E}_t(u) \ \textit{is l.s.c.}\ \forall\, t\in[0,T], \quad \exists C_0>0 :\ \forall (t,u)\in[0,T]\times D : \mathcal{E}_t(u) \geq C_0, \ \textit{and} \\
\operatorname{graph}(\partial\mathcal{E}) \ \textit{is a Borel set of}\ [0,T]\times X\times X^*.
\end{gathered}
$$

Coercivity. *Set* $\mathcal{G}(u) := \sup_{t\in[0,T]} \mathcal{E}_t(u)$ *for every* $u \in D$. *We require that*

$$u \mapsto \mathcal{G}(u) \text{ has compact sublevels.} \tag{3.$\mathcal{E}_1$}$$

Time-differentiability. *For any* $u \in D$ *the map* $t \mapsto \mathcal{E}_t(u)$ *is differentiable with derivative* $\partial_t\mathcal{E}_t(u)$*, and it holds that*

$$\exists C_1 > 0 : \forall u \in D : \quad |\partial_t\mathcal{E}_t(u)| \le C_1\mathcal{E}_t(u). \tag{3.$\mathcal{E}_2$}$$

Weak closedness. *For all* $t \in [0,T]$ *and for all sequences* $(u_n)_{n\in\mathbb{N}} \subset X$*,* $\xi_n \in \partial\mathcal{E}_t(u_n)$*,* $E_n = \mathcal{E}_t(u_n)$*, and* $p_n = \partial_t\mathcal{E}_t(u_n)$ *with*

$$u_n \to u \text{ in } X, \quad \xi_n \rightharpoonup \xi \text{ in } X^*, \quad p_n \to p, \quad \text{and} \quad E_n \to E \text{ in } \mathbb{R},$$

it holds that

$$(t,u) \in D(\partial\mathcal{E}),\ \xi \in \partial\mathcal{E}_t(u),\ p \le \partial_t\mathcal{E}_t(u), \quad \text{and} \quad E = \mathcal{E}_t(u)\,. \tag{3.$\mathcal{E}_3$}$$

Remark 3.10. In fact, up to a translation, we may always suppose that the constant involved in (3.$\mathcal{E}_0$) is strictly positive. As in [51], combining (3.$\mathcal{E}_2$) with the Gronwall lemma, we observe that

$$\exists\, C > 0 \qquad \forall\,(t,u) \in [0,T]\times D, \qquad \mathcal{G}(u) \le C \inf_{t\in[0,T]} \mathcal{E}_t(u). \tag{3.16}$$

Later, Assumption 3.9 will be complemented by a suitable version of the chain rule for $\mathcal{E}$; cf. Assumption 4.4 below. As already mentioned, in order to investigate the stability properties of the doubly nonlinear equation

$$\alpha(\dot u(t)) + \partial\mathcal{E}_t(u(t)) \ni 0 \text{ in } X^* \quad \text{for a.a. } t \in (0,T)\,, \tag{3.17}$$

under graph convergence of α, it is essential to resort to the Fitzpatrick function f_α associated with α. In the following lines, we will therefore shed light on how (3.17) can in fact be reformulated in terms of an *energy identity* (cf. (3.20) below) featuring f_α. At first, we will confine the discussion to the case of *absolutely continuous* solutions u to (3.17).

Reformulations of (3.17) in the absolutely continuous case. Preliminarily, let us precisely define what we understand by an *absolutely continuous* solution to (3.17).

Definition 3.11 (absolutely continuous solution). *In the framework of* (3.15)*,* (3.α_0)*, and* (3.$\mathcal{E}_0$)*, we say that a curve* $u \in W^{1,1}(0,T;X)$ *is a solution to* (3.17) *if there exists* $\xi \in L^1(0,T;X^*)$ *with*

$$\xi(t) \in (-\alpha(\dot u(t))) \cap \partial\mathcal{E}_t(u(t)) \qquad \text{for a.a. } t \in (0,T)\,. \tag{3.18}$$

In what follows, with a slight abuse of notation we will sometimes say that (u,ξ) is a solution to (3.17), meaning that (3.18) holds.

In Proposition 3.12, we reformulate (3.18) by means of an energy identity involving the Fitzpatrick function f_α. In the proof, a key role is played by the chain-rule condition (3.19) below on the energy $\mathcal{E}$, whereas note that not all of the conditions collected in Assumption 3.9 are needed.

Proposition 3.12 (variational reformulation). *In the framework of* (3.15)*, let* $\alpha : X \rightrightarrows X^*$ *fulfill* (3.α_0)*, and suppose that* $\mathcal{E} : [0,T]\times X \to (-\infty,+\infty]$ *complies with*

(3.$\mathcal{E}_0$), (3.$\mathcal{E}_1$), (3.$\mathcal{E}_2$), *and the following chain rule: for every* $u \in W^{1,1}(0,T;X)$ *and* $\xi \in L^1(0,T;X^*)$ *such that*

$$\sup_{t\in[0,T]} \mathcal{E}_t(u(t)) < \infty, \quad \xi(t) \in \partial\mathcal{E}_t(u(t)) \text{ for a.a. } t \in (0,T)\,, \quad \int_0^T f_\alpha(\dot u(t), -\xi(t))\,\mathrm{d}t < \infty$$

(observe that, thanks to (3.$\mathcal{E}_2$)*, the first of the conditions above guarantees that* $\int_0^T |\partial_t \mathcal{E}_t(u(t))|\,\mathrm{d}t < \infty$ *as well), there holds*

$$\text{(3.19)}\qquad \begin{aligned} &\text{the map } t \mapsto \mathcal{E}_t(u(t)) \text{ is absolutely continuous and} \\ &\frac{\mathrm{d}}{\mathrm{d}t}\mathcal{E}_t(u(t)) = \langle \xi(t), \dot u(t)\rangle + \partial_t\mathcal{E}_t(u(t)) \qquad \text{for a.a. } t \in (0,T). \end{aligned}$$

Then, the following implications hold:

1. *If* $(u,\xi) \in W^{1,1}(0,T;X) \times L^1(0,T;X^*)$ *fulfills the energy identity*
(3.20)
$$\mathcal{E}_t(u(t)) + \int_0^t f_\alpha\left(\dot u(s), -\xi(s)\right)\,\mathrm{d}s = \mathcal{E}_0(u(0)) + \int_0^t \partial_t\mathcal{E}_s(u(s))\,\mathrm{d}s \qquad \forall\, t \in (0,T],$$
then (u,ξ) *is a solution to* (3.17) *in the sense of Definition* 3.11.
2. *Every solution* (u,ξ) *to* (3.17) *(in the sense of Definition* 3.11*) fulfilling*
$$\text{(3.21)}\qquad \sup_{t\in[0,T]} \mathcal{E}_t(u(t)) < \infty, \qquad \int_0^T |\,\langle \xi(t), \dot u(t)\rangle\,|\,\mathrm{d}t < \infty$$
complies in addition with the energy identity (3.20).

Observe that, for every solution (u,ξ) to (3.17), since $-\xi \in \alpha(\dot u)$ a.e. in $(0,T)$ and $0 \in \alpha(0)$, we have $\langle -\xi, \dot u\rangle \geq 0$ a.e. in $(0,T)$. Hence the second inequality of (3.21) in fact reduces to $\int_0^T \langle -\xi, \dot u\rangle\,\mathrm{d}t < \infty$.

Proof. Let (u,ξ) fulfill (3.20). Taking into account that $f_\alpha(\dot u, -\xi) \geq 0$ a.e. in $(0,T)$ thanks to (2.9), and exploiting (3.$\mathcal{E}_2$), we gather that

$$\text{(3.22)}\qquad \mathcal{E}_t(u(t)) \leq \mathcal{E}_0(u(0)) + C_1 \int_0^t \mathcal{E}_s(u(s))\,\mathrm{d}s \qquad \forall\, t \in (0,T],$$

whence $\sup_{t\in[0,T]} \mathcal{E}_t(u(t)) < \infty$. Therefore, a fortiori (3.20) yields $f_\alpha(\dot u, -\xi) \in L^1(0,T)$. Hence the pair (u,ξ) fulfills the conditions for the chain rule (3.19), which yields for all $t \in (0,T]$

$$\text{(3.23)}\qquad \begin{aligned} \int_0^t f_\alpha\left(\dot u(s), -\xi(s)\right)\,\mathrm{d}s &\leq \mathcal{E}_0(u(0)) - \mathcal{E}_t(u(t)) + \int_0^t \partial_t\mathcal{E}_s(u(s))\,\mathrm{d}s \\ &\leq \int_0^T \langle -\xi(s), \dot u(t)\rangle\,\mathrm{d}s. \end{aligned}$$

Using that f_α represents α, it is immediate to deduce from the above inequality that $-\xi(t) \in \alpha\left(\dot u(t)\right)$ for almost all $t \in (0,T)$; thus (u,ξ) is a solution to (3.17) in the sense of Definition 3.11.

Conversely, let $(u,\xi) \in W^{1,1}(0,T;X) \times L^1(0,T;X^*)$ be a solution to (3.17) (in the sense of Definition 3.11) fulfilling in addition (3.21). Then, since $f_\alpha(\dot u, -\xi) = \langle -\xi, \dot u\rangle$, the chain rule (3.19) applies, yielding, for all $t \in [0,T]$, the energy identity

$$\int_0^t f_\alpha\left(\dot u(s), -\xi(s)\right) \mathrm{d}s = \int_0^t \langle -\xi(s), \dot u(t)\rangle \, \mathrm{d}s$$
$$= \mathcal{E}_0(u(0)) - \mathcal{E}_t(u(t)) + \int_0^T \partial_t \mathcal{E}_s(u(s)), \xi(s))\, \mathrm{d}s\,. \qquad \square$$

Remark 3.13. A few comments on Proposition 3.12 are in order.

1. It is not difficult to check that in Proposition 3.12 the Fitzpatrick function f_α could be replaced by *any* representative functional for α.
2. Observe that, in the chain of inequalities (3.23) leading to the proof of part (1) of Proposition 3.12, the fact that (3.20) and the chain rule (3.19) hold as inequalities only is used. The proof of part (2) requires (3.19) to hold as an equality instead.

4. Main results. Before stating Theorem 4.5, let us specify our hypothesis on the sequence (α_n) of maximal monotone operators.

ASSUMPTION 4.1. *Let* $\alpha_n : X \rightrightarrows X^*$ *fulfill* $(3.\alpha_0)$ *for all* $n \in \mathbb{N}$ *and*

$$(3.\alpha_1) \qquad \exists\, c_1, c_2, c_3 > 0, \quad p \geq 1, \quad q > 1 \qquad \forall\, n \in \mathbb{N} \quad \forall (x,y) \in \alpha_n : \\ \langle y, x\rangle \geq c_1 \|x\|^p + c_2 \|y\|_*^q - c_3\,.$$

Furthermore, there exists $\alpha : X \rightrightarrows X^*$ *fulfilling* $(3.\alpha_0)$ *such that* $\alpha_n \xrightarrow{\mathrm{g}} \alpha$.

Remark 4.2. Combining $(3.\alpha_1)$ with the graph convergence of (α_n) to α, it is immediate to conclude

$$\langle y, x\rangle \geq c_1 \|x\|^p + c_2 \|y\|_*^q - c_3 \qquad \forall\,(x,y) \in \alpha\,. \tag{4.1}$$

The following example guarantees that our analysis encompasses the $p_n \to 1$ vanishing-viscosity limit.

Example 4.3. Let $(p_n) \subset [1, +\infty)$ fulfill $p_n \downarrow 1$ as $n \to \infty$, and let us set

$$\psi_n(x) = \frac{1}{p_n}\|x\|^{p_n}, \qquad \alpha_n = \partial\psi_n : X \rightrightarrows X^*\,.$$

Clearly, (ψ_n) Mosco-converges to $\psi(x) = \|x\|$; hence (α_n) converges in the sense of graphs to $\alpha = \partial\psi$. Observe that $\psi_n^*(y) = \frac{1}{q_n}\|y\|_*^{q_n}$ with $q_n = p_n/(p_n - 1) \in [2, \infty]$ for all $n \in \mathbb{N}$ and that

$$\langle y, x\rangle = \frac{1}{p_n}\|x\|^{p_n} + \frac{1}{q_n}\|y\|_*^{q_n} = \|x\|^{p_n} = \|y\|_*^{q_n} \quad \forall\,(x,y) \in \alpha_n\,.$$

Therefore, Assumption 4.1 is satisfied.

The main result of this section addresses the passage to the limit as $n \to \infty$ in the doubly nonlinear equations

$$\alpha_n(\dot u(t)) + \partial\mathcal{E}_t(u(t)) \ni 0 \text{ in } X^* \quad \text{for a.a. } t \in (0,T)\,. \tag{4.2}$$

In particular, we will assume to be given a sequence (u_n) of absolutely continuous solutions to (4.2), and we will show that, if the sequence (α_n) complies with Assumption 4.1, up to a subsequence, then (u_n) converges to a curve u fulfilling a suitable generalized formulation of (3.17).

Observe that $(3.\alpha_1)$ in principle allows only for a bound of the type $\|\dot{u}_n\|_{L^1(0,T;X)} \leq C$. That is why we can expect only a $BV([0,T];X)$-regularity for the limiting curve u, and (3.17) has to be weakly formulated accordingly. This will be done through an energy *inequality* akin to (3.20); cf. (4.5) below. Therein, suitable replacements of the "time-derivative" of u are suitably handled in terms of the Fitzpatrick function f_α, of its recession function f_α^∞ (cf. Definition 3.6), and of the absolutely continuous and singular parts of the Radon derivative $\mathrm{d}u$ of u. Having in mind the role of the chain rule (3.19) relating (3.17) and the energy identity (3.20), we expect that a suitable BV *version* of (3.19) will play a relevant role. We state it in the following.

ASSUMPTION 4.4. *Let* $u \in BV([0,T];X)$ *and* $\xi \in L^1(0,T;X^*)$ *fulfill*

$$\sup_{t\in[0,T]} \mathcal{E}_t(u(t)) < \infty, \quad \xi(t) \in \partial\mathcal{E}_t(u(t)) \text{ for a.a. } t \in (0,T),$$

$$\int_0^T f_\alpha(\dot{u}(t), -\xi(t))\,\mathrm{d}t < \infty,$$

and suppose that the map $t \mapsto \mathcal{E}_t(u(t))$ *is a.e. equal on* $(0,T)$ *to a function* $E \in BV([0,T])$. *Furthermore, let* $\mathrm{d}u$ *and* $\mathrm{d}E$ *denote the Radon derivatives of* u *and* E.

Then, for almost all Lebesgue points t_0 *of the absolutely continuous parts* $\dot{u}_{\mathrm{ac}}$ *and* $\dot{E}_{\mathrm{ac}}$ *of* $\mathrm{d}u$ *and* $\mathrm{d}E$ *there holds*

$$(3.\mathcal{E}_4) \qquad \dot{E}_{\mathrm{ac}}(t_0) \geq \langle \xi(t_0), \dot{u}_{\mathrm{ac}}(t_0)\rangle + \partial_t\mathcal{E}_{t_0}(u(t_0)) \ \forall\, \xi(t_0) \in \partial\mathcal{E}_{t_0}(u(t_0)).$$

Observe that, since X has the Radon–Nikodým property, the set of Lebesgue points of $\dot{u}_{\mathrm{ac}}$ and $\dot{E}_{\mathrm{ac}}$ has full Lebesgue measure in $(0,T)$.

As will be clear from the proof of Theorem 4.5 below, Assumption 4.4 not only provides a motivation for the energy inequality (4.5) but also has a key role in the proof of the passage to the limit as $n \to \infty$ in (4.2).

THEOREM 4.5. *Assume* (3.15). *Let* $\alpha_n, \alpha : X \rightrightarrows X^*$ *fulfill Assumption* 4.1, *and suppose that* $\mathcal{E} : [0,T] \times X \to (-\infty, +\infty]$ *complies with Assumptions* 3.9 *and* 4.4. *Let us consider a sequence* $(u_0^n) \subset D$ *of initial data such that*

$$(4.3) \qquad u_0^n \rightharpoonup u_0 \quad \text{in } X, \qquad \mathcal{E}_0(u_0^n) \to \mathcal{E}_0(u_0),$$

and let $(u_n, \xi_n) \subset W^{1,1}(0,T;X) \times L^1(0,T;X^*)$ *be solutions to* (4.2) *in the sense of Definition* 3.11, *fulfilling the initial conditions* $u_n(0) = u_0^n$. *Suppose that, in addition, for all* $n \in \mathbb{N}$ *the functions* (u_n, ξ_n) *comply with the energy identity* (3.20).

Then, there exist functions $u \in BV([0,T];X)$ *and* $\xi \in L^q(0,T;X^*)$ *(with* $q > 1$ *from* $(3.\alpha_1)$*) satisfying* $u(0) = u_0$, $\xi(t) \in \partial\mathcal{E}_t(u(t))$ *for almost all* $t \in (0,T)$, *and such that up to a (not relabeled) subsequence*

$$(4.4) \qquad u_n(t) \to u(t)\ \forall t \in [0,T], \ \ \mathrm{d}u_n = (\dot{u}_n)_{\mathrm{ac}} \cdot \mathcal{L}|_{[0,T]} \overset{*}{\rightharpoonup} \mathrm{d}u \in \mathcal{M}(0,T;X),$$

and (u, ξ) *satisfies the energy inequality*

$$(4.5) \qquad \begin{aligned} &\mathcal{E}_t(u(t)) + \int_0^t f_\alpha\left(\dot{u}_{\mathrm{ac}}(s), -\xi(s)\right)\mathrm{d}s + \int_0^t f_\alpha^\infty\left(\dot{u}_{\sin}(s), 0\right)\|(\mathrm{d}u)_{\sin}\|(s) \\ &\qquad \leq \mathcal{E}_0(u(0)) + \int_0^t \partial_t\mathcal{E}_s(u(s))\,\mathrm{d}s \qquad \forall\, t \in [0,T] \end{aligned}$$

as well as

$$(4.6) \qquad \xi(t) \in (-\alpha(\dot{u}_{\mathrm{ac}}(t))) \cap \partial\mathcal{E}_t(u(t)) \qquad \text{for a.a. } t \in (0,T).$$

Furthermore, there exists $E \in BV([0,T])$ *such that*

$$E(t) = \mathcal{E}_t(u(t)) \quad \text{for a.a. } t \in (0,T), \qquad E(t) \geq \mathcal{E}_t(u(t)) \quad \forall\, t \in [0,T], \tag{4.7}$$

and we have the pointwise energy identity

$$\dot{E}_{\rm ac}(t) + f_\alpha\left(\dot{u}_{\rm ac}(t), -\xi(t)\right) = \partial_t \mathcal{E}_t(u(t)) \qquad \text{for a.a. } t \in (0,T). \tag{4.8}$$

Remark 4.6. In view of Proposition 3.12, a sufficient condition for functions (u_n, ξ_n) solving (4.2) to comply with the energy identity (3.20), is that they fulfill

$$\sup_{t\in(0,T)} \mathcal{E}_t(u_n(t)) < \infty \qquad \text{and} \qquad \langle -\xi_n, \dot{u}_n\rangle \in L^1(0,T).$$

This is provided that the *absolutely continuous* version (3.19) of the chain rule holds.

In section 4.2, we will discuss some sufficient conditions on $\mathcal{E}$ for both the chain rule (3.19) and its BV-version of Assumption 4.4 to hold.

4.1. Further results. We conclude this section with some results which shed light on the interpretation of the energy identities (4.5) and (4.8) satisfied by the pair (u, ξ). More precisely, we have the following:

- Proposition 4.7 focuses on the case in which we have the additional information that u is absolutely continuous. For instance, this is granted whenever u occurs as limiting curve of a sequence $(u_n) \subset W^{1,1}(0,T;X)$ of solutions to the differential inclusions (4.2), driven by operators (α_n) which fulfill a stronger version of condition $(3.\alpha_1)$; cf. Theorem 4.8.
- In Proposition 4.9 we address the special case in which $\alpha = \partial\psi$, with $\psi : X \to [0,+\infty)$ a convex, lower semicontinuous, and 1-homogeneous dissipation potential. We show that in this case any $u \in BV([0,T];X)$ complying with the energy inequality (4.5) is a *local solution* (cf. [49, 50]) to the rate-independent system $(X, \mathcal{E}, \psi)$.

The absolutely continuous case. Under a slightly stronger version of the chain rule of Assumption 4.4, Proposition 4.7 shows that, if in addition we have that the curve u is absolutely continuous on $(0,T)$, then $f_\alpha^\infty\left(\dot{u}_{\rm sin}(t), 0\right) = 0$ for $\|(\mathrm{d}u)_{\rm sin}\|$-a.a., $t \in (0,T)$, and (4.5) holds on every subinterval $[s,t] \subset [0,T]$. Furthermore, the pair (u,ξ) solves (3.17) in the sense of Definition 3.11; cf. (4.11) below.

PROPOSITION 4.7. *In the framework of* (3.15)*, let* $\alpha : X \rightrightarrows X^*$ *fulfill* $(3.\alpha_0)$*, and let* $\mathcal{E} : [0,T] \times X \to (-\infty, +\infty]$ *comply with Assumption* 3.9 *and with the following chain rule: for every* $u \in W^{1,1}(0,T;X)$ *and* $\xi \in L^1(0,T;X^*)$ *such that*

$$\sup_{t\in[0,T]} \mathcal{E}_t(u(t)) < \infty, \quad \xi(t) \in \partial\mathcal{E}_t(u(t)) \text{ for a.a. } t \in (0,T),$$

$$\int_0^T f_\alpha(\dot{u}(t), -\xi(t))\,\mathrm{d}t < \infty,$$

then

$$(\mathrm{d}u)_{\rm sin} = 0 \quad \Rightarrow \quad (\mathrm{d}E)_{\rm sin} = 0, \tag{4.9}$$

and the chain rule inequality $(3.\mathcal{E}_4)$ *holds.*

Let $(u, \xi, E) \in BV([0,T];X) \times L^1(0,T;X^*) \times BV([0,T])$ *fulfill* (4.7) *and* (4.8)*. Suppose in addition that* $u \in W^{1,1}(0,T;X)$*. Then,*

$$E \in W^{1,1}(0,T). \tag{4.10}$$

Furthermore, the pair (u,ξ) *fulfills*

$$-\xi(t) \in \alpha(\dot{u}(t)) \qquad \text{for a.a. } t \in (0,T), \tag{4.11}$$

and there holds the improved *energy inequality*

$$\begin{aligned}\mathcal{E}_t(u(t)) &+ \int_s^t f_\alpha\left(\dot{u}(r), -\xi(r)\right) \mathrm{d}r \\ &\leq \mathcal{E}_s(u(s)) + \int_s^t \partial_t \mathcal{E}_r(u(r))\, \mathrm{d}r \ \forall\, t \in (0,T],\ \text{for a.a. } s \in (0,t),\ \text{and for } s=0.\end{aligned} \tag{4.12}$$

Finally, if $\mathcal{E}$ *also fulfills the enhanced chain rule* (3.19)*, then* (4.12) *holds as an* equality *for every* $0 \leq s \leq t \leq T$.

Proof. Since $u \in W^{1,1}(0,T;X)$, its distributional derivative $\mathrm{d}u$ has zero singular part, viz. $\mathrm{d}u = \dot{u}_{\mathrm{ac}}\mathcal{L}$. Then, it follows from (4.9) that $\mathrm{d}E = \dot{E}_{\mathrm{ac}}\mathcal{L}$, viz. E is absolutely continuous. Therefore, (4.8) becomes

$$\dot{E}(t) + f_\alpha\left(\dot{u}(t), -\xi(t)\right) = \partial_t \mathcal{E}_t(u(t)) \qquad \text{for a.a. } t \in (0,T). \tag{4.13}$$

Now, combining this with the chain rule inequality $(3.\mathcal{E}_4)$, we conclude that $f_\alpha(\dot{u}(t), -\xi(t)) \leq \langle -\xi(t), \dot{u}(t)\rangle$ for almost all $t \in (0,T)$; hence (4.11) holds. Then, to prove (4.12) we integrate (4.13), thus obtaining

$$E(t) + \int_s^t f_\alpha\left(\dot{u}(r), -\xi(r)\right)\, \mathrm{d}r = E(s) + \int_s^t \partial_t \mathcal{E}_r(u(r))\, \mathrm{d}r \qquad \forall\, 0 \leq s \leq t \leq T, \tag{4.14}$$

and we use (4.7).

If, moreover, $\mathcal{E}$ complies with the chain rule (3.19), then $E(t) = \mathcal{E}_t(u(t))$ for all $t \in [0,T]$, since both functions $t \mapsto E(t)$ and $t \mapsto \mathcal{E}_t(u(t))$ are continuous on $[0,T]$ and coincide on a set of full Lebesgue measure. Therefore, from (4.14) we get (4.12) for $t \mapsto \mathcal{E}_t(u(t))$. This concludes the proof. ∎

As a straightforward consequence of Proposition 4.7 we have the following result, showing that, under a stronger coercivity assumption on the sequence of maximal monotone operators (α_n) (cf. (4.15) below), any sequence (u_n) of solutions to (4.2) converges up to a subsequence to a curve complying with (4.10)–(4.12). In particular, observe that, unlike in $(3.\alpha_1)$, in (4.15) we do not allow the "degenerate" value 1 for exponent p. Indeed, Theorem 4.8 below, for instance, applies to a sequence of operators $\alpha_n = \partial\psi_n$, with $\psi_n(v) = 1/p_n\|v\|^{p_n}$ and $p_n \downarrow p > 1$ as $n \to \infty$. In this way, we obtain a stability result for doubly nonlinear differential inclusions driven by *viscous* dissipation potentials, which generalizes the results in [1, Thms. 3.1, 3.2].

THEOREM 4.8. *In the frame of* (3.15)*, suppose that* $\mathcal{E} : [0,T] \times X \to (-\infty,+\infty]$ *complies with Assumptions* 3.9 *and* 4.4*. Let* $\alpha_n : X \rightrightarrows X^*$ *fulfill* $(3.\alpha_0)$ *for all* $n \in \mathbb{N}$ *and*

$$\begin{aligned}&\exists\, c_1, c_2, c_3 > 0, \quad p > 1, \quad q > 1 \qquad \forall\, n \in \mathbb{N} \quad \forall (x,y) \in \alpha_n\ : \\ &\qquad\qquad\qquad\qquad\qquad\qquad \langle y, x\rangle \geq c_1\|x\|^p + c_2\|y\|_*^q - c_3.\end{aligned} \tag{4.15}$$

Suppose that there exists $\alpha : X \rightrightarrows X^*$ *fulfilling* $(3.\alpha_0)$ *such that* $\alpha_n \xrightarrow{\text{g}} \alpha$*. Let* $(u_0^n) \subset D$ *be a sequence of initial data fulfilling* (4.3)*, and let* $(u_n,\xi_n) \subset W^{1,1}(0,T;X) \times L^1(0,T;X^*)$ *be solutions to* (4.2)*, fulfilling* $u_n(0) = u_0^n$ *and* (3.21) *for every* $n \in \mathbb{N}$*.*

Then, there exists $u \in W^{1,p}(0,T;X)$ *with* $u(0) = u_0$ *such that up to a (not relabeled) subsequence*

$$u_n(t) \to u(t)\ \forall\, t \in [0,T]\,, \quad u_n \rightharpoonup u \quad \text{in } W^{1,p}(0,T;X), \tag{4.16}$$

and there exists $\xi \in L^q(0,T;X^*)$ *such that the pair* (u,ξ) *is a solution to* (3.17) *in the sense of Definition* 3.11*, fulfilling the improved energy inequality* (4.12)*.*

The *proof* is outlined at the end of section 5.

The rate-independent case. Let us now focus on the case in which

$$\alpha = \partial\psi \text{ with } \psi : X \to [0,+\infty) \text{ convex, l.s.c., and 1-positively homogeneous} \tag{4.17}$$

with associated polar set $K^* \subset X^*$. In this case, the energy inequality (4.5) rephrases in a more explicit way.

PROPOSITION 4.9. *Assume* (3.15)*. Let* α *fulfill* (4.17)*, and let* $(u,\xi) \in BV(0,T;X) \times L^1(0,T : X^*)$ *satisfy the energy inequality* (4.5)*. Then,* (u,ξ) *fulfill*

$$-\,\xi(t) \in K^* \quad \text{for a.a. } t \in (0,T), \tag{4.18}$$

$$\mathcal{E}_t(u(t)) + \mathrm{Var}_\psi(u;[0,t]) \le \mathcal{E}_0(u(0)) + \int_0^t \partial_t \mathcal{E}_s(u(s))\,\mathrm{d}s \quad \forall\, t \in [0,T]\,, \tag{4.19}$$

with Var_ψ *from* (3.6)*.*

In the framework of rate-independent evolution, (4.18) is interpreted as a *local stability condition*, while the energy inequality (4.19) balances the stored energy $\mathcal{E}_t(u(t))$ and the dissipated energy $\mathrm{Var}_\psi(u;[0,t])$, with the initial energy and the work of the external forces $\int_0^t \partial_t \mathcal{E}_s(u(s))\,\mathrm{d}s$. In fact, the local stability (4.18) and the energy inequality (4.19) yield (a slightly weaker version of) the notion of *local solution* to the rate-independent system $(X, \mathcal{E}, \psi)$ from [49, 50]. Therein, it was observed that this concept is the weakest among all notions of rate-independent evolution, in that it yields the least precise information on the behavior of the solution at jump points. On the other hand, *local solutions* arise in the limit of a very broad class of approximations of rate-independent systems. This is in the same spirit as the stability results of this work. In particular, notice that the maximal monotone operators α_n converging in the sense of graphs to $\alpha = \partial\psi$ need not be cyclically monotone. An example in this direction in the plane $X = \mathbb{R}^2$ is given by the graphs $\alpha_n = \partial\psi + (1/n)Q$, where Q is a rotation of $\pi/2$. In this case $\alpha_n \xrightarrow{\mathrm{g}} \partial\psi$ but each α_n is noncyclic.

We now proceed with the proof.

Proof of Proposition 4.9. Let $(u,\xi) \in BV(0,T;X) \times L^1(0,T : X^*)$ fulfill (4.5). Now, in view of Proposition 2.5 and of formula (2.12b), we have

$$f_\alpha\left(\dot u_{\mathrm{ac}}(t), -\xi(t)\right) = \psi(\dot u_{\mathrm{ac}}(t)) + \psi^*(-\xi(t)) = \psi(\dot u_{\mathrm{ac}}(t)) + I_{K^*}(-\xi(t)) \quad \text{for a.a. } t \in (0,T)\,. \tag{4.20}$$

Furthermore, we have that

$$f_\alpha^\infty\left(\dot u_{\sin}(t), 0\right) \ge \sup_{\xi_0^* \in K^*} \langle \xi_0^*, \dot u_{\sin}(t)\rangle = \psi(\dot u_{\sin}(t)) \quad \text{for } \|(\mathrm{d}u)_{\sin}\|\text{-a.a. } t \in (0,T), \tag{4.21}$$

where the first inequality is due to (3.14) and the second identity to (2.12d).

Then, taking into account formula (3.9) for Var_ψ, (4.5) yields

$$\mathcal{E}_t(u(t)) + \mathrm{Var}_\psi(u;[0,t]) + \int_0^t I_{K^*}(-\xi(s))\,\mathrm{d}s \le \mathcal{E}_0(u(0)) + \int_0^t \partial_t \mathcal{E}_s(u(s))\,\mathrm{d}s \quad \forall\, t \in [0,T],$$

which is equivalent to (4.18)–(4.19). $\square$

4.2. Sufficient conditions for closedness and chain rule. Following [51], we now show that conditions of λ-convexity type on the energy functional $\mathcal{E}$ ensure the validity of the closedness property (3.$\mathcal{E}_3$), of the chain rules (3.19), (3.$\mathcal{E}_4$), and of property (4.9).

More precisely, in [51, sec. 2] the following *subdifferentiability property* was introduced.

DEFINITION 4.10. *Let* $\mathcal{E} : [0,T]\times X \to (-\infty, +\infty]$ *fulfill* (3.$\mathcal{E}_0$). *For every* $R > 0$, *set*

$$D_R = \{u \in D \; : \; \mathcal{G}(u) \le R\} \, .$$

We say that $\mathcal{E}$ *is* uniformly subdifferentiable *(w.r.t. the variable* u*) if for all* $R > 0$ *there exists a* modulus of subdifferentiability $\omega^R : [0,T] \times D_R \times D_R \to [0, +\infty)$ *such that for all* $t \in [0,T]$

(4.22)

$$\omega_t^R(u,u) = 0 \text{ for every } u \in D_R,$$
$$\text{the map } (t,u,v) \mapsto \omega_t^R(u,v) \text{ is upper semicontinuous, and}$$
$$\mathcal{E}_t(v) - \mathcal{E}_t(u) - \langle \xi, v-u \rangle \ge -\omega_t^R(u,v)\|v-u\| \quad \forall\, u,\, v \in D_R \text{ and } \xi \in \partial\mathcal{E}_t(u).$$

It was shown in [51, sec. 2] that a sufficient condition for (4.22) is that the map $u \mapsto \mathcal{E}_t(u)$ is λ-*convex* uniformly in $t \in [0,T]$, namely

(4.23)

$$\exists\, \lambda \in \mathbb{R} \quad \forall\, t \in [0,T] \; \forall\, u_0,\, u_1 \in D \; \forall\, \theta \in [0,1] \; :$$
$$\mathcal{E}_t((1-\theta)u_0 + \theta u_1) \le (1-\theta)\mathcal{E}_t(u_0) + \theta\mathcal{E}_t(u_1) - \frac{\lambda}{2}\theta(1-\theta)\|u_0 - u_1\|^2 \, .$$

Suitable perturbations of λ-convex functionals also fulfill the closedness and the chain rule properties: we refer the reader to [67, 51, 66] for more details and explicit examples.

We have the following proposition.

PROPOSITION 4.11. *Let* $\mathcal{E} : [0,T]\times X \to (-\infty, +\infty]$ *fulfill* (3.$\mathcal{E}_0$), (3.$\mathcal{E}_2$), *and the* uniform subdifferentiability *condition* (4.22). *Then,* $\mathcal{E}$ *complies with the closedness condition* (3.$\mathcal{E}_3$), *with the chain rules* (3.19) *and* (3.$\mathcal{E}_4$), *and with property* (4.9).

Proof. In [51, Prop. 2.4], it was proved that condition (4.22) implies (3.$\mathcal{E}_3$) and (3.19). The validity of (3.$\mathcal{E}_4$) and (4.9) can be checked trivially adapting the arguments developed for the proof of [51, Prop. 2.4], to which the reader is referred. ∎

4.3. Examples of quasistatic limits. Our approach to the approximation of doubly nonlinear evolution equations in particular allows us to discuss quasistatic limits of dynamical problems. Indeed, the flexibility in the choice of the approximating graphs α_n, possibly noncyclic monotone, makes it possible to take rate-independent limits of Hamiltonian systems. We shall provide here some examples of ODEs and PDEs that can be reformulated within our framework.

Let us start by considering the case of a *nonlinearly damped oscillator*. In particular, let $q = q(t)$ represent the set of generalized coordinates of the system, M be the mass matrix, and $U = U(q)$ its smooth and coercive potential energy. Assume, moreover, that the system dissipates energy in terms of a positively 1-homogeneous and nondegenerate dissipation potential $D = D(\dot{q})$. By rescaling time t as εt, the quasistatic limit of the system corresponds to the limit as $\varepsilon \to 0$ in the equation

$$\varepsilon^2 M\ddot{q} + \partial D(\dot{q}) + \nabla U(q) \ni 0. \tag{4.24}$$

The latter can be rephrased as a single doubly nonlinear Hamiltonian system in the pair $v = (p, q)$, by introducing the Hamiltonian $H(p, q) = U(p) + q{\cdot}M^{-1}q/2$, the symplectic operator

$$J = \begin{pmatrix} 0 & 1 \\ -1 & 0 \end{pmatrix},$$

and the dissipation potential $\widehat{D}(\dot{p}, \dot{q}) = D(\dot{p})$. Then, (4.24) reads

$$\partial\widehat{D}(\dot{p}, \dot{q}) + \varepsilon J(\dot{p}, \dot{q}) + \nabla H(p, q) \ni (0, 0), \tag{4.25}$$

which can be equivalently rewritten as

$$\begin{aligned} \partial D(\dot{p}) + \varepsilon\dot{q} + \nabla U(p) \ni 0, \\ -\varepsilon\dot{p} + M^{-1}q = 0. \end{aligned}$$

Taking the quasistatic limit $\varepsilon \to 0$ in relation (4.25) requires dealing with the graphs $\alpha_\varepsilon = \partial\widehat{D} + \varepsilon J$, which are noncyclic monotone for all $\varepsilon > 0$. On the other hand, for $D(\dot{p}) = |\dot{p}|$ (for simplicity) the graph α_ε is represented, for instance, by

$$f_{\alpha_\varepsilon}(\dot{p}, \dot{q}, \dot{p}^*, \dot{q}^*) = \begin{cases} |\dot{p}| & \text{if } |\dot{p}^* - \varepsilon\dot{q}| \le 1,\ \dot{q}^* + \varepsilon\dot{p} = 0, \\ \infty & \text{else.} \end{cases}$$

Apart from the coercivity assumption (3.α_1) (which can, however, be relaxed in this case), this situation fits into our theory. In particular, solution trajectories to the dynamic problem (4.24) converge to solutions of the corresponding quasistatic limit. By generalizing the choice of the graphs α_ε, convergence can be obtained for a large class of different approximating problems.

The nonlinear oscillator example can be turned into a first PDE example by considering the nonlinearly damped semilinear wave equation

$$\varepsilon^2 u_{tt} + \partial D(u_t) - \Delta u + f(u) = 0. \tag{4.26}$$

This is to be posed in the cylinder $\Omega \times (0, T)$ for some smoothly bounded open set $\Omega \subset \mathbb{R}^n$, along with the positively 1-homogeneous and nondegenerate dissipation potential D, the smooth and polynomially bounded function f, and suitable initial and homogeneous Dirichlet boundary conditions (for simplicity). Equation (4.26) can be variationally reformulated in terms of a first-order system as

$$\partial\mathcal{D}(u_t, v_t) + \varepsilon\mathcal{J}(u_t, v_t) + \partial\mathcal{H}(u, v) \ni (0, 0) \quad \text{in } \mathcal{U}^*{\times}\mathcal{V}^* \quad \text{for a.a. } t \in (0, T), \tag{4.27}$$

where $\mathcal{U} = H^1_0(\Omega)$, $\mathcal{V} = L^2(\Omega)$, the functionals $\mathcal{D} : \mathcal{V}^2 \to [0, \infty]$, and $\mathcal{H} : \mathcal{U}{\times}\mathcal{V} \to (-\infty, \infty]$ are given by

$$\mathcal{D}(u_t, v_t) = \int_\Omega D(u_t)\,\mathrm{d}x, \quad \mathcal{H}(u, v) = \int_\Omega \left(\frac{1}{2}|\nabla u|^2 + \widehat{f}(u) + \frac{1}{2}|v|^2\right)\mathrm{d}x$$

for $\widehat{f}' = f$, and $\mathcal{J}(u_t, v_t)(x) = J(u_t(x), v_t(x))$ for a.e. $x \in \Omega$. Equation (4.27) fits in our framework along with the choice $\alpha_\varepsilon = \partial\mathcal{D} + \varepsilon\mathcal{J}$, which are noncyclic monotone for all $\varepsilon > 0$. Again, for $D(p_t) = |p_t|$ these graphs can be represented by

$$f_{\alpha_\varepsilon}(p_t, q_t, p^*_t, q^*_t) = \begin{cases} \|p_t\|_1 & \text{if } |p^*_t - \varepsilon q_t| \le 1,\ q^*_t + \varepsilon p_t = 0 \ \text{ a.e. in } \Omega, \\ \infty \text{ else,} \end{cases}$$

where $\|\cdot\|_1$ stands for the $L^1(\Omega)$-norm. In particular, owing to our analysis we can take the quasistatic limit $\varepsilon \to 0$ in the latter (again by suitably circumventing the lack of coercivity, which is inessential here).

Let us now provide a second PDE example by considering the quasistatic limit in *linearized elastoplasticity* with linear kinematic hardening [40]. We let $\Omega \subset \mathbb{R}^3$ be the reference configuration of an elastoplastic body which is subject to a displacement $u : \Omega \to \mathbb{R}^3$ and a *plastic strain* $p : \Omega \to \mathbb{R}^{3\times 3}_{\rm dev}$ (traceless or *deviatoric* symmetric $3{\times}3$ tensors). Then, the evolution of the elastoplastic medium is described by the system of the (time-rescaled) momentum balance (in $\mathbb{R}^3$) and constitutive equation (in $\mathbb{R}^{3\times 3}_{\rm dev}$) as

$$\begin{aligned}&\varepsilon^2 \rho u_{tt} - \nabla{\cdot}(\mathbb{C}(\varepsilon(u){-}p)) = b,\\ &\partial D(p_t) + \mathbb{H}p = \mathbb{C}(\varepsilon(u){-}p)\end{aligned}$$

in $\Omega \times (0,T)$, where $\rho = \rho(x)$ stands for the material density, $\mathbb{C}$ is the elasticity tensor (symmetric, positive definite), $\varepsilon(u) = (\nabla u{+}\nabla u^\top)/2$ is the symmetrized strain gradient, $b = b(t,x)$ denotes some body force density, $\mathbb{H}$ is the hardening tensor, and D is a positively 1-homogeneous and nondegenerate dissipation potential. The choice $D(p_t) = R|p_t|$ for some $R > 0$ corresponds to the classical *Von Mises* plasticity. We shall close the latter elastoplasticity system by imposing homogeneous Dirichlet conditions on u and no-traction conditions at the boundary (for simplicity). Then, the system can be recast in the form of a first-order system by augmenting the variables, including the momentum $v_t = \rho u_t$. In particular, we can variationally reformulate the system as

$$\begin{aligned}&\partial\mathcal{D}(u_t,v_t,p_t) + \varepsilon\mathcal{J}(u_t,v_t,p_t) + \partial\mathcal{H}(u,v,p) \ni (b,0,0)\\ &\qquad \text{in } \mathcal{U}^*{\times}\mathcal{V}^*{\times}\mathcal{P}^* \quad \text{for a.a. } t\in(0,T),\end{aligned} \tag{4.28}$$

where now the spaces are defined as $\mathcal{U} = \{u \in H^1(\Omega;\mathbb{R}^3) : u = 0 \text{ in } \partial\Omega\}$, $\mathcal{V} = L^2(\Omega;\mathbb{R}^3)$, $\mathcal{P} = L^2(\Omega;\mathbb{R}^{3\times3}_{\rm dev})$. The functionals and the operator are given by

$$\begin{aligned}\mathcal{D}(u_t,v_t,p_t) &= \int_\Omega D(p_t)\,\mathrm{d}x \quad \forall p_t \in L^1(\Omega;\mathbb{R}^{3\times3}_{\rm dev}),\\ \mathcal{H}(u,v,p) &= \int_\Omega \left(\frac12(\varepsilon(u){-}p){:}\mathbb{C}(\varepsilon(u){-}p) + \frac12 p{:}\mathbb{H}p + \frac{1}{2\rho}|v|^2\right)\mathrm{d}x\\ &\qquad \forall (u,v,p) \in \mathcal{U}{\times}\mathcal{V}{\times}\mathcal{P},\\ \mathcal{J}(u_t,v_t,p_t) &= \begin{pmatrix} v_t\\ -u_t\\ 0\end{pmatrix} \quad \forall (u,v,p)\in\mathcal{U}{\times}\mathcal{V}{\times}\mathcal{P}.\end{aligned}$$

Once again, the operators $\alpha_\varepsilon = \partial\mathcal{D} + \varepsilon\mathcal{J}$ are noncyclic monotone for all $\varepsilon > 0$ and can be represented by

$$f_{\alpha_\varepsilon}(u_t,v_t,p_t,u_t^*,v_t^*,p_t^*) = \begin{cases} \int_\Omega D(p_t) \text{ if } |u_t^* - \varepsilon v_t| \le 1,\ v_t^* + \varepsilon u_t = 0,\ p_t^* = 0 \ \text{ a.e. in } \Omega,\\ \infty \text{ else.}\end{cases}$$

In particular, our analysis is suited to analyzing the quasistatic limit $\varepsilon \to 0$ in the elastoplastic system (4.28). This clearly distinguishes our framework from the former variational principle from [74], which is of no use in the dynamical case.

5. Proof of Theorem 4.5. *Outline* Our starting point is the fact that the functions (u_n, ξ_n) fulfill for every $n \in \mathbb{N}$ the energy identity

$$\mathcal{E}_t(u_n(t)) + \int_0^t f_{\alpha_n}\left(\dot{u}_n(s), -\xi_n(s)\right) \mathrm{d}s = \mathcal{E}_0(u_0^n) + \int_0^t \partial_t \mathcal{E}_s(u_n(s))\, \mathrm{d}s \qquad \forall\, t \in (0,T]. \tag{5.1}$$

From (5.1), we will deduce a priori estimates on the sequence (u_n, ξ_n). Relying on well-known strong and weak compactness results, we will then prove the convergence (up to a subsequence) of (u_n, ξ_n) to a limit pair $(u, \hat{\xi})$. Hence we will pass to the limit as $n \to \infty$ in (5.1), following the lines of the proof of [51, Thm. 4.4]. Namely, we will combine the finite-dimensional lower semicontinuity theorem [35, Theorem 5.27] with tools from infinite-dimensional Young measure theory (see Appendix A for some basic recaps) and refined selection arguments mutuated from the proof of [51, Thm. 4.4]. Such arguments will yield the existence of a function $\xi \in L^1(0,T;X^*)$ such that the pair (u, ξ) fulfill the energy inequality (4.5).

Notation 5.1. Hereafter we will denote by the symbols C, C' various positive constants, which may changefrom line to line, depending only on known quantities and in particular independent of $n \in \mathbb{N}$. We will also use the place-holders

$$E_n(t) := \mathcal{E}_t(u_n(t)), \qquad P_n(t) := \partial_t \mathcal{E}_t(u_n(t)). \tag{5.2}$$

Step 1: A priori estimates and compactness. It follows from (5.1) and (3.$\mathcal{E}_2$) (cf. also estimate (3.22)) that $E_n(t) \leq E_n(0) + C_1 \int_0^t E_n(s)\, \mathrm{d}s$ for all $t \in [0,T]$. Since $\sup_{n \in \mathbb{N}} E_n(0) \leq C$ by (4.3), applying the Gronwall lemma we deduce $\sup_{t \in [0,T]} \{E_n(t) : t \in [0,T]\} \leq C$. Therefore, in view of assumption (3.$\mathcal{E}_2$) and property (3.16), we conclude that

$$\exists\, C > 0 \ \ \forall\, n \in \mathbb{N} : \quad \sup_{t \in [0,T]} \left(\mathcal{G}(u_n(t)) + |P_n(t)|\right) \leq C. \tag{5.3}$$

Thanks to (3.$\mathcal{E}_1$) we then infer that

$$\exists\, K \Subset X \quad \forall\, n \in \mathbb{N}\ \forall\, t \in [0,T] : \quad u_n(t) \in K. \tag{5.4}$$

Then, taking into account that $f_{\alpha_n}(\dot{u}_n, -\xi_n) \geq 0$ a.e. in $(0,T)$ in view of (2.9), (5.1) yields

$$\exists\, C > 0 \ \ \forall\, n \in \mathbb{N} : \qquad \|f_{\alpha_n}(\dot{u}_n, -\xi_n)\|_{L^1(0,T)} \leq C. \tag{5.5}$$

In view of assumption (3.α_1), from (5.5) we conclude

$$\int_0^T c_1 \|\dot{u}_n(s)\| + c_2 \|\xi_n(s)\|_*^q \, \mathrm{d}s \leq C.$$

Also, due to (5.4), we ultimately deduce that

$$\exists\, C > 0 \ \ \forall\, n \in \mathbb{N} : \quad \|u_n\|_{BV([0,T];X)} + \|\xi_n(s)\|_{L^q(0,T;X^*)} \leq C\,. \tag{5.6}$$

Furthermore, from the energy identity (5.1) we immediately infer that, setting $h_n(t) := E_n(t) - \int_0^t P_n(s)\, \mathrm{d}s$, there holds

$$h_n(t) - h_n(s) = -\int_s^t f_{\alpha_n}(\dot{u}_n(r), -\xi_n(r))\, \mathrm{d}r \leq 0 \quad \forall\, 0 \leq s \leq t \leq T.$$

Therefore, we have $\mathrm{Var}(h_n;[0,T]) = E_n(0) - E_n(T) + \int_0^T P_n(s)\,\mathrm{d}s \le C$ thanks to (5.3) and (4.3). Since (P_n) is uniformly bounded in $L^\infty(0,T)$, we conclude that

$$\exists\, C>0 \ \ \forall\, n\in\mathbb{N}: \quad \mathrm{Var}(E_n;[0,T]) \le C\,. \tag{5.7}$$

Estimates (5.4), (5.6), and (5.7) and the Helly principle (see, e.g., [43, Thm. 3.1]) guarantee that there exist a subsequence (n_k) and functions $u \in BV([0,T];X)$ and $E \in BV([0,T])$ such that, as $k\to\infty$,

$$(u_{n_k}(t), \mathcal{E}_t\,(u_{n_k}(t))) \to (u(t),E(t)) \ \text{ in } X\times\mathbb{R}\ \forall\, t\in[0,T], \tag{5.8}$$

$$\mathrm{d}u_{n_k} = \dot{u}_{n_k}\cdot\mathcal{L} \overset{*}{\rightharpoonup} \mathrm{d}u \ \text{ in } \mathcal{M}(0,T;X), \tag{5.9}$$

where the last convergence is classically defined as

$$\mu_k \overset{*}{\rightharpoonup} \mu \ \text{ in } \mathcal{M}(0,T;X) \quad\Longleftrightarrow\quad \int_0^T \langle\phi(t),\mathrm{d}\mu_k(t)\rangle \to \int_0^T\langle\phi(t),\mathrm{d}\mu(t)\rangle \quad \forall\,\phi\in C^0([0,T];X^*).$$

Exploiting Theorem 3.4, we decompose $\mathrm{d}u$ as

$$\mathrm{d}u = (\mathrm{d}u)_{\mathrm{ac}} + (\mathrm{d}u)_{\mathrm{sin}} = \dot{u}_{\mathrm{ac}}\,\mathcal{L} + \dot{u}_s\,\|(\mathrm{d}u)_{\mathrm{sin}}\|.$$

Observe that, by the lower semicontinuity $(3.\mathcal{E}_0)$,

$$E(t) \ge \mathcal{E}_t(u(t)) \qquad \forall\, t\in[0,T]. \tag{5.10}$$

Further, in view of estimate (5.5), there exists $\mu\in\mathcal{M}(0,T)$ such that (up to a not relabeled subsequence)

$$f_{\alpha_k}\,(\dot{u}_{n_k}(\cdot), -\xi_{n_k}(\cdot))\cdot\mathcal{L} \overset{*}{\rightharpoonup} \mu \ \text{ in } \mathcal{M}(0,T). \tag{5.11}$$

Moreover, by an infinite-dimensional version of the fundamental compactness theorem of Young measure theory (cf. Theorem A.3 in Appendix A), we can associate with (possibly a subsequence of) (ξ_{n_k}, P_{n_k}) a limiting Young measure $(\sigma_t)_{t\in(0,T)} \in \mathscr{Y}(0,T;X\times\mathbb{R})$ such that, for almost all $t\in(0,T)$ it holds that $\sigma_t(X\times\mathbb{R}) = 1$ and σ_t is supported on the set of the limit points of $(\xi_{n_k}(t), P_{n_k}(t))$ w.r.t. the weak topology on $X^*\times\mathbb{R}$, viz.

$$\mathrm{supp}\,(\sigma_t) \subset \bigcap_{j\in\mathbb{N}} \overline{\{(\xi_{n_k}(t), P_{n_k}(t)) : k\ge j\}}^{\mathrm{weak}} \tag{5.12}$$

(where by $\overline{B}^{\mathrm{weak}}$ we denote the closure of a set $B\subset X^*\times\mathbb{R}$ w.r.t. the weak topology). Furthermore, it holds that

$$\xi_{n_k} \rightharpoonup \int_{X^*\times\mathbb{R}} \zeta\,\mathrm{d}\sigma_t(\zeta,p) =: \hat{\xi} \quad \text{ in } L^q(0,T;X^*) \ \text{ and} \tag{5.13}$$

$$P_{n_k} \overset{*}{\rightharpoonup} \int_{X^*\times\mathbb{R}} p\,\mathrm{d}\sigma_t(\zeta,p) =: \hat{P} \quad \text{ in } L^\infty(0,T)\,. \tag{5.14}$$

Step 2: Nonemptyness of admissible sets. From now on, for the sake of simplicity, we shall write k instead of n_k. There exists a negligible set $N\subset(0,T)$ such that for every $t\in(0,T)\setminus N$ convergences (5.8) and the support property (5.12) hold. Taking into account the closedness condition $(3.\mathcal{E}_3)$, it can be easily checked (cf. also [51, sec. 6]) that for almost all $t \in (0,T)$ there holds

$$
\begin{array}{lll}
(t,u(t)) & \in & D(\partial\mathcal{E}), \\
\mathcal{E}_t(u(t)) & = & E(t), \qquad \mathcal{E}_0(u(0)) = E(0), \\
\mathrm{supp}\,(\sigma_t) & \subset & \{(\zeta,p) \in X^* \times \mathbb{R} : \zeta \in \partial\mathcal{E}_t(u(t))\,,\ p \le \partial_t\mathcal{E}_t(u(t))\}\,.
\end{array} \tag{5.15}
$$

In particular, from (5.14) and the third equation of (5.15) it follows that

$$
\hat{P}(t) \le \partial_t\mathcal{E}_t(u(t)) \qquad \text{for a.a. } t \in (0,T)\,. \tag{5.16}
$$

Step 3: lim inf result for the Fitzpatrick function. In this section, we are going to prove that

$$
\begin{aligned}
&\liminf_{k\to\infty} \int_0^t f_{\alpha_k}\left(\dot{u}_k(r), -\xi_k(r)\right)\,\mathrm{d}r \\
&\quad \ge \int_0^t \int_{X^*\times\mathbb{R}} f_\alpha\left(\dot{u}_{\mathrm{ac}}(r), -\zeta\right)\,\mathrm{d}\sigma_r(\zeta,p)\,\mathrm{d}r + \int_0^t f_\alpha^\infty\left(\dot{u}_s(r), 0\right)\|(\mathrm{d}u)_{\mathrm{sin}}\|(r)\,.
\end{aligned} \tag{5.17}
$$

In order to do so, employing [35, Corollary 1.116, p. 75], we decompose the measure μ from (5.11) as follows: there exist μ_{ac}, μ_{sin}, $\mu_\perp$ in $\mathcal{M}(0,T)$ such that

$$
\begin{gathered}
\mu_{\mathrm{ac}} \ll \|(\mathrm{d}u)_{\mathrm{ac}}\|\,,\ \mu_{\mathrm{sin}} \ll \|(\mathrm{d}u)_{\mathrm{sin}}\|\,,\ \mu_\perp \perp \|(\mathrm{d}u)_{\mathrm{ac}}\| + \|(\mathrm{d}u)_{\mathrm{sin}}\|, \text{ and} \\
\mu = \mu_{\mathrm{ac}} + \mu_{\mathrm{sin}} + \mu_\perp.
\end{gathered} \tag{5.18}
$$

In particular, μ_{ac} is absolutely continuous w.r.t. the Lebesgue measure $\mathcal{L}$. Since $f_{\alpha_k}(\dot{u}_k, -\xi_k) \ge 0$ a.e. in $(0,T)$, we obtain $\mu_\perp \ge 0$. We will split the proof of (5.17) into two steps.

First step. Now, it follows from (3.8) and (5.18) and the Radon–Nikodým property of X that the set of the points $t_0 \in (0,T)$ such that $\sigma_{t_0}(X^* \times \mathbb{R}) = 1$ and

$$
\begin{array}{lll}
\dot{u}_{\mathrm{ac}}(t_0) & = & \displaystyle\lim_{\varepsilon\to0} \frac{(\mathrm{d}u)\left([t_0-\varepsilon, t_0+\varepsilon]\cap[0,T]\right)}{2\varepsilon}\,, \\
\hat{\xi}(t_0) & = & \displaystyle\lim_{\varepsilon\to0} \frac{1}{\varepsilon}\int_{t_0-\varepsilon/2}^{t_0+\varepsilon/2} \hat{\xi}(t)\,\mathrm{d}t, \text{ and} \\
\displaystyle\frac{\mathrm{d}\mu_{\mathrm{ac}}}{\mathrm{d}\mathcal{L}}(t_0) & = & \displaystyle\lim_{\varepsilon\to0} \frac{\mu\left([t_0-\varepsilon, t_0+\varepsilon]\cap[0,T]\right)}{2\varepsilon} < \infty
\end{array} \tag{5.19}
$$

has full Lebesgue measure. From now on, we shall use the notation

$$
Q_\varepsilon(t_0) := [t_0-\varepsilon, t_0+\varepsilon]\cap[0,T] \qquad \text{with } t_0 \in (0,T) \text{ such that (5.19) holds.} \tag{5.20}
$$

We then prove that

$$
\frac{\mathrm{d}\mu_{\mathrm{ac}}}{\mathrm{d}\mathcal{L}}(t_0) \ge \int_{X^*\times\mathbb{R}} f_\alpha\left(\dot{u}_{\mathrm{ac}}(t_0), -\zeta\right)\,\mathrm{d}\sigma_{t_0}(\zeta,p) \tag{5.21}
$$

with $t_0 \in (0,T)$ such that (5.19) holds. For any such t_0, it is also possible to choose a vanishing sequence $(\varepsilon_m)_m$ such that for all $m \in \mathbb{N}$ there holds

$$
\mu\left(\{t_0-\varepsilon_m, t_0+\varepsilon_m\}\cap[0,T]\right) = (\mathrm{d}u)\left(\{t_0-\varepsilon_m, t_0+\varepsilon_m\}\cap[0,T]\right) = 0. \tag{5.22}
$$

The latter will be exploited in combination with (5.19) to establish (5.26) later. In order to show (5.21), we will use (2.7), which yields

$$
f_\alpha\left(\dot{u}_{\mathrm{ac}}(t_0), -\zeta\right) = \sup\{\langle x^*, \dot{u}_{\mathrm{ac}}(t_0)\rangle - \langle\zeta, x\rangle - \rho_{\alpha^{-1}}(x^*, x)\ :\ (x,x^*) \in X\times X^*\}\,. \tag{5.23}
$$

In view of (5.23), we thus confine ourselves to showing that

$$\frac{\mathrm{d}\mu_{\mathrm{ac}}}{\mathrm{d}\mathcal{L}}(t_0) \geq \langle x^*, \dot{u}_{\mathrm{ac}}(t_0)\rangle - \int_{X^*\times\mathbb{R}} \langle \zeta, x\rangle \,\mathrm{d}\sigma_{t_0}(\zeta,p) - \rho_{\alpha^{-1}}(x,x^*) \quad \forall\,(x,x^*)\in X\times X^* \text{ with } \rho_{\alpha^{-1}}(x,x^*)<\infty, \tag{5.24}$$

for relation (5.21) then follows by Fatou's lemma. Now, to check (5.24) we observe that, since $\alpha_k \xrightarrow{\mathrm{g}} \alpha$ and $\alpha_k^{-1} \xrightarrow{\mathrm{g}} \alpha^{-1}$ in the graph sense, we can apply Theorem 2.10 to $\rho_{\alpha_k^{-1}} = f^*_{\alpha_k}$. Therefore, for any $(x,x^*) \in X\times X^*$ there exists a sequence $(x_k,x_k^*)_{k\in\mathbb{N}}$ such that

$$(x_k,x_k^*) \to (x,x^*) \text{ and } \limsup_{n\to\infty} \rho_{\alpha_k^{-1}}(x_k^*,x_k) \leq \rho_{\alpha^{-1}}(x^*,x). \tag{5.25}$$

Combining (5.11) with the third equation of (5.19) (for the sequence $(\varepsilon_m)_m$ fulfilling (5.22)), we have that

$$\begin{aligned} &\frac{\mathrm{d}\mu_{\mathrm{ac}}}{\mathrm{d}\mathcal{L}}(t_0) \\ &= \lim_{m\to\infty}\lim_{k\to\infty} \frac{1}{2\varepsilon_m}\int_{Q_{\varepsilon_m}(t_0)} f_{\alpha_k}(\dot{u}_k(t),-\xi_k(t))\,\mathrm{d}t \\ &\geq \liminf_{m\to\infty}\liminf_{k\to\infty} \frac{1}{2\varepsilon_m}\int_{Q_{\varepsilon_m}(t_0)} \Big(\langle x_k^*, \dot{u}_k(t)\rangle + \langle -\xi_k(t), x_k\rangle - \rho_{\alpha_k^{-1}}(x_k^*,x_k)\Big)\,\mathrm{d}t, \end{aligned} \tag{5.26}$$

where in the latter inequality we have plugged in the sequence (x_k,x_k^*) from (5.25) and applied formula (5.23) for f_{α_k}. On account of convergences (5.9) and (5.13), and of the fact that $(x_k,x_k^*)\to(x,x^*)$, we have for every $m\in\mathbb{N}$

$$\int_{Q_{\varepsilon_m}(t_0)} \langle x_k^*, \dot{u}_k(t)\rangle\,\mathrm{d}t \to \int_{Q_{\varepsilon_m}(t_0)} \langle x^*, \,\mathrm{d}u(t)\rangle\,. \tag{5.27}$$

$$\begin{aligned} \int_{Q_{\varepsilon_m}(t_0)} \langle \xi_k(t), x_k\rangle\,\mathrm{d}t &\to \int_{Q_{\varepsilon_m}(t_0)} \Big\langle \hat{\xi}(t), x\Big\rangle\,\mathrm{d}t \\ &= \int_{Q_{\varepsilon_m}(t_0)} \left(\int_{X^*\times\mathbb{R}} \langle\zeta, x\rangle\,\mathrm{d}\sigma_t(\zeta,p)\right)\mathrm{d}t. \end{aligned} \tag{5.28}$$

Inserting (5.27)–(5.28) into (5.26) and using (5.25), we thus get

$$\begin{aligned} &\frac{\mathrm{d}\mu_{\mathrm{ac}}}{\mathrm{d}\mathcal{L}}(t_0) \\ &\geq \liminf_{m\to\infty} \frac{1}{2\varepsilon_m}\left(\int_{Q_{\varepsilon_m}(t_0)} \langle x^*, \,\mathrm{d}u(t)\rangle \right. \\ &\left. + \int_{Q_{\varepsilon_m}(t_0)} \left(\int_{X^*\times\mathbb{R}} \langle -\zeta, x\rangle\,\mathrm{d}\sigma_t(\zeta,p) - \rho_{\alpha^{-1}}(x,x^*)\right)\mathrm{d}t\right), \end{aligned}$$

and in view of (5.19) we infer (5.24), whence the desired (5.21).

Second step. Choose $t_0\in(0,T)$ such that it satisfies

$$\begin{aligned} \dot{u}_{\mathrm{sin}}(t_0) &= \lim_{\varepsilon\to 0}\frac{(\mathrm{d}u)\,([t_0-\varepsilon,t_0+\varepsilon]\cap[0,T])}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,([t_0-\varepsilon,t_0+\varepsilon]\cap[0,T])}\,, \\ 0 &= \lim_{\varepsilon\to 0}\frac{\mathcal{L}\,([t_0-\varepsilon,t_0+\varepsilon]\cap[0,T])}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,([t_0-\varepsilon,t_0+\varepsilon]\cap[0,T])}, \quad \text{and} \\ \frac{\mathrm{d}\mu_{\mathrm{sin}}}{\|(\mathrm{d}u)_{\mathrm{sin}}\|}(t_0) &= \lim_{\varepsilon\to 0}\frac{\mu\,([t_0-\varepsilon,t_0+\varepsilon]\cap[0,T])}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,([t_0-\varepsilon,t_0+\varepsilon]\cap[0,T])} < \infty\,. \end{aligned} \tag{5.29}$$

The set of all t_0 failing any of (5.29) is a $\|(\mathrm{d}u)_{\mathrm{sin}}\|$-null set. We are now going to prove that

$$\frac{\mathrm{d}\mu_{\mathrm{sin}}}{\|(\mathrm{d}u)_{\mathrm{sin}}\|}(t_0) \geq f_\alpha^\infty(\dot{u}_{\mathrm{sin}}(t_0), 0) \tag{5.30}$$

for any $t_0 \in (0,T)$ complying with (5.29). As before, we will use the notation of (5.20) for the set $Q_\varepsilon(t_0)$ with any such t_0, and we choose correspondingly a vanishing sequence (ε_m) such that (5.22) is satisfied. In order to show (5.30), in view of the representation formula (3.13) for f_α^∞ it is sufficient to show that

$$\frac{\mathrm{d}\mu_{\mathrm{sin}}}{\|(\mathrm{d}u)_{\mathrm{sin}}\|}(t_0) \geq \langle x^*, \dot{u}_{\mathrm{sin}}(t_0)\rangle \ \ \forall\,(x,x^*) \text{ such that } \rho_{\alpha^{-1}}(x,x^*) < \infty. \tag{5.31}$$

With the same argument as in the previous lines, we see that

$$\begin{aligned}
&\frac{\mathrm{d}\mu_{\mathrm{sin}}}{\|(\mathrm{d}u)_{\mathrm{sin}}\|}(t_0)\\
&= \lim_{m\to\infty}\lim_{k\to\infty} \frac{1}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,(Q_{\varepsilon_m}(t_0))} \int_{Q_{\varepsilon_m}(t_0)} f_{\alpha_k}(\dot{u}_k(t), -\xi_k(t))\,\mathrm{d}t\\
&\geq \liminf_{m\to\infty}\liminf_{k\to\infty} \frac{1}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,(Q_{\varepsilon_m}(t_0))} \int_{Q_{\varepsilon_m}(t_0)} \Big(\langle x_k^*, \dot{u}_k(t)\rangle + \langle -\xi_k(t), x_k\rangle - \rho_{\alpha_k^{-1}}(x_k^*, x_k)\Big)\,\mathrm{d}t,
\end{aligned} \tag{5.32}$$

where (x_k, x_k^*) as in (5.25) approximates (x,x^*) from (5.31). Once again, due to (5.9) and (5.13), we have for every fixed $m \in \mathbb{N}$ that

$$\begin{aligned}
&\int_{Q_{\varepsilon_m}(t_0)} \langle x_k^*, \dot{u}_k(t)\rangle\ \mathrm{d}t \to \int_{Q_{\varepsilon_m}(t_0)} \langle x^*,\, \mathrm{d}u(t)\rangle \text{ and}\\
&\int_{Q_{\varepsilon_m}(t_0)} \langle \xi_k(t), x_k\rangle\ \mathrm{d}t \to \int_{Q_{\varepsilon_m}(t_0)} \Big\langle \hat{\xi}(t), x\Big\rangle\ \mathrm{d}t.
\end{aligned}$$

By construction (cf. (5.29)), there holds

$$\begin{aligned}
&\lim_{m\to\infty} \frac{1}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,(Q_{\varepsilon_m}(t_0))} \int_{Q_{\varepsilon_m}(t_0)} \langle x^*,\, \mathrm{d}u(t)\rangle = \langle x^*, \dot{u}_{\mathrm{sin}}(t_0)\rangle\ ,\\
&\lim_{m\to\infty} \frac{1}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,(Q_{\varepsilon_m}(t_0))} \int_{Q_{\varepsilon_m}(t_0)} \Big\langle -\hat{\xi}(t), x\Big\rangle\ \mathrm{d}t = 0,\\
&\lim_{m\to\infty} \frac{\mathcal{L}(Q_{\varepsilon_m}(t_0))}{\|(\mathrm{d}u)_{\mathrm{sin}}\|\,(Q_{\varepsilon_m}(t_0))} \rho_{\alpha^{-1}}(x,x^*) = 0.
\end{aligned}$$

We thus conclude (5.31), whence (5.30).

In conclusion, passing to the limit as $n_k \to \infty$ in (5.1) and relying on the initial data convergence (4.3), the energy convergence (5.8) joint with (5.10), (5.14), and the lower semicontinuity (5.17), we have obtained

$$\begin{aligned}
&\mathcal{E}_t(u(t)) + \int_0^t \int_{X^*\times\mathbb{R}} f_\alpha\left(\dot{u}_{\mathrm{ac}}(s), -\zeta\right)\ \mathrm{d}\sigma_s(\zeta,p)\,\mathrm{d}s + \int_0^t f_\alpha^\infty\left(\dot{u}_{\mathrm{sin}}(s), 0\right) \|(\mathrm{d}u)_{\mathrm{sin}}\|(s)\\
&\leq E(t) + \int_0^t \int_{X^*\times\mathbb{R}} f_\alpha\left(\dot{u}_{\mathrm{ac}}(s), -\zeta\right)\ \mathrm{d}\sigma_s(\zeta,p)\,\mathrm{d}s + \int_0^t f_\alpha^\infty\left(\dot{u}_{\mathrm{sin}}(s), 0\right) \|(\mathrm{d}u)_{\mathrm{sin}}\|(s)\\
&\leq \mathcal{E}_0(u(0)) + \int_0^t \int_{X^*\times\mathbb{R}} p\,\mathrm{d}\sigma_s(\zeta,p)\,\mathrm{d}s \qquad \forall\, t \in (0,T]\,.
\end{aligned} \tag{5.33}$$

Step 4: Enhanced support properties of the Young measure $(\sigma_t)_{t\in(0,T)}$. We can now improve the third equation of (5.15), showing that indeed

$$\operatorname{supp}(\sigma_t) \subset \{(\zeta,p)\in X^*\times\mathbb{R} : \zeta\in\partial\mathcal{E}_t(u(t))\,,\ -\zeta\in\alpha(u(t))\,,\ p\leq\partial_t\mathcal{E}_t(u(t))\} \quad \text{for a.a. } t\in(0,T). \tag{5.34}$$

To this end, observe that, passing to the limit as $n_k\to\infty$ in (5.1) (written on the interval (s,t)) yields, in view of convergences (5.8), (5.11), and (5.14), the following energy identity:

$$E(t)+\mu([s,t]) = E(s)+\int_s^t p(r)\,\mathrm{d}r \qquad \forall\, 0\leq s\leq t\leq T \qquad \text{with} \tag{5.35a}$$

$$\mu([s,t]) \geq \int_s^t \int_{X^*\times\mathbb{R}} f_\alpha\left(\dot u_{\mathrm{ac}}(r),-\zeta\right)\,\mathrm{d}\sigma_r(\zeta,p)\,\mathrm{d}r, \tag{5.35b}$$

the latter inequality due to (5.21). In particular, observe that

$$\int_0^T \int_{X^*\times\mathbb{R}} f_\alpha\left(\dot u_{\mathrm{ac}}(t),-\zeta\right)\,\mathrm{d}\sigma_t(\zeta,p)\,\mathrm{d}t<\infty. \tag{5.36}$$

Let $\mathcal{T}\subset[0,T]$ be the set of all Lebesgue points t_0 of $\hat P$ (5.14), such that relations (5.19) and (5.21) hold, and

$$\dot E_{\mathrm{ac}}(t_0) = \lim_{\varepsilon\to0}\frac{E\left(t_0+\frac{\varepsilon}{2}\right)-E\left(t_0-\frac{\varepsilon}{2}\right)}{\varepsilon}\,.$$

Then $\mathcal{T}$ has full measure. Let us now choose a sequence (ε_m), $\varepsilon_m\downarrow0$, such that (5.22) holds. Then, passing to the limit as $\varepsilon_m\downarrow0$ in (5.35) (written for $s=t_0-\varepsilon_m/2$ and $t=t_0+\varepsilon_m/2$), we obtain

$$\dot E_{\mathrm{ac}}(t_0)+\int_{X^*\times\mathbb{R}} f_\alpha(\dot u_{\mathrm{ac}}(t_0),-\zeta)\,\mathrm{d}\sigma_{t_0}(\zeta,p) \leq \int_{X^*\times\mathbb{R}} p\,\mathrm{d}\sigma_{t_0}(\zeta,p) \quad \forall\, t_0\in\mathcal{T} \tag{5.37}$$

(up to removing from $\mathcal{T}$ a set of zero Lebesgue measure). Now, observe that thanks to the third of (5.15) and (5.36), the Young measure $(\sigma_t)_{t\in(0,T)}$ satisfies the assumptions of the forthcoming Lemma A.4. Therefore, in view of the Young measure version of the chain rule inequality $(3.\mathcal{E}_4)$ therein, we find that

$$-\dot E_{\mathrm{ac}}(t_0)+\int_{X^*\times\mathbb{R}} p\,\mathrm{d}\sigma_{t_0}(\zeta,p) \leq \int_{X^*\times\mathbb{R}} \langle-\zeta,\dot u_{\mathrm{ac}}(t_0)\rangle\,\mathrm{d}\sigma_{t_0}(\zeta,p) \quad \text{for a.a. } t_0\in\mathcal{T}. \tag{5.38}$$

Combining (5.37) and (5.38), we deduce that for almost all $t_0\in\mathcal{T}$ (and hence for almost all $t_0\in(0,T)$) it holds that

$$\int_{X^*\times\mathbb{R}} \left(f_\alpha(\dot u_{\mathrm{ac}}(t_0),-\zeta)-\langle-\zeta,\dot u_{\mathrm{ac}}(t_0)\rangle\right)\mathrm{d}\sigma_{t_0}(\zeta,p)\leq0. \tag{5.39}$$

Since f_α is a representative function for α, we easily see that (5.39) holds as an equality, and that in fact

$$-\zeta\in\alpha(\dot u_{\mathrm{ac}}(t_0)) \qquad \text{for } \sigma_{t_0}\text{-a.a. } (\zeta,p)\in\operatorname{supp}(\sigma_{t_0}).$$

Since $t_0 \in (0,T)$ is arbitrary out of a Lebesgue-null set, we have ultimately proved the desired support property (5.34). Furthermore, as a by-product of (5.37)–(5.39) holding as equalities, we infer the following pointwise energy equality:

$$\dot{E}_{\rm ac}(t) + \int_{X^*\times\mathbb{R}} f_\alpha(u_{\rm ac}(t), -\zeta)\,\mathrm{d}\sigma_t(\zeta,p) = \int_{X^*\times\mathbb{R}} p\,\mathrm{d}\sigma_t(\zeta,p) \qquad \text{for a.a. } t\in(0,T). \tag{5.40}$$

Step 5: Selection argument and conclusion of the proof. We can now apply Lemma A.5 and deduce that there exist measurable functions $\xi:(0,T)\to X$ and $P:(0,T)\to\mathbb{R}$ such that

$$(\xi(t),P(t)) \in \operatorname{argmin}\left\{f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)-p : (\zeta,p)\in \mathcal{S}\left(t,u(t),\dot{u}_{\rm ac}(t)\right)\right\} \;\; \text{for a.a. } t\in(0,T), \tag{5.41}$$

with $\mathcal{S}(t,u(t),\dot{u}_{\rm ac}(t)) := \{(\zeta,p)\in X^*\times\mathbb{R} : \zeta\in\partial\mathcal{E}_t(u(t)),\ -\zeta\in\alpha(\dot{u}_{\rm ac}(t)),\ p\le \partial_t\mathcal{E}_t(u(t))\}$. In particular,

$$\xi(t)\in\partial\mathcal{E}_t(u(t))\,, \qquad -\xi(t)\in\alpha(\dot{u}_{\rm ac}(t)), \qquad \text{and } P(t)\le\partial_t\mathcal{E}_t(u(t)) \quad \text{for a.a. } t\in(0,T). \tag{5.42}$$

We then have the following chain of inequalities for almost all $t\in(0,T)$:

$$\begin{aligned} -\dot{E}_{\rm ac}(t) &= \int_{X^*\times\mathbb{R}} \left(f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)-p\right)\,\mathrm{d}\sigma_t(\zeta,p) \\ &\ge f_\alpha(\dot{u}_{\rm ac}(t),-\xi(t)) - P(t) \\ &\ge \langle -\xi(t),\dot{u}_{\rm ac}(t)\rangle - \partial_t\mathcal{E}_t(u(t)) \ge -\dot{E}_{\rm ac}(t), \end{aligned} \tag{5.43}$$

where the first identity follows from (5.40), the second from (5.41), the third from the fact that f_α is a representative function for α and from (5.42), and the last from the chain rule inequality (3.$\mathcal{E}_4$). Therefore, we infer that all inequalities in (5.43) hold as equalities, which proves (4.8). In particular, from (5.43) we derive

$$\int_{X^*\times\mathbb{R}} f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)\,\mathrm{d}\sigma_t(\zeta,p) - \langle-\xi(t),\dot{u}_{\rm ac}(t)\rangle = \int_{X^*\times\mathbb{R}} p\,\mathrm{d}\sigma_t(\zeta,p) - \partial_t\mathcal{E}_t(u(t)).$$

Since the left-hand side is nonnegative and the right-hand side is nonpositive (cf. (5.41)), we conclude that both sides are equal to zero. With the same argument, we ultimately infer that for almost all $t\in(0,T)$

$$\begin{aligned} P(t) &= \int_{X^*\times\mathbb{R}} p\,\mathrm{d}\sigma_t(\zeta,p) = \partial_t\mathcal{E}_t(u(t)), \\ f_\alpha(\dot{u}_{\rm ac}(t),-\xi(t)) &= \int_{X^*\times\mathbb{R}} f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)\,\mathrm{d}\sigma_t(\zeta,p) = \langle-\xi(t),\dot{u}_{\rm ac}(t)\rangle\,. \end{aligned} \tag{5.44}$$

Combining (5.44) with (5.33), we ultimately deduce (4.5). Now, from (4.5) with (4.6) we have that $\int_0^T |\langle-\xi(t),\dot{u}_{\rm ac}(t)\rangle|\,\mathrm{d}t<\infty$. Since α complies with (4.1), we then conclude that $\xi\in L^q(0,T;X^*)$. This completes the proof. ☐

Remark 5.2 (the role of the Fitzpatrick function). As pointed out in Remark 3.13, the variational reformulation of the doubly nonlinear differential inclusion (3.17) could be given in terms of *any* representative functional for α. The distinguished role of the Fitzpatrick function f_α is apparent in the passage to the limit argument developed in

Step 3 of the proof of Theorem 4.5. Therein (cf. (5.21)–(5.28)), we exploit the duality formula (2.7) for f_α, as well as Theorem 2.10.

Remark 5.3 (refinement of the measurable selection argument). A close perusal of Step 5 in the above proof reveals that, in principle, it should be sufficient to select the functions $t \mapsto (\xi(t), p(t))$ in the set $\tilde{\mathcal{S}}(t, u(t)) := \{(\zeta, p) \in X^* \times \mathbb{R} : \zeta \in \partial\mathcal{E}_t(u(t)),\ p \leq \partial_t\mathcal{E}_t(u(t))\}$, i.e., dropping the requirement $-\zeta \in \alpha(\dot{u}_{\mathrm{ac}}(t))$. Indeed, if we were in the position of applying Lemma A.5 to the set $\tilde{\mathcal{S}}$, from the chain of inequalities (5.43) the second equation of (5.44) would still follow, yielding $-\xi(t) \in \alpha(\dot{u}_{\mathrm{ac}}(t))$ for almost all $t \in (0, T)$, i.e., (4.6).

Nonetheless, the extension of Lemma A.5 to the set $\tilde{\mathcal{S}}$ seems to be an open problem at the moment; cf. the upcoming Remark A.6.

We conclude this section with the proof of Theorem 4.8.

Proof of Theorem 4.8. Repeating the calculations from Step 1 of the proof of Theorem 4.5, we prove that the sequence (u_n) is bounded in $W^{1,p}(0, T; X)$ and in addition fulfills estimate (5.3). Therefore, convergence (4.16) holds. We use the arguments from the above Steps 1 and 4 to infer that there exist $(\xi, E) \in L^1(0, T; X^*) \times BV([0, T])$ complying with (4.7) and (4.8). Since $u \in W^{1,p}(0, T; X)$, Proposition 4.7 applies, and we conclude the proof. ☐

Appendix A. Young measure results. We fix here some definitions and results on parameterized (or Young) measures (see, e.g., [12, 13, 14, 78]) with values in a reflexive Banach space $\mathcal{Y}$. In particular, in section 5 the upcoming results are applied to the space $\mathcal{Y} = X^* \times \mathbb{R}$.

Notation A.1. In what follows, we will denote by $\mathscr{L}_{(0,T)}$ the σ-algebra of the Lebesgue measurable subsets of $(0, T)$ and by $\mathscr{B}(\mathcal{Y})$ the Borel σ-algebra of $\mathcal{Y}$. We use the symbol $\otimes$ for product σ-algebrae. We recall that an $\mathscr{L}_{(0,T)} \otimes \mathscr{B}(\mathcal{Y})$-measurable function $h : (0, T) \times \mathcal{Y} \to (-\infty, +\infty]$ is a *normal integrand* if for almost all $t \in (0, T)$ the map $y \mapsto h_t(y) = h(t, y)$ is lower semicontinuous on $\mathcal{Y}$.

We consider the space $\mathcal{Y}$ endowed with the *weak* topology, and say that an $\mathscr{L}_{(0,T)} \otimes \mathscr{B}(\mathcal{Y})$-measurable functional $h : (0, T) \times \mathcal{Y} \to (-\infty, +\infty]$ is a *weakly normal integrand* if for almost all $t \in (0, T)$ the map

$$y \mapsto h(t, y) \text{ is sequentially l.s.c. on } \mathcal{Y} \text{ w.r.t. the weak topology.} \tag{A.1}$$

We denote by $\mathscr{M}(0, T; \mathcal{Y})$ the set of all $\mathscr{L}_{(0,T)}$-measurable functions $y : (0, T) \to \mathcal{Y}$. A sequence $(y_n) \subset \mathscr{M}(0, T; \mathcal{Y})$ is said to be *weakly tight* if there exists a weakly normal integrand $h : (0, T) \times \mathcal{Y} \to (-\infty, +\infty]$ such that the map

$$y \mapsto h_t(y) \text{ has compact sublevels w.r.t. the weak topology of } \mathcal{Y} \text{ for a.e. } t \in (0, T), \text{ and}$$
$$\sup_n \int_0^T h(t, y_n(t))\,\mathrm{d}t < \infty.$$

Definition A.2 (Young measures with values in $\mathcal{Y}$). *A (time-dependent)* Young measure *in the space $\mathcal{Y}$ is a family $\boldsymbol{\sigma} := \{\sigma_t\}_{t\in(0,T)}$ of Borel probability measures on $\mathcal{Y}$ parameterized by $t \in (0, T)$, such that the map on $(0, T)$*

$$t \mapsto \sigma_t(B) \quad \textit{is} \quad \mathscr{L}_{(0,T)}\textit{-measurable} \quad \forall\, B \in \mathscr{B}(\mathcal{Y}). \tag{A.2}$$

We denote by $\mathscr{Y}(0, T; \mathcal{Y})$ the set of all Young measures in $\mathcal{Y}$.

The following result is taken from [51] (cf. Theorems A.2 and A.3 therein). It is a generalization of the so-called *fundamental theorem of Young measures* (cf. the classical results [12, Thm. 1] [13, Thm. 2.2], [14], [78, Thm. 16]) to the case of Young measures with values in $\mathcal{Y}$ endowed with the weak topology (see also [67, Thm. 3.2] for the case in which $\mathcal{Y}$ is a Hilbert space endowed with the weak topology).

THEOREM A.3 (the fundamental theorem for weak topologies). *Let $(y_n) \subset \mathscr{M}(0,T;\mathcal{Y})$ be a weakly tight sequence. Then the following hold.*

1. *There exist a subsequence (y_{n_k}) and a Young measure $\boldsymbol{\sigma} = (\sigma_t)_{t\in(0,T)} \in \mathscr{Y}(0,T;\mathcal{Y})$ such that*

$$\limsup_{k\uparrow\infty} \|y_{n_k}(t)\|_{\mathcal{Y}} < \infty \quad \text{and} \quad \operatorname{supp}(\sigma_t) \subset \bigcap_{j=1}^{\infty} \overline{\{y_{n_k}(t) : k \geq j\}}^{\text{weak}} \quad \text{for a.a. } t \in (0,T) \tag{A.3}$$

(where $\overline{B}^{\text{weak}}$ denotes the closure of a set $B \subset \mathcal{Y}$ w.r.t. the weak topology), and such that for every weakly normal integrand $h : [0,T]\times\mathcal{Y} \to (-\infty,\infty]$ such that $h^-(\cdot, y_{n_k}(\cdot)) = \max\{0; -h(\cdot, y_{n_k}(\cdot))\}$ is uniformly integrable it holds that

$$\liminf_{k\to\infty} \int_0^T h\left(t, y_{n_k}(t)\right) \mathrm{d}t \geq \int_0^T \int_{\mathcal{Y}} h\left(t, y\right) \mathrm{d}\sigma_t(y)\,\mathrm{d}t. \tag{A.4}$$

2. *In particular, let $(y_n) \subset L^q(0,T;\mathcal{Y})$ be a bounded sequence, with $q \in (1,+\infty]$. Then, there exists a further (not relabeled) subsequence (y_{n_k}) and a Young measure $\boldsymbol{\sigma} = \{\sigma_t\}_{t\in(0,T)} \in \mathscr{Y}(0,T;\mathcal{Y})$ such that for almost all $t \in (0,T)$ properties* (A.3) *hold. Setting $\mathrm{y}(t) := \int_{\mathcal{Y}} y \,\mathrm{d}\sigma_t(y)$ for almost all $t \in (0,T)$, there holds*

$$y_{n_k} \overset{*}{\rightharpoonup} \mathrm{y} \quad \text{in } L^p(0,T;\mathcal{Y}). \tag{A.5}$$

A.1. A Young measure version of the chain rule. In what follows, we will work with Young measures with values in the space $\mathcal{Y} = X^* \times \mathbb{R}$. Our first result, a small variation of [51, Prop. B.1], provides the version of the chain rule inequality (3.$\mathcal{E}_4$) in terms of Young measures used in Step 4 of the proof of Theorem 4.5.

LEMMA A.4. *In the framework of* (3.15)*, let $\alpha : X \rightrightarrows X^*$ fulfill* (3.α_0) *and the coercivity condition* (4.1)*, and let $\mathcal{E} : [0,T]\times X \to (-\infty,+\infty]$ comply with Assumption* 4.4*. Let $u \in BV([0,T];X)$ satisfy*

$$\begin{aligned} &\sup_{t\in[0,T]} \mathcal{E}_t(u(t)) < \infty\,, \quad (t,u(t)) \in \operatorname{dom}(\partial\mathcal{E}) \text{ for a.a. } t \in (0,T), \\ &\int_0^T |\partial_t \mathcal{E}_t(u(t))|\,\mathrm{d}t < \infty, \exists E \in BV([0,T]) \text{ such that } E(t) = \mathcal{E}_t(u(t)) \\ &\text{for a.a. } t \in (0,T), \end{aligned} \tag{A.6}$$

and let $(\sigma_t)_{t\in(0,T)} \in \mathscr{Y}(0,T;X^*\times\mathbb{R})$ be a Young measure such that

$$\forall (\xi,p) \in \operatorname{supp}(\sigma_t) : \xi \in \partial\mathcal{E}_t(u(t))\,,\ p \leq \partial_t\mathcal{E}_t(u(t)) \text{ for a.a. } t \in (0,T), \tag{A.7}$$

$$\int_0^T \int_{X^*\times\mathbb{R}} f_\alpha(\dot{u}_{\mathrm{ac}}(s), -\zeta)\,\mathrm{d}\sigma_s(\zeta,p)\,\mathrm{d}s < \infty. \tag{A.8}$$

Then, for almost all $t \in (0,T)$ such that t is Lebesgue point of $\dot{E}_{\mathrm{ac}}$ and $\dot{u}_{\mathrm{ac}}$, there holds

$$\dot{E}_{\mathrm{ac}}(t) \geq \int_{X^*\times\mathbb{R}} \left(\langle \zeta, \dot{u}_{\mathrm{ac}}(t)\rangle + p\right) \mathrm{d}\sigma_t(\zeta,p)\,. \tag{A.9}$$

Proof. We consider the set $K(t,u(t)) := \{(\xi,p) \in X^* \times \mathbb{R} \, : \, \xi \in \partial\mathcal{E}_t(u(t)),\ p \leq \partial_t\mathcal{E}_t(u(t))\}$. Repeating the very same arguments as in the proof of [51, Prop. B.1], we can show that there exists a sequence (ξ_n, p_n) of strongly measurable functions $(\xi_n, p_n) : (0,T) \to X^* \times \mathbb{R}$ such that

$$\{(\xi_n(t), p_n(t)) \, : \, n \in \mathbb{N}\} \subset K(t,u(t)) \subset \overline{\{(\xi_n(t), p_n(t)) \, : \, n \in \mathbb{N}\,\}} \quad \text{for a.a. } t \in (0,T) \tag{A.10}$$

(where $\overline{B}$ denotes the closure of $B \subset X^* \times \mathbb{R}$ w.r.t. the strong topology of $X^* \times \mathbb{R}$).

We now claim that the sequence (ξ_n, p_n) can be chosen such that

$$\forall n \in \mathbb{N} : \xi_n \in L^1(0,T;X^*) \ \text{ and } \sup_{n\in\mathbb{N}} \int_0^T f_\alpha(\dot{u}_{\mathrm{ac}}(t), -\xi_n(t))\,\mathrm{d}t < \infty. \tag{A.11}$$

To this aim, we define the function $g(t) := \inf\{f_\alpha(\dot{u}_{\mathrm{ac}}(t), -\zeta) : (\zeta,p) \in K(t,u(t))\}$ for almost all $t \in (0,T)$. Notice that due to (A.10) it holds that

$$g(t) := \inf_{n\in\mathbb{N}}\{f_\alpha(\dot{u}_{\mathrm{ac}}(t), -\xi_n(t))\} \quad \text{for a.a. } t \in (0,T), \tag{A.12}$$

and hence g is measurable on $(0,T)$. Moreover,

$$\int_0^T g(t)\,\mathrm{d}t \leq \int_0^T \int_{X^*\times\mathbb{R}} f_\alpha(\dot{u}_{\mathrm{ac}}(t), -\zeta)\,\mathrm{d}\sigma_t(\zeta,p)\,\mathrm{d}t < \infty. \tag{A.13}$$

With a straightforward adaptation of the argument of [51, Prop. B.1] (see also [67, Lemma 3.4]), from (A.12) and (A.13) we deduce (A.11).

In view of the obtained (A.10) and (A.11), we are in the position to apply the chain rule inequality ($3.\mathcal{E}_4$) to the pair (u,ξ_n) for every $n \in \mathbb{N}$. Therefore, for every $n \in \mathbb{N}$ there exists a set $\mathcal{T}_n \subset (0,T)$ of full measure such that $\dot{E}_{\mathrm{ac}}(t) \geq \langle \xi_n(t), \dot{u}_{\mathrm{ac}}(t)\rangle + p_n(t)$ for all $t \in \mathcal{T}_n$, where we have also used that $p_n(t) \leq \partial_t\mathcal{E}_t(u(t))$. The set $\mathcal{T} = \bigcap_{n\in\mathbb{N}} \mathcal{T}_n$ still has full measure, and there holds for all $t \in \mathcal{T}$

$$\dot{E}_{\mathrm{ac}}(t) \geq \langle \zeta, \dot{u}_{\mathrm{ac}}(t)\rangle + p \quad \forall\,(\zeta,p) \in \overline{\mathrm{conv}\ K(t,u(t))}, \tag{A.14}$$

the latter set denoting the closed convex hull of $K(t,u(t))$. Integrating (A.14) w.r.t. the measure σ_t, we obtain (A.9). ∎

We conclude with the measurable selection result exploited in Step 5 of the proof of Theorem 4.5.

LEMMA A.5. *In the framework of* (3.15)*, let* $\alpha : X \rightrightarrows X^*$ *fulfill* ($3.\alpha_0$) *and the coercivity condition* (4.1)*, and let* $\mathcal{E} : [0,T] \times X \to (-\infty,+\infty]$ *comply with Assumptions* 3.9 *and* 4.4*. Furthermore, let* $u \in BV([0,T];X)$ *fulfill* (A.6)*. Suppose that for almost all* $t \in (0,T)$

$$\begin{aligned}\mathcal{S}(t,u(t),\dot{u}_{\mathrm{ac}}(t)) := \{(\zeta,p) \in X^* \times \mathbb{R} : \zeta \in \partial\mathcal{E}_t(u(t))\,,\\ -\zeta \in \alpha(\dot{u}_{\mathrm{ac}}(t))\,, p \leq \partial_t\mathcal{E}_t(u(t))\} \neq \emptyset\,.\end{aligned} \tag{A.15}$$

Then, there exist measurable functions $\xi : (0,T) \to X^*$ *and* $P : (0,T) \to \mathbb{R}$ *such that*

$$\begin{aligned}&(\xi(t),P(t)) \in \mathrm{argmin}\,\{f_\alpha(\dot{u}_{\mathrm{ac}}(t), -\zeta) - p : (\zeta,p) \in \mathcal{S}(t,u(t),\dot{u}_{\mathrm{ac}}(t))\}\\ &\quad \textit{for a.a. } t \in (0,T)\,.\end{aligned} \tag{A.16}$$

Proof. The argument follows the very same lines of [51, Lemma B.2]. First, we observe that

(A.17)
$$\operatorname{argmin}\left\{f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)-p:(\zeta,p)\in\mathcal{S}(t,u(t),\dot{u}_{\rm ac}(t))\right\}\neq\emptyset \quad \text{for a.a. } t\in(0,T)\,.$$

To this aim, let $(\zeta_n,p_n)\subset\mathcal{S}(t,u(t),\dot{u}_{\rm ac}(t))$ be an infimizing sequence; then there exist constants C, $C'>0$ such that for every $n\in\mathbb{N}$

$$C\geq f_\alpha(\dot{u}_{\rm ac}(t),-\zeta_n)-p_n=\langle-\zeta_n,\dot{u}_{\rm ac}(t)\rangle-p_n\geq c_1\|\dot{u}_{\rm ac}(t)\|^p+c_2\|\zeta_n\|_*^q-c_3-C', \tag{A.18}$$

where we have used that $-\zeta_n\in\dot{u}_{\rm ac}(t)$, the coercivity property (4.1) of α, and that $p_n\leq\partial_t\mathcal{E}_t(u(t))\leq C$ due to the fact that $\sup_{t\in[0,T]}\mathcal{E}_t(u(t))<\infty$ and to (3.$\mathcal{E}_2$). Therefore, we infer that $\sup_{n\in\mathbb{N}}(\|\zeta_n\|_*^q+|p_n|)<\infty$. Hence, there exist $(\zeta,p)\in X^*\times\mathbb{R}$ such that, up to a not relabeled subsequence, $\zeta_n\rightharpoonup\zeta$ in X^* and $p_n\to p$. Thanks to the closedness condition (3.$\mathcal{E}_3$) and to the weak closedness of $\alpha(\dot{u}_{\rm ac}(t))$, we have $(\zeta,p)\in\mathcal{S}(t,u(t),\dot{u}_{\rm ac}(t))$. Using that $\zeta\mapsto f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)$ is (sequentially) weakly lower semicontinuous, we conclude that

$$\liminf_{n\to\infty}\left(f_\alpha(\dot{u}_{\rm ac}(t),-\zeta_n)-p_n\right)\geq f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)-p,$$

and (A.17) ensues.

Once (A.17) is obtained, the argument for (A.16) is a straightforward adaptation of the proof of [51, Lemma B.2], to which we refer the reader for the details. Let us only mention here that the existence of (ξ,P) as in (A.16) is a consequence of the measurable selection results [24, Cor. III.3, Thm. III.6]. ∎

Remark A.6. Let us stress that the requirement $\zeta\in\alpha(\dot{u}_{\rm ac}(t))$ in the definition (A.15) of the set $\mathcal{S}(t,u(t),\dot{u}_{\rm ac}(t))$ has a crucial role in proving that

$$\operatorname{argmin}\{f_\alpha(\dot{u}_{\rm ac}(t),-\zeta)-p:(\zeta,p)\in\mathcal{S}(t,u(t),\dot{u}_{\rm ac}(t))\}$$

is nonempty. In fact, it ensures the estimates in (A.18) for any infimizing sequence (ζ_n,p_n).

Acknowledgments. Thomas Roche would like to thank IMATI Pavia and the University of Brescia, where part of this work was conducted, for the kind hospitality. We wish to thank the anonymous referees for their many insightful and pointed remarks.

REFERENCES

[1] S. Aizicovici and Q. Yan, *Convergence theorems for abstract doubly nonlinear differential equations*, Panamer. Math. J., 7 (1997), pp. 1–17.
[2] G. Akagi, *Doubly nonlinear evolution equations in reflexive Banach spaces*, in Nonlinear Phenomena with Energy Dissipation, GAKUTO Internat. Ser. Math. Sci. Appl. 29, Gakkōtosho, Tokyo, 2008, pp. 19–36.
[3] G. Akagi, *Doubly nonlinear evolution equations with non-monotone perturbations in reflexive Banach spaces*, J. Evol. Equ., 11 (2011), pp. 1–41.
[4] G. Akagi, *Global attractors for doubly nonlinear evolution equations with non-monotone perturbations*, J. Differential Equations, 250 (2011), pp. 1850–1875.
[5] G. Akagi and M. Ôtani, *Time-dependent constraint problems arising from macroscopic critical-state models for type*-II *superconductivity and their approximations*, Adv. Math. Sci. Appl., 14 (2004), pp. 683–712.
[6] G. Akagi and U. Stefanelli, *A variational principle for doubly nonlinear evolution*, Appl. Math. Lett., 23 (2010), pp. 1120–1124.
[7] G. Akagi and U. Stefanelli, *Periodic solutions for doubly nonlinear evolution equations*, J. Differential Equations, 251 (2011), pp. 1790–1812.

[8] G. Akagi and U. Stefanelli, *Weighted energy-dissipation functionals for doubly nonlinear evolution*, J. Funct. Anal., 260 (2011), pp. 2541–2578.

[9] T. Arai, *On the existence of the solution for* $\partial\varphi(u'(t)) + \partial\psi(u(t)) \ni f(t)$, J. Fac. Sci. Univ. Tokyo Sect. IA Math., 26 (1979), pp. 75–96.

[10] H. Attouch, *Variational Convergence for Functions and Operators*, Pitman (Advanced Publishing Program), Boston, MA, 1984.

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

[12] E. J. Balder, *A general approach to lower semicontinuity and lower closure in optimal control theory*, SIAM J. Control Optim., 22 (1984), pp. 570–598.

[13] E. J. Balder, *An extension of Prohorov's theorem for transition probabilities with applications to infinite-dimensional lower closure problems*, Rend. Circ. Mat. Palermo (2), 34 (1985), pp. 427–447.

[14] J. M. Ball, *A version of the fundamental theorem for Young measures*, in PDEs and Continuum Models of Phase Transitions (Nice, 1988), Lecture Notes in Phys. 344, Springer, Berlin, 1989, pp. 207–215.

[15] H. H. Bauschke and X. Wang, *The kernel average for two convex functions and its applications to the extension and representation of monotone operators*, Trans. Amer. Math. Soc., 361 (2009), pp. 5947–5965.

[16] H. Brezis, *Opérateurs maximaux monotones et semi-groupes de contractions dans les espaces de Hilbert*, North-Holland Math. Stud. 5, North-Holland, Amsterdam, 1973.

[17] H. Brezis 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), pp. A1197–A1198.

[18] H. Brezis 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), pp. A971–A974.

[19] M. Brokate and J. Sprekels, *Hysteresis and Phase Transitions*, Appl. Math. Sci. 121, Springer, Berlin, 1996.

[20] M. Buliga, G. de Saxcé, and C. Vallée, *Existence and construction of bipotentials for graphs of multivalued laws*, J. Convex Anal., 15 (2008), pp. 87–104.

[21] M. Buliga, G. de Saxcé, and C. Vallée, *Bipotentials for non-monotone multivalued operators: Fundamental results and applications*, Acta Appl. Math., 110 (2010), pp. 955–972.

[22] R. S. Burachik and B. F. Svaiter, *Maximal monotone operators, convex functions and a special family of enlargements* Set-Valued Anal., 10 (2002), pp. 297–316.

[23] R. S. Burachik and B. F. Svaiter, *Maximal monotonicity, conjugation and the duality product*, Proc. Amer. Math. Soc., 131 (2003), pp. 2379–2383.

[24] C. Castaing and M. Valadier, *Convex Analysis and Measurable Multifunctions*, Lecture Notes in Math. 580, Springer-Verlag, Berlin, 1977.

[25] P. Colli, *On some doubly nonlinear evolution equations in Banach spaces*, Japan J. Indust. Appl. Math., 9 (1992), pp. 181–203.

[26] P. Colli and A. Visintin, *On a class of doubly nonlinear evolution equations*, Comm. Partial Differential Equations, 15 (1990), pp. 737–756.

[27] G. Dal Maso, *An Introduction to* Γ*-Convergence*, Progress in Nonlinear Differential Equations and Their Applications 8, Birkhäuser Boston, Boston, MA, 1993.

[28] G. Dal Maso, A. DeSimone, and M. G. Mora, *Quasistatic evolution problems for linearly elastic-perfectly plastic materials*, Arch. Ration. Mech. Anal., 180 (2006), pp. 237–291.

[29] G. Dal Maso, G. A. Francfort, and R. Toader, *Quasistatic crack growth in nonlinear elasticity*, Arch. Ration. Mech. Anal., 176 (2005), pp. 165–225.

[30] J. Diestel and J. J. Uhl, Jr, *Vector Measures*, AMS, Providence, RI, 1977.

[31] M. Efendief and S. Zelik, *Finite-dimensional attractors and exponential attractors for degenerate doubly nonlinear equations*, Math. Methods Appl. Sci., 32 (2009), pp. 1638–1668.

[32] M. A. Efendiev and A. Mielke, *On the rate-independent limit of systems with dry friction and small viscosity*, J. Convex Anal., 13 (2006), pp. 151–167.

[33] E. Emmrich and M. Thalhammer, *Doubly nonlinear evolution equations of second order: Existence and fully discrete approximation*, J. Differential Equations, 251 (2011), pp. 82–118.

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

[35] I. Fonseca and G. Leoni, *Modern Methods in the Calculus of Variations:* L^p *Spaces*, Springer Monographs in Mathematics, Springer, New York, 2007.

[36] G. Francfort and A. Mielke, *Existence results for a class of rate-independent material models with nonconvex elastic energies*, J. Reine Angew. Math., 595 (2006), pp. 55–91.
[37] P. Germain, *Cours de mécanique des milieux continus. Tome* I*: Théorie Générale*, Masson et Cie, Éditeurs, Paris, 1973.
[38] N. Ghoussoub, *A variational theory for monotone vector fields*, J. Fixed Point Theory Appl., 4 (2008), pp. 107–135.
[39] N. Ghoussoub, *Selfdual Partial Differential Systems and Their Variational Principles*, Springer, New York, 2009.
[40] W. Han and B. D. Reddy, *Plasticity, Mathematical Theory and Numerical Analysis*, Springer, New York, 1999.
[41] N. Kraynyukova and H.-D. Alber, *A doubly nonlinear problem associated with a mathematical model for piezoelectric material behavior*, ZAMM Z. Angew. Math. Mech., 92 (2012), pp. 141–159.
[42] P. Krejčí, *Evolution variational inequalities and multidimensional hysteresis operators*, in Nonlinear Differential Equations (Chvalatice, 1998), Chapman & Hall/CRC Res. Notes Math. 404, Chapman & Hall/CRC, Boca Raton, FL, 1999, pp. 47–110.
[43] A. Mainik and A. Mielke, *Existence results for energetic models for rate-independent systems*, Calc. Var. Partial Differential Equations, 22 (2005), pp. 73–99.
[44] M. Marques Alves and B. F. Svaiter, *Maximal monotonicity, conjugation and the duality product in non-reflexive Banach spaces*, J. Convex Anal., 17 (2010), pp. 553–563.
[45] J. E. Martínez-Legaz and B. F. Svaiter, *Monotone operators representable by l.s.c. convex functions*, Set-Valued Anal., 13 (2005), pp. 21–46.
[46] J. E. Martínez-Legaz and B. F. Svaiter, *Minimal convex functions bounded below by the duality product*, Proc. Amer. Math. Soc., 136 (208), pp. 873–878.
[47] A. Mielke, *Existence of minimizers in incremental elasto-plasticity with finite strains*, SIAM J. Math. Anal., 36 (2004), pp. 384–404.
[48] A. Mielke, *Evolution of rate-independent systems*, in Handbook of Differential Equations, Evolutionary Equations 2, C. Dafermos and E. Feireisl, eds., Elsevier, New York, 2005, pp. 461–559.
[49] A. Mielke, R. Rossi, and G. Savaré, *Modeling solutions with jumps for rate-independent systems on metric spaces*, Discrete Contin. Dyn. Syst. Ser. A, 25 (2009), pp. 585–615.
[50] A. Mielke, R. Rossi, and G. Savaré, *BV solutions and viscosity approximations of rate-independent systems*, ESAIM Control Optim. Calc. Var., 18 (2012), pp. 36–80.
[51] A. Mielke, R. Rossi, and G. Savaré, *Nonsmooth analysis of doubly nonlinear evolution equations*, Calc. Var. Partial Differential Equations, 46 (2013), pp. 253–310.
[52] A. Mielke, R. Rossi, and G. Savaré, *Variational convergence of gradient flows and rate-independent evolutions in metric spaces*, Milan J. Math., 80 (2012), pp. 381–410.
[53] A. Mielke and T. Roubíček, *Rate-independent damage processes in nonlinear elasticity*, Math. Models Methods Appl. Sci., 16 (2006), pp. 177–209.
[54] A. Mielke, T. Roubíček, and U. Stefanelli, Γ*-limits and relaxations for rate-independent evolutionary problems*, Calc. Var. Partial Differential Equations, 31 (2008), pp. 387–416.
[55] A. Mielke and F. Theil, *On rate-independent hysteresis models*, NoDEA Nonlinear Differential Equations Appl., 11 (2004), pp. 151–189.
[56] A. Mielke, F. Theil, and V. I. Levitas, *A variational formulation of rate-independent phase transformations using an extremum principle*, Arch. Ration. Mech. Anal., 162 (2002), pp. 137–177.
[57] J.-J. Moreau, *Sur les lois de frottement, de viscosité et plasticité*, C. R. Acad. Sci. Paris Sér. II Méc. Phys. Chim. Sci. Univers Sci. Terre, 271 (1970), pp. 608–611.
[58] J.-J. Moreau, *Sur l'évolution d'un système élasto-visco-plastique*, C. R. Acad. Sci. Paris Sér. A-B, 273 (1971), pp. A118–A121.
[59] J.-J. Moreau, *Bounded variation in time*, in Topics in Nonsmooth Mechanics, J.-J. Moreau, P. D. Panagiotopoulos, G. Strang, eds.. Birkhäuser, Basel, 1988, pp. 1–74.
[60] B. Nayroles, *Deux théorèmes de minimum pour certains systèmes dissipatifs*, C. R. Acad. Sci. Paris Sér. A-B, 282 (1976), pp. A1035–A1038.
[61] B. Nayroles, *Un théorème de minimum pour certains systèmes dissipatifs. Variante hilbertienne*, Travaux Sém. Anal. Convexe, 6 (1976), 2.
[62] J.-P. Penot, *Autoconjugate functions and representations of monotone operators*, Bull. Austral. Math. Soc., 67 (2003), pp. 277–284.
[63] J.-P. Penot, *The relevance of convex analysis for the study of monotonicity*, Nonlinear Anal., 58 (2004), pp. 855–871.
[64] J.-P. Penot, *A representation of maximal monotone operators by closed convex functions and its impact on calculus rules*, C. R. Math. Acad. Sci. Paris, 338 (2004), pp. 853–858.

[65] J.-P. Penot and C. Zălinescu, *Some problems about the representation of monotone operators by convex functions*, ANZIAM J., 47 (2005), pp. 1–20.

[66] R. Rossi, A. Mielke, and G. Savaré, *A metric approach to a class of doubly nonlinear evolution equations and applications*, Ann. Sc. Norm. Super. Pisa Cl. Sci. (5), 7 (2008), pp. 97–169.

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

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

[69] G. Schimperna and A. Segatti, *Attractors for the semiflow associated with a class of doubly nonlinear parabolic equations*, Asymptot. Anal., 56 (2008), pp. 61–86.

[70] G. Schimperna, A. Segatti, and U. Stefanelli, *Well-posedness and long-time behavior for a class of doubly nonlinear equations*, Discrete Contin. Dyn. Syst., 18 (2007), pp. 15–38.

[71] A. Segatti, *Global attractor for a class of doubly nonlinear abstract evolution equations*, Discrete Contin. Dyn. Syst., 14 (2006), pp. 801–820.

[72] T. Senba, *On some nonlinear evolution equation*, Funkcial. Ekval., 29 (1986), pp. 243–257.

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

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

[75] U. Stefanelli, *Some remarks on convergence and approximation for a class of hysteresis problems*, Istit. Lombardo Accad. Sci. Lett. Rend. A, 140 (2009), pp. 81–108.

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

[77] B. F. Svaiter, *Fixed points in the family of convex representations of a maximal monotone operator*, Proc. Amer. Math. Soc., 131 (2003), pp. 3851–3859.

[78] M. Valadier, *Young measures*, in Methods of Nonconvex Analysis (Varenna, 1989), Lecture Notes in Math. 1446, Springer, Berlin, 1990, pp. 152–188.

[79] A. Visintin, *Differential Models of Hysteresis*, Appl. Math. Sci. 111, Springer, Berlin, 1994.

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

[81] A. Visintin, *Structural stability of doubly-nonlinear flows*, Boll. Unione Mat. Ital. (9), 4 (2011), pp. 363–391.

[82] A. Visintin, *Structural stability of rate-independent nonpotential flows*, Discrete Contin. Dyn. Syst. S, 6 (2013), pp. 257–275.

[83] A. Visintin, *Variational formulation and structural stability of monotone equations*, Calc. Var. Partial Differential Equations, 47 (2013), pp. 273–317.