Asymptotic Behavior of Penalty Dynamics for Constrained Variational Inequalities

Juan Peypouquet[1] · Siqi Qu[1,2] · Mathias Staudigl[2]

Abstract

We propose a comprehensive framework for solving constrained variational inequalities via various classes of evolution equations displaying multi-scale aspects. In an infinite-dimensional Hilbertian framework, the class of dynamical systems we propose combine Tikhonov regularization and exterior penalization terms in order to induce strong convergence of trajectories to least norm solutions in the constrained domain. Our construction thus unifies the literature on regularization methods and penalty-based dynamical systems. An extension to a full splitting formulation of the constrained domain is also provided, with associated weak convergence results involving the Attouch-Czarnecki condition.

1 Introduction

Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and corresponding norm $\|\cdot\|$.

Problem 1 We are given

- $A : \mathcal{H} \to 2^{\mathcal{H}}$ a maximally monotone operator;
- $D : \mathcal{H} \to \mathcal{H}$ is monotone and $\frac{1}{\eta}$-Lipschitz;

✉ Siqi Qu
qu.siqi@uni-mannheim.de

Juan Peypouquet
j.g.peypouquet@rug.nl

Mathias Staudigl
m.staudigl@uni-mannheim.de

[1] Faculty of Science and Engineering, Systems, Control and Optimization — Bernoulli Institute, Rijksuniversiteit Groningen, Nijenborgh 9, Groningen 9747 AG, The Netherlands

[2] Department of Mathematics, University of Mannheim, B6 26, Mannheim 68159, Germany

- $\mathcal{C} \triangleq \mathrm{zer}(\mathsf{B}) \neq \varnothing$ a nonempty closed convex set which can be represented as the zeros of a cocoercive operator $\mathsf{B} : \mathcal{H} \to \mathcal{H}$.

Our aim is to find $x \in \mathcal{H}$ such that

$$0 \in \Phi(x) \triangleq \mathsf{A}(x) + \mathsf{D}(x) + \mathsf{N}_{\mathcal{C}}(x). \tag{P}$$

This is a three-operator formulation of a general class of variational problems, in which a constrained equilibrium of the sum of two maximally monotone operators $\mathsf{A} + \mathsf{D}$ is requested over a closed convex domain $\mathcal{C} \subset \mathcal{H}$. In this paper we approach this problem via a combination of an exterior penalization method and Tikhonov regularization, essentially leading to a hierarchical equilibrium formulation. A key assumption in our approach is thus that the feasible set $\mathcal{C}$ admits a simple representation in terms of the set of zeros of another single-valued monotone operator B. This general formulation admits many applications in optimal control and optimization, in particular those problems featuring a hierarchical structure. We briefly describe some examples below, and provide a more detailed exposition on applications in Section 5.

Example 1 *(Simple Bilevel Optimization)* A simple bilevel optimization problem is formulated as

$$\min f(x)$$
$$\text{s.t.: } x \in \mathcal{C} = \mathrm{argmin}\{\Psi(y) : y \in \mathcal{H}\}$$

where Ψ is a convex and Fréchet differentiable function, with L-Lipschitz continuous gradient, i.e. $\Psi \in \mathbf{C}_L^{1,1}(\mathcal{H})$. By Fermat's optimality condition

$$x \in \mathcal{C} \Leftrightarrow 0 = \nabla\Psi(x).$$

Hence, $\mathcal{C} = \mathrm{zer}(\nabla\Psi)$ defines the set of solutions to the lower-level optimization problem over which a minimizer of the function f is searched for. These are simplest hierarchical optimization problems, in which the decision variables in the lower and an upper level variational problems are completely decoupled. This class of problems has been studied extensively. See, for instance [22, 36, 45] in an infinite-dimensional setting, or [41] in finite dimension. For a recent dynamical systems perspective with acceleration, we refer to [14]

Example 2 *(Constrained Variational Inequalities)* Let $\mathcal{C} \subset \mathcal{H}$ be a set for which we know a convex function $\Psi \in \mathbf{C}_{L_\Psi}^{1,1}(\mathcal{H})$ having the properties that $\Psi \geq 0$ and $\mathcal{C} = \Psi^{-1}(0)$. Then, $\mathsf{B} = \nabla\Psi$ is a maximally monotone operator whose zero set is $\mathcal{C}$. Given a mapping $\mathsf{D} : \mathcal{H} \to \mathcal{H}$ and a proper convex and lower semi-continuous function $h : \mathcal{H} \to (-\infty, \infty]$, we search for a solution of the variational inequality of the second kind

$$\text{Find } \bar{x} \in \mathcal{C} \text{ such that } \langle \mathsf{D}(\bar{x}), x - \bar{x} \rangle + h(x) - h(\bar{x}) \geq 0 \quad \forall x \in \mathcal{C}. \tag{1}$$

This is equivalent to the inclusion problem

$$\text{Find } \bar{x} \in \mathcal{C} \text{ such that } 0 \in \mathsf{D}(\bar{x}) + \partial h(\bar{x}) + \mathsf{N}_{\mathcal{C}}(x), \qquad \forall x \in \mathcal{C}.$$

and arises frequently in optimal control problems [11, 39, 40].

Our Approach

To tackle problem class (P), we follow a recent trend in variational analysis by approaching the general resolution of problem (P) via dynamical systems derived from operator splitting methods featuring multiscale aspects. Our schemes can be considered as hybrid versions of penalty-based methods, inspired by [5, 6, 10, 16, 17, 35], and Tikhonov regularization of dynamical systems [4, 24, 29]. Specifically, if D is cocoercive, we propose a forward-backward dynamical system of the form

$$\dot{x}(t) + x(t) = \mathsf{J}_{\lambda(t)\mathsf{A}}\left(x(t) - \lambda(t)V_t(x)\right), \tag{FB}$$

where $V_t : \mathcal{H} \to \mathcal{H}$ is a time-dependent vector field, defined as

$$V_t(x) \triangleq \mathsf{D}(x) + \varepsilon(t)x + \beta(t)\mathsf{B}(x).$$

The evolution equation (FB) generalizes most of the first-order dynamical systems that have been studied so far in the literature. Indeed, if $\mathsf{J}_{\lambda(t)\mathsf{A}} = \mathrm{Id}_{\mathcal{H}}$ (arising if $\mathsf{A} = \{0\}$), then (FB) simplifies to

$$\dot{x}(t) + \lambda(t)(\mathsf{D}(x(t)) + \varepsilon(t)x(t) + \beta(t)\mathsf{B}(x(t))) = 0.$$

In the full potential case, i.e. when $\mathsf{D} = \nabla f$ and $\mathsf{B} = \nabla\Psi$, for convex potentials $f, \Psi : \mathcal{H} \to \mathbb{R}$ having a Lipschitz continuous gradient, with $\mathcal{C} = \Psi^{-1}(0) = \mathrm{argmin}(\Psi) = \mathrm{zer}(\nabla\Psi)$, we obtain an exterior penalization of the gradient method, as studied in [38], augmented with Tikhonov regularization, for solving set-constrained convex minimization problems of the form

$$\min f(x) \quad \text{s.t. } x \in \mathcal{C}.$$

Setting $\mathsf{B} = 0$ (no penalty), the dynamics reduce to a gradient system with Tikhonov regularization, a class of dynamical systems which has received enormous attention in the literature (references [2, 3, 5, 9, 26, 29] give a partial overview on this literature). In the context of full splitting and maximal monotone operators instead of potentials, particular instances of (FB) have been studied in [19] without explicit constraints and Tikhonov regularization, in [17] for the case with a penalty term but no Tikhonov regularization, and in [24] in the context of Tikhonov regularization and without penalty terms.

For large-scale equilibrium and minmax problems, the cocoercivity of the single-valued operator D is typically an incompatible hypothesis. For such problems, we study

an extragradient-based dynamical system in the spirit of Tseng [44], featuring multi-scale aspects. This dynamical system is defined in terms of the projection-differential dynamical system

$$\begin{cases} p(t) = \mathsf{J}_{\lambda(t)\mathsf{A}}\left(x(t) - \lambda(t)V_t(x(t))\right) \\ \dot{x}(t) = p(t) - x(t) + \lambda(t)[V_t(x(t)) - V_t(p(t))]. \end{cases} \tag{FBF}$$

The asymptotic properties of this evolution equation with Tikhonov regularization has been studied in [24], while the penalty case has been investigated in [16], although in discrete time. While we do not treat this problem directly here, it should be mentioned that our approach can be used to solve constrained min-max problems in infinite dimensions [31, 32]. A thorough analysis of this important problem class in our multi-scale dynamical system framework is left for future research.

Contributions and Related Literature

Our dynamical system combines exterior-penalty methods with Tikhonov regularization. With this hybrid construction, we generalize the pure penalization-based dynamical systems studied in [6, 38], who concentrate on the case where B is the gradient of a convex function $\Psi : \mathcal{H} \to [0, \infty]$, satisfying $\mathcal{C} = \Psi^{-1}(0) = \operatorname{argmin}_{\mathcal{H}}(\Psi)$, and [16] who extended this to the monotone inclusion setting. Adding the Tikhonov term to the dynamical system allows us to enforce strong convergence to the least-norm solution. We borrow ideas from [24] to analyze the effects of the Tikhonov term, and extend this technique to deal with constrained variational inequalities. To the best of our knowledge this is a new result in the literature since the analysis of penalty dynamics, *in tandem* with Tikhonov regularization, has not been done so far. In particular, this combined approach gains relevance in inverse problems and PDE constrained optimization; See [33] for a recent approach in this direction.

Our work is closely related to the hierarchical minimization formulation of Cabot [26]. He proposed an inexact proximal point method for the hierarchical minimization of a finitely many convex potentials. Specifically, in the case of three convex functions f_0, f_1, f_2, the method of [26] is designed for solving the nested optimization problems $S_0 = \operatorname{argmin}_{\mathcal{H}} f_0$, $S_1 = \operatorname{argmin}_{S_0} f_1$ and $S_2 = \operatorname{argmin}_{S_1} f_2$. Under time rescaling and constraint qualification, the exact version of the proximal point method of [26] thus becomes

$$-\frac{x_n - x_{n+1}}{\lambda_n} \in \beta_n \partial f_0 + \partial f_1 + \varepsilon_n \partial f_2.$$

In our work, we are interested in continuous-time systems and general monotone operators. Both directions lie out of the scope of [26].

In parallel to this work, a series of papers studied penalty methods and Tikhonov regularization from the lens of dynamical systems (separately). In particular, [15] establishes an interesting connection between the Tikhonov regularization and the Halpern schemes and thereby shows the acceleration potential of the Tikhonov regularization. Inertial dynamics with Tikhonov regularization have been studied in [7,

25], among many others. Penalty dynamics have been extended to second order in time systems in [20, 23].[1]

Organization of the Paper

After briefly recalling some known facts from convex analysis and monotone operator theory, Section 2 presents a detailed analysis of the *central paths*, which are curves parameterized by regularization variables, representing solutions to auxiliary monotone inclusion problems. In particular, we show that central paths are absolutely continuous, differentiable - an important ingredient in our proof building on Lyapunov analysis - and approximate the least-norm solution of (P). Section 3 is concerned with two dynamical systems intended to approximate solutions of (P). The nature of these systems, of either forward-backward or forward-backward-forward type, depends on whether the operator D is cocoercive. For each system, and under suitable assumptions on the regularization parameters, we show that every trajectory convergence strongly to the least norm solution of (P). This is achieved by establishing a tracking property of the dynamics with respect to the central paths. Section 4 extends our analysis of the (FB) dynamical systems to a domain decomposition problem, modeled via multiple penalty functions. Section 5 describes some scenarios in optimal control and non-linear analysis to which our method naturally applies. We also present implementations on a class of image deblurring problems, to illustrate the computational efficacy of the method.

2 Preliminaries

We follow the standard notation from convex and non-smooth analysis (see e.g. [12, 37]). Let $\mathcal{H}$ be a real Hilbert space with inner product $\langle \cdot, \cdot \rangle$ and associated norm $\|\cdot\|$. The symbols $\rightharpoonup$ and $\rightarrow$ denote weak and strong convergence, respectively. For an extended-real valued function $f : \mathcal{H} \rightarrow \bar{\mathbb{R}} \triangleq \mathbb{R} \cup \{-\infty, \infty\}$, we denote by $\mathrm{dom}(f) = \{x \in \mathcal{H} | f(x) < \infty\}$ its effective domain and say that f is proper if $\mathrm{dom}(f) \neq \varnothing$ and $f(x) \neq -\infty$ for all $x \in \mathcal{H}$. If f is convex, we let $\partial f(x) \triangleq \{u \in \mathcal{H} | f(y) \geq f(x) + \langle y - x, v \rangle \ \forall y \in \mathcal{H}\}$ the subdifferential of f at $x \in \mathrm{dom}(f)$.

Let $\mathcal{C} \subseteq \mathcal{H}$ be a nonempty set. The indicator function $\delta_{\mathcal{C}} : \mathcal{H} \rightarrow \bar{\mathbb{R}}$, is the function satisfying $\delta_{\mathcal{C}}(x) = 0$ if $x \in \mathcal{C}$ and $+\infty$ otherwise. The subdifferential of the indicator function is the normal cone

$$N_{\mathcal{C}}(x) \triangleq \begin{cases} \{\xi \in \mathcal{H} | \langle \xi, y - x \rangle \leq 0 \ \forall y \in \mathcal{C}\} & \text{if } x \in \mathcal{C}, \\ \varnothing & \text{else.} \end{cases}$$

For a set-valued operator $M : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ we denote by $\mathrm{gr}(M) \triangleq \{(x, u) \in \mathcal{H} \times \mathcal{H} | u \in M(x)\}$ its graph, $\mathrm{dom}(M) \triangleq \{x \in \mathcal{H} | M(x) \neq \varnothing\}$ its domain, and by $M^{-1} : \mathcal{H} \rightarrow 2^{\mathcal{H}}$ its

---

[1] Parts of the results reported in this paper are published in the conference proceeding [43]. Besides giving detailed proofs of all results, which partly were missing from the proceedings, we extend the approach by a new splitting scheme with multiple penalty functions, and present applications and numerical examples.

inverse, defined by $(u, x) \in \mathrm{gr}(M^{-1}) \Leftrightarrow (x, u) \in \mathrm{gr}(M)$. We let $\mathrm{zer}(M) = \{x \in \mathcal{H} | 0 \in M(x)\}$ denote the set of zeros of M. A operator M is *monotone* if $\langle x - y, u - v \rangle \geq 0$ for all $(x, u), (y, v) \in \mathrm{gr}(M)$. A monotone operator M is maximally monotone if there exists no proper monotone extension of the graph of M on $\mathcal{H} \times \mathcal{H}$. A single-valued operator $M : \mathcal{H} \to \mathcal{H}$ is *cocoercive* (inverse strongly monotone) if there exists $\mu > 0$ such that

$$\langle M(x) - M(y), x - y \rangle \geq \mu \, \|M(x) - M(y)\|^2 \,.$$

A cocoercive operator is Lipschitz, but the converse is not true in general.

Fact 1 [12, Proposition 23.39] If M is maximally monotone, then $\mathrm{zer}(M)$ is convex and closed.

Fact 2 If M is maximally monotone, then

$$p \in \mathrm{zer}(M) \Leftrightarrow \langle u - p, w \rangle \geq 0 \quad \forall (u, w) \in \mathrm{gr}(M).$$

The resolvent of M, $J_M : \mathcal{H} \to 2^{\mathcal{H}}$ is defined by $J_M = (\mathrm{Id} + M)^{-1}$. If M is maximally monotone, then J_M is single-valued and maximally monotone. In particular, the resolvent of the normal cone mapping $M = N_C$ of a closed convex set $C \subset \mathcal{H}$ is the orthogonal projection Π_C. The support function of a closed set $C \subset \mathcal{H}$ is $\sigma_C(u) = \sup_{y \in C} \langle y, u \rangle$. Clearly, $p \in N_C(u)$ if, and only, if $\sigma_C(p) = \langle p, u \rangle$. According to [12, Proposition 23.31], we have the relation

$$\|J_{\lambda M}(x) - J_{\alpha M}(x)\| \leq |\lambda - \alpha| \cdot \|M_\lambda(x)\| \,, \tag{2}$$

where $M_\lambda \triangleq \frac{1}{\lambda}(\mathrm{Id}_{\mathcal{H}} - J_{\lambda M})$ denotes the *Yosida approximation*, with parameter λ, of the maximally monotone operator M. The Fitzpatrick function associated to a monotone operator M is the convex and lower semi-continuous function defined as

$$\varphi_M(x, u) = \sup_{(y, v) \in \mathrm{gr}(M)} \{\langle x, v \rangle + \langle y, u \rangle - \langle y, v \rangle\}.$$

Lemma 1 *Let $x, y, z \in \mathcal{H}$ and $\alpha \in \mathbb{R}$. Then*

$$2\langle x - y, z - y \rangle = \|x - y\|^2 - \|x - z\|^2 + \|z - y\|^2 \tag{3}$$

$$\|\alpha x + (1 - \alpha) y\|^2 + \alpha(1 - \alpha) \|x - y\|^2 = \alpha \|x\|^2 + (1 - \alpha) \|y\|^2 \tag{4}$$

To assess the global asymptotic stability of the dynamical systems we consider, we recall the following central result (see e.g. [1], Lemma 5.1):

Lemma 2 *Suppose that $F : \mathbb{R}_{\geq 0} \to \mathbb{R}$ is locally absolutely continuous and bounded below and that there exists $G \in L^1(\mathbb{R}_{\geq 0})$ such that*

$$\frac{\mathrm{d}}{\mathrm{d}t} F(t) \leq G(t) \quad a.e. \ t \in \mathbb{R}_{\geq 0}.$$

Then $\lim_{t \to \infty} F(t)$ exists in $\mathbb{R}$.

2.1 Perturbed Solutions and the Central Funnel

We follow a *double penalization* approach to solve the constrained variational inequality (P). We work under following standing assumption:

Assumption 1 The main set of standing hypothesis in this paper is the following:

(i) $A + N_C$ is maximally monotone and $\mathrm{zer}(A + D + N_C) \neq \varnothing$.
(ii) $B : \mathcal{H} \to \mathcal{H}$ is μ-cocoercive (for some $\mu > 0$), and coercive:

$$\lim_{\|x\| \to \infty} \frac{\langle x, B(x) \rangle}{\|x\|} = \infty$$

(iii) $D : \mathcal{H} \to \mathcal{H}$ is maximally monotone and $\frac{1}{\eta}$-Lipschitz.

Since A is maximally monotone and C is a nonempty convex and closed set, $A + N_C$ is maximally monotone if additional regularity conditions are satisfied. A comprehensive list of conditions guaranteeing the maximal monotonicity of the sum of two maximally monotone operators can be found in [12, 13]. In particular, if $\mathrm{dom}(D) = \mathcal{H}$ and $\mathrm{dom}(B) = \mathcal{H}$ (which we assume throughout this manuscript), Corollary 25.5 of [12] implies that the operators $\Phi = A + D + N_C$ and $F_n = A + D + \beta_n B$ are maximally monotone. In particular, $\mathrm{zer}(A + D + N_C)$ is a nonempty closed convex set.

To ensure strong convergence, we follow the classical iterative Tikhonov regularization approach. However, besides strong convergence, we have to respect the constraints imposed on the variational inequality. To enforce these constraints, we augment our operator by a time-varying penalty term which regulates over time the importance we attach to constraint violation of the generated trajectory. The combination of these two dynamic effects leads us to study a family of *auxiliary problems*, formulated as:

Problem 2 Given $(\varepsilon, \beta) \in \mathbb{R}_{\geq 0}^2$ find $x \in \mathcal{H}$ such that

$$0 \in \Phi_{\varepsilon,\beta}(x) \triangleq (A + D + \varepsilon \, \mathrm{Id}_\mathcal{H} + \beta B)(x). \tag{5}$$

The logic behind this model is as follows: Suppose $0 \in \Phi_{\varepsilon,\beta}(\hat{x}_{\varepsilon,\beta})$. On the one hand, the presence of the Tikhonov term $\varepsilon \, \mathrm{Id}_\mathcal{H}$ will attract the vector $\hat{x}_{\varepsilon,\beta}$ towards the origin and, as $\varepsilon \to 0$, $\hat{x}_{\varepsilon,\beta}$ will approach the set of solutions of (P), selecting its least-norm element. On the other hand, the penalization term βB will force $B(\hat{x}_{\varepsilon,\beta}) \to 0$, as $\beta \to \infty$. As a result, the limit of $\hat{x}_{\varepsilon,\beta}$ will be a zero of B, thus an element of C.

In the remainder of this section, we show that solutions of (5) do approximate the least-norm solution of (P), as $(\varepsilon, \beta) \in \mathbb{R}_{\geq 0}^2$, and establish some important structural properties of this approximation. For $(\varepsilon, \beta) \in \mathbb{R}_{\geq 0}^2$, we define the mapping $V_{\varepsilon,\beta} : \mathcal{H} \to \mathcal{H}$ by

$$V_{\varepsilon,\beta}(x) \triangleq D(x) + \varepsilon x + \beta B(x) \tag{6}$$

for all $x \in \mathcal{H}$. Assumption 1 immediately implies the following regularity properties of the parameterized family $(V_{\varepsilon,\beta})_{(\varepsilon,\beta) \in \mathbb{R}_{\geq 0}^2}$.

Lemma 3 *For each $(\varepsilon, \beta) \in \mathbb{R}^2_{\geq 0}$, the mapping $x \mapsto V_{\varepsilon,\beta}(x)$ is Lipschitz continuous with modulus $L_{\varepsilon,\beta} \triangleq \frac{1}{\eta} + \varepsilon + \frac{\beta}{\mu}$. Additionally, if $\varepsilon > 0$, $V_{\varepsilon,\beta}(x)$ is ε-strongly monotone. Furthermore, $\mathrm{dom}(V_{\varepsilon,\beta}) = \mathcal{H}$.*

Hence, the auxiliary problems (5) constitute a family of strongly monotone inclusions. Since $\Phi_{\varepsilon,\beta}$ is strongly monotone, the set $\mathrm{zer}(\Phi_{\varepsilon,\beta})$ is a singleton whose unique element we denote by $\bar{x}(\varepsilon, \beta)$. The function $(\varepsilon, \beta) \mapsto \bar{x}(\varepsilon, \beta)$ maps the positive quadrant of $\mathbb{R}^2$ to a region in $\mathcal{H}$, which we call the *central funnel* for the reasons we will explain in the following.

2.2 Central Paths Approximate the Least-Norm Zero of Φ

Given absolutely continuous functions $\varepsilon, \beta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$, such that $\lim_{t\to\infty} \varepsilon(t) = 0$ and $\lim_{t\to\infty} \beta(t) = \infty$, the curve $t \mapsto \bar{x}(\varepsilon(t), \beta(t))$ is called the *central path* for the approximation of Problem (P) given by (5). Similarly, given positive sequences $(\varepsilon_n)_{n\in\mathbb{N}}$ and $(\beta_n)_{n\in\mathbb{N}}$, such that $\varepsilon_n \to 0$ and $\beta_n \to \infty$, the sequence $\left(\bar{x}(\varepsilon_n, \beta_n)\right)_{n\in\mathbb{N}}$ is a *(discrete) central path*. As shown below, every central path converges strongly to the least-norm zero of Φ, which motivates our choice of the word funnel.

Proposition 4 *Let $(\varepsilon_n)_{n\in\mathbb{N}}$, $(\beta_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}_{>0}$ such that $\varepsilon_n \to 0$, $\beta_n \to \infty$, and $\lim_{n\to\infty} \varepsilon_n \beta_n = \infty$. Then $\bar{x}(\varepsilon_n, \beta_n) \to \Pi_{\mathrm{zer}(\Phi)}(0)$ as $n \to \infty$.*

We prove this proposition via a sequence of Lemmas.

Lemma 5 *Let $(\varepsilon_n)_{n\in\mathbb{N}}$, $(\beta_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}_{>0}$ such that $\varepsilon_n \to 0$, $\beta_n \to \infty$, and $\lim_{n\to\infty} \varepsilon_n \beta_n = \infty$. Denote by $(\bar{x}_n)_{n\in\mathbb{N}}$ the discrete central path obtained by $\bar{x}_n = \bar{x}(\varepsilon_n, \beta_n)$. Any weak limit point of $(\bar{x}_n)_{n\in\mathbb{N}}$ is contained in $\mathcal{C}$.*

Proof For every $n \in \mathbb{N}$, define $\bar{x}_n \equiv \bar{x}(\varepsilon_n, \beta_n)$, the unique root of the set-valued mapping $\Phi_n \equiv \Phi_{\varepsilon_n,\beta_n} : \mathcal{H} \to 2^{\mathcal{H}}$. To simplify the notation, we define the set-valued operator $\mathsf{F}_n : \mathcal{H} \to 2^{\mathcal{H}}$ by

$$\mathsf{F}_n \triangleq \mathsf{A} + \mathsf{D} + \beta_n \mathsf{B}.$$

Hence, for each $n \geq 1$, we have the inclusion

$$
\begin{aligned}
0 \in \mathsf{F}_n(\bar{x}_n) + \varepsilon_n \bar{x}_n &\Leftrightarrow (1 - \varepsilon_n)\bar{x}_n \in (\mathrm{Id}_{\mathcal{H}} + \mathsf{F}_n)(\bar{x}_n) \\
&\Leftrightarrow \bar{x}_n = \mathsf{J}_{\mathsf{F}_n}((1 - \varepsilon_n)\bar{x}_n)
\end{aligned}
$$

At the same time, for every $z \in \mathrm{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$, it holds that $z \in \mathcal{C}$, and therefore $\beta_n \mathsf{B}(z) = 0$ for all $n \geq 1$. This implies that there exists $p \in \mathsf{N}_{\mathcal{C}}(z)$ such that for all $n \in \mathbb{N}$

$$z - p \in (\mathrm{Id}_{\mathcal{H}} + \mathsf{F}_n)(z) \Leftrightarrow z = \mathsf{J}_{\mathsf{F}_n}(z - p).$$

Since F_n is maximally monotone, $\mathsf{J}_{\mathsf{F}_n}$ is firmly non-expansive. Hence,

$$
\begin{aligned}
\|\bar{x}_n - z\|^2 &\leq \langle \mathsf{J}_{\mathsf{F}_n}((1 - \varepsilon_n \bar{x}_n)) - \mathsf{J}_{\mathsf{F}_n}(z - p), (1 - \varepsilon_n)\bar{x}_n - (z - p) \rangle \\
&= \langle \bar{x}_n - z, (1 - \varepsilon_n)\bar{x}_n - (z - p) \rangle
\end{aligned}
$$

Rearranging yields the inequality

$$\langle \bar{x}_n - z, p \rangle \geq \varepsilon_n \langle \bar{x}_n, \bar{x}_n - z \rangle \geq \varepsilon_n \|\bar{x}_n\| (\|\bar{x}_n\| - \|z\|). \tag{7}$$

$\mathsf{J}_{\mathsf{F}_n}$ is also non-expansive, so that

$$\begin{aligned}
\|\bar{x}_n - z\| &\leq \|(1 - \varepsilon_n)\bar{x}_n - z + p\| \\
&= \|(1 - \varepsilon_n)(\bar{x}_n - z) + p - \varepsilon_n z\| \\
&\leq (1 - \varepsilon_n) \|\bar{x}_n - z\| + \|p - \varepsilon_n z\|.
\end{aligned}$$

Hence,

$$\varepsilon_n \|\bar{x}_n - z\| = \|\varepsilon_n \bar{x}_n - \varepsilon_n z\| \leq \|p - \varepsilon_n z\| \qquad \forall n \geq 1.$$

It follows that $(\varepsilon_n \bar{x}_n)_{n \in \mathbb{N}}$ is bounded.

We can also reorganize the equilibrium condition defining the sequence $\bar{x}_n$ and the point z as

$$-\varepsilon_n \bar{x}_n - \beta_n \mathsf{B}(\bar{x}_n) \in \mathsf{A}(\bar{x}_n) + \mathsf{D}(\bar{x}_n), \text{ and} \tag{8}$$

$$-p \in \mathsf{A}(z) + \mathsf{D}(z) \tag{9}$$

Monotonicity of $\mathsf{A} + \mathsf{D}$ implies

$$\beta_n \langle \mathsf{B}(\bar{x}_n), z - \bar{x}_n \rangle \geq \langle \varepsilon_n \bar{x}_n, \bar{x}_n - z \rangle + \langle p, z - \bar{x}_n \rangle. \tag{10}$$

The μ-co-coercivity of the penalty operator B yields

$$\mu \|\mathsf{B}(\bar{x}_n)\|^2 \leq \langle p - \varepsilon_n \bar{x}_n, \frac{\varepsilon_n(\bar{x}_n - z)}{\beta_n \varepsilon_n} \rangle.$$

If $\varepsilon_n \to 0$ and $\varepsilon_n \beta_n \to \infty$ as $n \to \infty$, we conclude from the above relation that

$$\limsup_{n \to \infty} \|\mathsf{B}(\bar{x}_n)\|^2 \leq 0.$$

Since the operator B is assumed to be coercive, the sequence $(\bar{x}_n)_n$ is bounded. Let $\bar{x}_\infty$ be a weak accumulation point of $(\bar{x}_n)_{n \in \mathbb{N}}$. Pick a subsequence (x_{n_j}) with $x_{n_j} \rightharpoonup \bar{x}_\infty$ as $j \to \infty$. (10) implies

$$0 \leq \langle \mathsf{B}(\bar{x}_{n_j}), \bar{x}_{n_j} - z \rangle \leq \frac{\varepsilon_{n_j}}{\beta_{n_j}} \langle \bar{x}_{n_j}, z - \bar{x}_{n_j} \rangle + \frac{1}{\beta_{n_j}} \langle p, \bar{x}_{n_j} - z \rangle.$$

Hence,

$$\limsup_{n \to \infty} \langle \mathsf{B}(\bar{x}_n), \bar{x}_n - z \rangle = 0,$$

and for all $w \in \mathcal{H}$, we deduce

$$\langle \mathsf{B}(w), w - \bar{x}_\infty \rangle = \lim_{j \to \infty} \langle \mathsf{B}(w), w - \bar{x}_{n_j} \rangle \geq \lim_{j \to \infty} \langle \mathsf{B}(\bar{x}_{n_j}), w - \bar{x}_{n_j} \rangle = 0.$$

Fact 2, implies $\bar{x}_\infty \in \mathcal{C}$. □

Next to the feasibility of weak accumulation points of $(\bar{x}_n)_{n\in\mathbb{N}}$, we want these limit points to belong to the set of solutions to (P).

Lemma 6 *Let $(\varepsilon_n)_{n\in\mathbb{N}}, (\beta_n)_{n\in\mathbb{N}}$ be sequences in $\mathbb{R}_{>0}$ such that $\varepsilon_n \to 0, \beta_n \to \infty,$ and $\lim_{n\to\infty} \varepsilon_n\beta_n = \infty$. Any weak limit point of the central path $(\bar{x}_n)_{n\in\mathbb{N}}$ is contained in $\mathrm{zer}(\Phi)$. In particular, $\liminf_{n\to\infty} \|\bar{x}_n\| \geq \inf\{\|z\| : z \in \mathrm{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_\mathcal{C})\}$.*

Proof Let $(u, w) \in \mathrm{gr}(\Phi)$ arbitrary. Then, there exists $\xi \in \mathsf{N}_\mathcal{C}(u)$ such that

$$w - \xi - \mathsf{D}(u) \in \mathsf{A}(u).$$

Together with (8), the monotonicity of A gives

$$\langle -\varepsilon_n\bar{x}_n - \beta_n\mathsf{B}(\bar{x}_n) - \mathsf{D}(\bar{x}_n) - w + \xi + \mathsf{D}(u), \bar{x}_n - u\rangle \geq 0.$$

Whence,

$$\begin{aligned}
\langle w, u - \bar{x}_n\rangle &\geq \varepsilon_n\langle\bar{x}_n, \bar{x}_n - u\rangle + \beta_n\langle\mathsf{B}(\bar{x}_n), \bar{x}_n - u\rangle \\
&\quad + \langle\mathsf{D}(\bar{x}_n) - \mathsf{D}(u), \bar{x}_n - u\rangle + \langle\xi, u - \bar{x}_n\rangle \\
&\geq \varepsilon_n\langle\bar{x}_n, \bar{x}_n - u\rangle + \langle\xi, u - \bar{x}_n\rangle,
\end{aligned}$$

where we have used the monotonicity of B and D, as well as $\mathsf{B}(u) = 0$. If $\bar{x}_\infty$ is a weak accumulation point of $(\bar{x}_n)$, then

$$\langle w, u - \bar{x}_\infty\rangle \geq \langle\xi, u - \bar{x}_\infty\rangle.$$

The right-hand side is nonnegative since $\xi \in \mathsf{N}_\mathcal{C}(u)$ and $\bar{x}_\infty \in \mathcal{C}$, thanks to Lemma 5. Using Fact 2, this means that $\bar{x}_\infty \in \mathrm{zer}(\Phi)$. In particular, $\|\bar{x}_\infty\| \geq \inf\{\|z\| : z \in \mathrm{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_\mathcal{C})\}$ for every weak limit point $\bar{x}_\infty$ of $(\bar{x}_n)_{n\in\mathbb{N}}$. □

With the help of Lemmas 5 and 6, we can proof Proposition 4.

Proof of Proposition 4 Let $z \in \mathrm{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_\mathcal{C})$ and $p \in \mathsf{N}_\mathcal{C}(z)$. Using relation (7), we obtain
$$\langle p, \bar{x}_n - z\rangle \geq \varepsilon_n\|\bar{x}_n\|(\|\bar{x}_n\| - \|z\|).$$

Since
$$\liminf_{n\to\infty}\langle p, \bar{x}_n - z\rangle \leq \limsup_{n\to\infty}\langle p, \bar{x}_n - z\rangle \leq 0,$$

it follows that $\|\bar{x}_n\| \leq \|z\|$ for all n sufficiently large. Hence, $\limsup_{n\to\infty} \|\bar{x}_n\| \leq \|z\|$ for all $z \in \mathrm{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_\mathcal{C})$. Let x, y be two different weak cluster points of $(\bar{x}_n)_{n\in\mathbb{N}}$. Then, $x, y \in \mathrm{zer}(\mathsf{A}+\mathsf{D}+\mathsf{N}_\mathcal{C})$, and there are weakly converging subsequences $\bar{x}_{k_n} \rightharpoonup x$ and $\bar{x}_{l_n} \rightharpoonup y$ for monotonically increasing and diverging sequences $(k_n)_n, (l_n)_n$

in $\mathbb{N}$. Consequently, for n sufficiently large, the Cauchy-Schwarz inequality and the minimum norm property of weak accumulation points imply

$$\left(\frac{x}{\|x\|}, \bar{x}_{k_n}\right) \leq \|\bar{x}_{k_n}\| \leq \|y\|, \text{ and } \left(\frac{y}{\|y\|}, \bar{x}_{l_n}\right) \leq \|\bar{x}_{l_n}\| \leq \|x\|.$$

It follows,

$$\lim_{n\to\infty}\left(\frac{x}{\|x\|}, \bar{x}_{k_n}\right) = \|x\| \leq \|y\| \text{ and } \lim_{n\to\infty}\left(\frac{y}{\|y\|}, \bar{x}_{l_n}\right) = \|y\| \leq \|x\|.$$

Hence, $\|x\| = \|y\| = \inf\{\|z\| : z \in \text{zer}(A + D + N_C)\}$. Since the zero set of a $A+D+N_C$ is closed convex and nonempty, the least-norm element is unique, implying $x = y = \Pi_{\text{zer}(\Phi)}(0)$. This shows that $(\bar{x}_n)_{n\in\mathbb{N}}$ has at most one weak accumulation point. Hence, the sequence $(\bar{x}_n)$ is weakly convergent, and the strong convergence to the least norm solution follows from Lemma 2.51 in [12]. □

The consequence for the continuous central paths is the following:

Corollary 7 *Let $\varepsilon, \beta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ be absolutely continuous functions such that $\lim_{t\to\infty} \varepsilon(t) = 0$ and $\lim_{t\to\infty} \beta(t) = \infty$. Then, $\bar{x}(\varepsilon(t), \beta(t)) \to \Pi_{\text{zer}(\Phi)}(0)$ as $t \to \infty$.*

We next prove that $(\varepsilon, \beta) \mapsto \bar{x}(\varepsilon, \beta)$ is a locally Lipschitz continuous function. This will be a fundamental result in the Lyapunov analysis of the dynamical systems.

Proposition 8 *The solution mapping $(\varepsilon, \beta) \mapsto \bar{x}(\varepsilon, \beta)$ is locally Lipschitz continuous. In particular, for all $\sigma_1 = (\varepsilon_1, \beta_1)$ and $\sigma_2 = (\varepsilon_2, \beta_2)$, we have*

$$\|\bar{x}(\sigma_2) - \bar{x}(\sigma_1)\| \leq \frac{\ell}{\varepsilon_1}\left(|\beta_2 - \beta_1| + |\varepsilon_2 - \varepsilon_1|\right), \tag{11}$$

where $r \triangleq \inf\{\|x\| : x \in \text{zer}(\Phi)\}$, and $\ell \triangleq \max\{r, \sup_{x\in\mathbb{B}(0,r)}\|B(x)\|\}$.

Proof Fix $\beta > 0$ and pick $\varepsilon_1, \varepsilon_2 > 0$, so that $z_1 = \bar{x}(\varepsilon_1, \beta)$ and $z_2 = \bar{x}(\varepsilon_2, \beta)$, denote the corresponding unique solutions of (5). Denoting by $D_\varepsilon \triangleq D + \varepsilon \text{Id}_{\mathcal{H}}$, we obtain

$$-D_{\varepsilon_1}(z_1) - \beta B(z_1) \in A(z_1), \text{ and } -D_{\varepsilon_2}(z_2) - \beta B(z_2) \in A(z_2).$$

Since A is maximally monotone, we have

$$\langle z_1 - z_2, -D_{\varepsilon_1}(z_1) - \beta B(z_1) + D_{\varepsilon_2}(z_2) + \beta B(z_2)\rangle \geq 0.$$

Since D and B are both maximally monotone, we conclude $\langle \varepsilon_1 z_1 - \varepsilon_2 z_2, z_1 - z_2\rangle \leq 0$. Assume first that $\varepsilon_2 > \varepsilon_1$. Then

$$0 \geq \langle \varepsilon_1 z_1 - \varepsilon_2 z_2, z_1 - z_2\rangle = \varepsilon_1 \|z_1 - z_2\|^2 + (\varepsilon_1 - \varepsilon_2)\langle z_2, z_1 - z_2\rangle,$$

which means $(\varepsilon_2 - \varepsilon_1)\langle z_2, z_1 - z_2 \rangle \geq \varepsilon_1 \|z_1 - z_2\|^2$. By Cauchy-Schwarz,

$$(\varepsilon_2 - \varepsilon_1) \|z_2\| \cdot \|z_1 - z_2\| \geq \varepsilon_1 \|z_1 - z_2\|^2,$$

so that

$$\|z_2 - z_1\| \leq \frac{\varepsilon_2 - \varepsilon_1}{\varepsilon_1} \|z_2\|.$$

Next, assuming $\varepsilon_1 > \varepsilon_2$. Then, interchanging the labels in the above inequality, we get

$$\|z_2 - z_1\| \leq \frac{\varepsilon_1 - \varepsilon_2}{\varepsilon_2} \|z_1\|.$$

Hence, $\|\bar{x}(\varepsilon_1, \beta) - \bar{x}(\varepsilon_2, \beta)\| \leq \frac{|\varepsilon_2 - \varepsilon_1|}{\min\{\varepsilon_1, \varepsilon_2\}} \max\{\|z_1\|, \|z_2\|\}$. This shows that $\varepsilon \mapsto \bar{x}(\varepsilon, \beta)$ is locally Lipschitz.

Now, fix $\varepsilon > 0$ and let $\beta_1, \beta_2 > 0$. Denote $z_1 = \bar{x}(\varepsilon, \beta_1)$ and $z_2 = \bar{x}(\varepsilon, \beta_2)$. By definition, we have

$$-\mathsf{D}_\varepsilon(z_1) - \beta_1 \mathsf{B}(z_1) \in \mathsf{A}(z_1), \text{ and } -\mathsf{D}_\varepsilon(z_2) - \beta_2 \mathsf{B}(z_2) \in \mathsf{A}(z_2).$$

It follows that $\beta_2 \langle \mathsf{B}(z_2), z_1 - z_2 \rangle - \beta_1 \langle \mathsf{B}(z_1), z_1 - z_2 \rangle \geq \varepsilon \|z_1 - z_2\|^2$. Now, if $\beta_2 > \beta_1$, then $(\beta_2 - \beta_1)\langle \mathsf{B}(z_1), z_1 - z_2 \rangle + \beta_2 \langle \mathsf{B}(z_2) - \mathsf{B}(z_1), z_1 - z_2 \rangle \geq \varepsilon \|z_1 - z_2\|^2$. Using the monotonicity of B, we conclude

$$\|z_1 - z_2\| \leq \frac{\beta_1 - \beta_2}{\varepsilon} \|\mathsf{B}(z_1)\|.$$

If $\beta_1 > \beta_2$, we repeat the above computation, and obtain

$$\|z_1 - z_2\| \leq \frac{\beta_1 - \beta_2}{\varepsilon} \|\mathsf{B}(z_2)\|.$$

This yields $\|z_1 - z_2\| \leq \frac{|\beta_1 - \beta_2|}{\varepsilon} \max\{\|\mathsf{B}(z_1)\|, \|\mathsf{B}(z_2)\|\}$, which shows that $\beta \mapsto \bar{x}(\varepsilon, \beta)$ is locally Lipschitz, for all $\varepsilon > 0$.

Next, we show the Lipschitz continuity of the bivariate map $(\varepsilon, \beta) \mapsto \bar{x}(\varepsilon, \beta)$. Let $\sigma_1 \triangleq (\varepsilon_1, \beta_1)$ and $\sigma_2 \triangleq (\varepsilon_2, \beta_2)$ with corresponding solutions $\bar{x}(\sigma_1)$ and $\bar{x}(\sigma_2)$. By definition of these points, we have

$$-V_{\sigma_1}(\bar{x}(\sigma_1)) \in \mathsf{A}(\bar{x}(\sigma_1)), \text{ and } -V_{\sigma_2}(\bar{x}(\sigma_2)) \in \mathsf{A}(\bar{x}(\sigma_2)).$$

Hence,

$$\langle \mathsf{D}_{\varepsilon_2}(\bar{x}(\sigma_2) + \beta_2 \mathsf{B}(\bar{x}(\sigma_2))) - \mathsf{D}_{\varepsilon_1}(\bar{x}(\varepsilon_1, \beta_1) - \beta_1 \mathsf{B}(\bar{x}(\varepsilon_1, \beta_1)), \bar{x}(\varepsilon_1, \beta_1) - \bar{x}(\sigma_2) \rangle \geq 0.$$

Rearranging, we obtain

$$\begin{aligned}\langle \varepsilon_2 \bar{x}(\sigma_2) - \varepsilon_1 \bar{x}(\sigma_1), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle &\geq \beta_1 \langle \mathsf{B}(\bar{x}(\sigma_1)), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle \\ &+ \beta_2 \langle \mathsf{B}(\bar{x}(\sigma_2)), \bar{x}(\sigma_2) - \bar{x}(\sigma_1) \rangle.\end{aligned}$$

This gives

$$
\begin{aligned}
\varepsilon_1 \langle \bar{x}(\sigma_2) - \bar{x}(\sigma_1), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle \geq{} & \beta_1 \langle \mathsf{B}(\bar{x}(\sigma_1)), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle \\
& + \beta_2 \langle \mathsf{B}(\bar{x}(\sigma_2)), \bar{x}(\sigma_2) - \bar{x}(\sigma_1) \rangle \\
& - (\varepsilon_2 - \varepsilon_1) \langle \bar{x}(\sigma_2), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle .
\end{aligned}
$$

Hence,

$$
\begin{aligned}
\varepsilon_1 \| \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \|^2 \leq{} & \beta_1 \langle \mathsf{B}(\bar{x}(\sigma_1)), \bar{x}(\sigma_2) - \bar{x}(\sigma_1) \rangle \\
& + \beta_2 \langle \mathsf{B}(\bar{x}(\sigma_2)), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle \\
& + (\varepsilon_2 - \varepsilon_1) \langle \bar{x}(\sigma_2), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle \\
={} & (\beta_1 - \beta_2) \langle \mathsf{B}(\bar{x}(\sigma_1)), \bar{x}(\sigma_2) - \bar{x}(\sigma_1) \rangle \\
& + \beta_2 \langle \mathsf{B}(\bar{x}(\sigma_2)) - \mathsf{B}(\bar{x}(\sigma_1)), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle \\
& + (\varepsilon_2 - \varepsilon_1) \langle \bar{x}(\sigma_2), \bar{x}(\sigma_1) - \bar{x}(\sigma_2) \rangle \\
\leq{} & |\beta_2 - \beta_1| \, \| \mathsf{B}(\bar{x}(\sigma_1)) \| \cdot \| \bar{x}(\sigma_2) - \bar{x}(\sigma_1) \| \\
& + |\varepsilon_2 - \varepsilon_1| \, \| \bar{x}(\sigma_2) \| \cdot \| \bar{x}(\sigma_2) - \bar{x}(\sigma_1) \| .
\end{aligned}
$$

We thus finally arrive at the estimate

$$
\| \bar{x}(\sigma_2) - \bar{x}(\sigma_1) \| \leq \frac{|\beta_2 - \beta_1|}{\varepsilon_1} \| \mathsf{B}(\bar{x}(\sigma_1)) \| + \frac{|\varepsilon_2 - \varepsilon_1|}{\varepsilon_1} \| \bar{x}(\sigma_2) \| . \tag{12}
$$

From the proof of Step (i) of the proof of Proposition 4, we deduce that $\| \bar{x}(\varepsilon, \beta) \| \leq \inf\{ \|x\| : x \in \mathrm{zer}(\Phi) \} \triangleq r$. Hence, defining $\ell \triangleq \max\{ \sup_{x \in \mathbb{B}(0,r)} \| \mathsf{B}(x) \|, r \}$, the claim follows. $\qquad\square$

2.2.1 Differentiability of Central Paths

We now assume that the parameters (ε, β) are defined in terms of real-valued functions $\varepsilon, \beta : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$. In terms of these functions, we define the time-dependent vector field

$$
V_t : \mathbb{R}_{\geq 0} \times \mathcal{H} \to \mathcal{H}, \quad t \mapsto V_t(x) \equiv V_{\varepsilon(t), \beta(t)}(x).
$$

For each t, we obtain the unique solution $\bar{x}(t) \equiv \bar{x}(\varepsilon(t), \beta(t)) \in \mathrm{zer}(\mathsf{A} + V_t)$. From Lemma 3, we know that $V_t : \mathcal{H} \to \mathcal{H}$ is $\varepsilon(t)$-strongly monotone and $L(t)$-Lipschitz continuous, where

$$
L(t) \triangleq L_{\varepsilon(t), \beta(t)} = \frac{1}{\eta} + \varepsilon(t) + \frac{\beta(t)}{\mu}. \tag{13}
$$

Assumption 2 The functions $t \mapsto \varepsilon(t), t \mapsto \beta(t)$ are absolutely continuous and satisfy the following properties:

(i) $t \mapsto \varepsilon(t)$ non-increasing and $\lim_{t \to \infty} \varepsilon(t) = 0$;
(ii) $t \mapsto \beta(t)$ non-decreasing and $\lim_{t \to \infty} \beta(t) = \infty$;
(iii) $\lim_{t \to \infty} \beta(t) \varepsilon(t) = \infty$.

Lemma 9 *Consider the central path $(\bar{x}(t))_{t \geq 0}$. Under Assumption 2, the curve $t \mapsto \bar{x}(t)$ is almost everywhere differentiable, with*

$$\left\| \frac{\mathrm{d}}{\mathrm{d}t} \bar{x}(t) \right\| \leq \frac{\dot{\beta}(t)}{\varepsilon(t)} \| B(\bar{x}(t)) \| + \frac{|\dot{\varepsilon}(t)|}{\varepsilon(t)} \| \bar{x}(t) \| \qquad a.e.\ t \geq 0. \tag{14}$$

Proof Let $v \in \mathcal{H}$ be an arbitrary unit norm vector of the real Hilbert space $\mathcal{H}$. Define the real-valued function $f_v : [0, \infty) \to \mathbb{R}$ by

$$f_v(t) \triangleq \langle v, \bar{x}(t) \rangle.$$

From inequality (12) in the proof of Proposition 8, we deduce that for all $0 \leq t_1 < t_2 \leq T_1 + h$ and $h > 0$,

$$|f_v(t_2) - f_v(t_1)| \leq \| \bar{x}(t_2) - \bar{x}(t_1) \| \leq \frac{|\beta(t_2) - \beta(t_1)|}{\varepsilon(t_1)} \| B(\bar{x}(t_1)) \| + \frac{|\varepsilon(t_2) - \varepsilon(t_1)|}{\varepsilon(t_1)} \| \bar{x}(t_2) \|.$$

Hence, for $t_1 = t \in [0, \infty)$ and $t_2 = t + h > t$, Assumption 2 implies that $\beta(t + h) - \beta(t) \leq h\dot{\beta}(t + h)$ and $\varepsilon(t + h) - \varepsilon(t) \leq h\dot{\varepsilon}(t)$. Using these estimates, we can continue with the above bound

$$|f_v(t + h) - f_v(t)| \leq \| \bar{x}(t + h) - \bar{x}(t) \| \leq \frac{|h\dot{\beta}(t + h)|}{\varepsilon(t)} \| B(\bar{x}(t)) \| + \frac{|h\dot{\varepsilon}(t)|}{\varepsilon(t)} \| \bar{x}(t + h) \|$$

$$\leq h \left(\frac{\dot{\beta}(T_1 + h)}{\varepsilon(0)} + 1 \right) \sup_{t \in [0, T_1 + h]} \max\{ \| B(\bar{x}(t)) \|, \| \bar{x}(t) \| \},$$

for all $t \in [0, T_1]$. Hence, $t \mapsto f_v(t)$ is locally Lipschitz and by the Rademacher theorem (see e.g. [27]) it is almost everywhere Fréchet differentiable, with the almost everywhere derivative $f_v'(t)$ satisfying the bound

$$|f_v'(t)| \leq \frac{\dot{\beta}(t)}{\varepsilon(t)} \| B(\bar{x}(t)) \| + \frac{|\dot{\varepsilon}(t)|}{\varepsilon(t)} \| \bar{x}(t) \| \qquad \text{a.e.}\ t \geq 0. \tag{15}$$

Let $\{e_i\}_i$ be an orthonormal basis of $\mathcal{H}$. This allows us to identify the time derivative $\frac{\mathrm{d}}{\mathrm{d}t} \bar{x}(t)$ with $\frac{\mathrm{d}}{\mathrm{d}t} \bar{x}(t) = \sum_i e_i f_{e_i}'(t)$ for almost every $t \geq 0$. Furthermore, we observe that

$$\left\| \frac{\mathrm{d}}{\mathrm{d}t} \bar{x}(t) \right\| = \sup_{v \in \mathcal{H}: \|v\|=1} |f_v'(t)| \leq \frac{\dot{\beta}(t)}{\varepsilon(t)} \| B(\bar{x}(t)) \| + \frac{|\dot{\varepsilon}(t)|}{\varepsilon(t)} \| \bar{x}(t) \| \qquad \text{a.e.}\ t \geq 0,$$

as stated. $\qquad \square$

3 Penalty-Regulated Dynamical Systems for Constrained Variational Inequalities

Recall that a function $f : [0, b] \to \mathcal{H}$ (where $b > 0$) is said to be absolutely continuous if there exists an integrable function $g : [0, b] \to \mathcal{H}$ such that

$$f(t) = f(0) + \int_0^t g(s)\, ds \quad \forall t \in [0, b].$$

Definition 1 Let $f : \mathbb{R}_{\geq 0} \times \mathcal{H} \to \mathcal{H}$ be a vector field depending on time and space, and let $x_0 \in \mathcal{H}$ be given. We say $x : \mathbb{R}_{\geq 0} \to \mathcal{H}$ is a strong global solution of

$$\begin{cases} \dot{x}(t) = f(t, x(t)) \\ x(0) = x_0 \in \mathcal{H} \end{cases} \tag{D}$$

(i) $x : \mathbb{R}_{\geq 0} \to \mathcal{H}$ is absolutely continuous on each interval $[0, b]$, $0 < b < \infty$;
(ii) $\dot{x}(t) = f(t, x(t))$ for almost every $t \in \mathbb{R}_{>0}$.

Existence and strong uniqueness of non-autonomous systems can be proven by means of the classical Cauchy-Lipschitz Theorem (see e.g. [42, Theorem 54]). To use this, we need to ensure the following properties enjoyed by the vector field $f(t, x)$.

Theorem 10 *Let $f : \mathbb{R}_{\geq 0} \times \mathcal{H} \to \mathcal{H}$ be a given function satisfying:*

(f1) $f(\cdot, x) : \mathbb{R}_{\geq 0} \to \mathcal{H}$ is measurable for each $x \in \mathcal{H}$;
(f2) $f(t, \cdot) : \mathcal{H} \to \mathcal{H}$ is continuous for each $t \geq 0$;
(f3) there exists a function $\ell(\bullet) \in L^1_{loc}(\mathbb{R}_{\geq 0}; \mathbb{R})$ such that

$$\|f(t, x) - f(t, y)\| \leq \ell(t) \|x - y\| \qquad \forall t \in [0, b], \forall b \in \mathbb{R}_{\geq 0}, \forall x, y \in \mathcal{H}; \tag{16}$$

(f4) for each $x \in \mathcal{H}$ there exists a function $\Delta(\bullet) \in L^1_{loc}(\mathbb{R}_{\geq 0}; \mathbb{R})$ such that

$$\|f(t, x)\| \leq \Delta(t) \qquad \forall t \in [0, b], \forall b \in \mathbb{R}_{\geq 0}. \tag{17}$$

Then, the dynamical system (D) admits a unique strong solution $t \mapsto x(t)$, $t \geq 0$.

3.1 Penalty Regulated Forward-Backward Dynamics

In this section we study explicitly the case where the involved single-valued operators D and B are *both* cocoercive.

Assumption 3 The operator $D : \mathcal{H} \to \mathcal{H}$ is η-cocoercive for some $\eta > 0$.

Under this additional assumption on the data, we investigate the asymptotic properties of the following evolution equation

$$\dot{x}(t) = \gamma(t) \left(J_{\lambda(t)A} \left(x(t) - \lambda(t) V_t(x(t)) \right) - x(t) \right), \quad x(0) = x_0 \in \mathcal{H} \tag{18}$$

Remark 1 The dynamical system (18) contains a small generalization compared to (FB) by including the time-scale parameter $t \mapsto \gamma(t)$. In a discrete-time formulation of the dynamical system, the scaling function $t \mapsto \gamma(t)$ can be interpreted as a relaxation parameter of the resulting numerical scheme. Let $\pi_\delta = \{0 = 0 < t_1 < t_2 < \ldots < t_{n^\delta} = T\}$ be a partition of $[0, T]$; let $h_k^\delta \triangleq t_{k+1}^\delta - t_k^\delta$, and let δ be the mesh size given by $\delta \triangleq \max\{h_k^\delta : 0 \le k \le n^\delta - 1\}$. We define a sequence $(X_k^\delta)_{k \in \{0,1,\ldots,n^\delta - 1\}}$ recursively by

$$X_{k+1}^\delta = (1 - \gamma_k^\delta h_k^\delta) X_k^\delta + h_k^\delta \gamma_k^\delta \, \mathsf{J}_{\lambda_k^\delta \mathsf{A}} \left(X_k^\delta - \lambda_k^\delta V_k(X_k^\delta) \right),$$
$$X_0^\delta = x_0 \in \mathcal{H}.$$

This is a relaxed forward-backward splitting method, involving the iteration-dependent mapping $V_k(x) = \mathsf{D}(x) + \varepsilon_k x + \beta_k \mathsf{B}(x)$. A version of this scheme, without relaxation and in presence of only a penalty term, has been studied in [16].

3.1.1 Existence and Uniqueness of Strong Solutions

In this section we verify that the data defining the evolution equation (18) satisfy all conditions stated in Theorem 10. Whence, existence and uniqueness of strong solutions for (18) follow.

Proposition 11 *Consider the dynamical system (18), where the parameter function* $\lambda : \mathbb{R}_{\ge 0} \to \mathbb{R}_{>0}$ *is continuous, and Assumption 3 being in place. Then, for every* $t \ge 0$ *and every* $x, y \in \mathcal{H}$ *we have*

$$\|f(t, x) - f(t, y)\| \le \gamma(t)(2 + \lambda(t)L(t)) \|x - y\|, \text{ and} \tag{19}$$
$$(\forall x \in \mathcal{H})(\forall b > 0), \quad f(\cdot, x) \in L^1([0, b], \mathcal{H}). \tag{20}$$

Proof Properties $(f1)$, $(f2)$ are clearly satisfied. To simplify the verification of the remaining properties, we set $\mathsf{J}_t \equiv \mathsf{J}_{\lambda(t)\mathsf{A}}$ and $R_t(x) \triangleq x - \lambda(t)V_t(x)$, so that $T_t \triangleq \mathsf{J}_t \circ R_t$. The evolution equation (18) admits then the simpler representation $f(t, x) = T_t(x) - x$. Since the resolvent J_t is firmly non-expansive, it follows that

$$\begin{aligned}
\|f(t, x) - f(t, y)\| &= \gamma(t) \|T_t(x) - x - T_t(y) + y\| \\
&\le \gamma(t) \big(\|T_t(x) - T_t(y)\| + \|x - y\| \big) \\
&= \gamma(t) \big(\|\mathsf{J}_t \circ R_t(x) - \mathsf{J}_t \circ R_t(y)\| + \|x - y\| \big) \\
&\le \gamma(t) \big(\|R_t(x) - R_t(y)\| + \|x - y\| \big) \\
&\le \gamma(t) \big(\|x - \lambda(t)V_t(x) - y + \lambda(t)V_t(y)\| + \|x - y\| \big) \\
&\le \gamma(t)(2 + \lambda(t)L(t)) \|x - y\|.
\end{aligned}$$

where $L(t) = \frac{1}{\eta} + \varepsilon(t) + \frac{\beta(t)}{\mu}$.

As $\lambda, \varepsilon, \beta : \mathbb{R}_{\ge 0} \to \mathbb{R}_{>0}$ are continuous on each interval $[0, b]$, where $0 < b < \infty$, we get

$$L_f : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}, \quad L_f(t) = \gamma(t)(2 + \lambda(t)L(t)),$$

which is clearly a locally function. This verifies condition (f3). It remains to establish condition (f4). From the continuity of $\lambda, \varepsilon, \beta$, there exist $\lambda_{min}, \varepsilon_{min}, \beta_{min}$, such that

$$0 < \lambda_{min} < \lambda(t), \quad 0 < \varepsilon_{min} < \varepsilon(t) \text{ and } 0 < \beta_{min} < \beta(t), \quad \forall t \in [0, b].$$

Hence, we have for all $t \in [0, b]$, using the triangle inequality, nonexpansiveness of J_t and eq. (2), we obtain

$$
\begin{aligned}
\|f(t, x)\| &\leq \gamma(t) \|T_t(x) - x\| \leq \gamma(t) \big(\|T_t(x)\| + \|x\| \big) \\
&\leq \gamma(t) \big(\|x\| + \big\| \mathsf{J}_{\lambda(t)\mathsf{A}}(x - \lambda_{min} V_{\varepsilon_{min}, \beta_{min}}(x)) \big\| \big) \\
&\quad + \gamma(t) \big\| \mathsf{J}_{\lambda(t)\mathsf{A}}(x - \lambda(t) V_{\varepsilon(t), \beta(t)}(x)) - \mathsf{J}_{\lambda(t)\mathsf{A}}(x - \lambda_{min} V_{\varepsilon_{min}, \beta_{min}}(x)) \big\| \\
&\leq \gamma(t) \|x\| + \gamma(t) \big\| \mathsf{J}_{\lambda(t)\mathsf{A}}(x - \lambda_{min} V_{\varepsilon_{min}, \beta_{min}}(x)) \big\| \\
&\quad + \gamma(t) \|x - \lambda(t) V_{\varepsilon(t), \beta(t)}(x) - x + \lambda_{min} V_{\varepsilon_{min}, \beta_{min}}(x)\| \\
&\leq \gamma(t) \|x\| + \gamma(t) \big\| \mathsf{J}_{\varepsilon_{min}\mathsf{A}}(x - \lambda_{min} V_{\varepsilon_{min}, \beta_{min}}(x)) \big\| \\
&\quad + \gamma(t)(\lambda(t) - \lambda_{min}) \big\| \mathsf{A}_{\lambda_{min}}(x - \lambda_{min} V_{\varepsilon_{min}, \beta_{min}}(x)) \big\| \\
&\quad + \gamma(t)(\lambda(t) - \lambda_{min}) \|\mathsf{D}(x)\| + \gamma(t)(\lambda(t)\varepsilon(t) - \lambda_{min}\varepsilon_{min}) \|x\| \\
&\quad + \gamma(t)(\lambda(t)\beta(t) - \lambda_{min}\beta_{min}) \|\mathsf{B}(x)\|.
\end{aligned}
$$

Property (f4) follows by integrating. $\qquad\square$

3.1.2 Strong Convergence of the Trajectories to the Least-Norm Solution of (P)

Our asymptotic analysis of the FB-dynamical system (18) relies on Lyapunov techniques, building on the following technical result.

Lemma 12 *Let Assumptions 1 and 3 be in place. Furthermore, we impose the conditions*

$$\lambda(t) < \frac{\eta}{1 + \eta\varepsilon(t)} \text{ and } \lambda(t) < \frac{\mu}{\mu\varepsilon(t) + \beta(t)} \quad \forall t \geq 0. \tag{21}$$

Then, we have

$$2\langle \dot{x}(t), x(t) - \bar{x}(t) \rangle \leq \gamma(t)\lambda(t)\varepsilon(t)(\lambda(t)\varepsilon(t) - 2) \|x(t) - \bar{x}(t)\|^2.$$

Proof Recall, that (18) can be written as $\dot{x}(t) = \gamma(t)(T_t(x) - x)$, with $T_t(x) = x \Leftrightarrow x \in \text{zer}(\Phi_t)$. With this remark in mind, and using the definition of the dynamics, we observe that

$$
\begin{aligned}
2\langle \dot{x}(t), x(t) - \bar{x}(t) \rangle &= \|\dot{x}(t) + x(t) - \bar{x}(t)\|^2 - \|\dot{x}(t)\|^2 - \|x(t) - \bar{x}(t)\|^2 \\
&= \|\gamma(t)(T_t(x(t)) - \bar{x}(t)) + (1 - \gamma(t))(x(t) - \bar{x}(t))\|^2 - \|\dot{x}(t)\|^2 - \|x(t) - \bar{x}(t)\|^2 \\
&= \gamma(t) \|T_t(x(t)) - \bar{x}(t)\|^2 + (1 - \gamma(t)) \|x(t) - \bar{x}(t)\|^2 \\
&\quad - \gamma(t)(1 - \gamma(t)) \|T_t(x(t)) - x(t)\|^2 ; - \|\dot{x}(t)\|^2 - \|x(t) - \bar{x}(t)\|^2
\end{aligned}
$$

where the last equality uses (4). On the other hand,

$$\|(\mathrm{Id}_{\mathcal{H}} - \lambda(t)V_t)(x) - (\mathrm{Id}_{\mathcal{H}} - \lambda(t)V_t)(y)\|^2$$
$$= \|(1 - \lambda(t)\varepsilon(t))(x - y) - \lambda(t)(\mathsf{D}(x) - \mathsf{D}(y) + \beta(t)(\mathsf{B}(x) - \mathsf{B}(y)))\|^2$$
$$= (1 - \lambda(t)\varepsilon(t))^2 \|x - y\|^2$$
$$\quad - 2\lambda(t)(1 - \lambda(t)\varepsilon(t))\langle x - y, \mathsf{D}(x) - \mathsf{D}(y) - \beta(t)(\mathsf{B}(x) - \mathsf{B}(y))\rangle$$
$$\quad + \lambda^2(t) \|\mathsf{D}(x) - \mathsf{D}(y) + \beta(t)(\mathsf{B}(x) - \mathsf{B}(y))\|^2$$

Since D and B are cocoercive, we have $\langle x - y, \mathsf{D}(x) - \mathsf{D}(y)\rangle \geq \eta \|\mathsf{D}(x) - \mathsf{D}(y)\|^2$ and $\langle x - y, \mathsf{B}(x) - \mathsf{B}(y)\rangle \geq \mu \|\mathsf{B}(x) - \mathsf{B}(y)\|^2$. Moreover

$$\|\mathsf{D}(x) - \mathsf{D}(y) + \beta(t)(\mathsf{B}(x) - \mathsf{B}(y))\|^2 \leq 2 \|\mathsf{D}(x) - \mathsf{D}(y)\|^2 + 2\beta(t)^2 \|\mathsf{B}(x) - \mathsf{B}(y)\|^2,$$

so that we obtain

$$\|(\mathrm{Id}_{\mathcal{H}} - \lambda(t)V_t)(x) - (\mathrm{Id}_{\mathcal{H}} - \lambda(t)V_t)(y)\|^2$$
$$\leq (1 - \lambda(t)\varepsilon(t))^2 \|x - y\|^2 + 2\lambda(t)(\lambda(t) - \eta(1 - \lambda(t)\varepsilon(t))) \|\mathsf{D}(x) - \mathsf{D}(y)\|^2$$
$$\quad + 2\lambda(t)\beta(t)(\lambda(t)\beta(t) - \mu(1 - \lambda(t)\varepsilon(t))) \|\mathsf{B}(x) - \mathsf{B}(y)\|^2.$$

Thanks to (21), we remain with

$$\|(\mathrm{Id}_{\mathcal{H}} - \lambda(t)V_t)(x) - (\mathrm{Id}_{\mathcal{H}} - \lambda(t)V_t)(y)\|^2 \leq (1 - \lambda(t)\varepsilon(t))^2 \|x - y\|^2 \qquad \forall t \geq 0.$$

Therefore, using the non-expansiveness of the resolvent, we can continue the previous estimate to obtain

$$2\langle \dot{x}(t), x(t) - \bar{x}(t)\rangle = \gamma(t) \|T_t(x(t)) - T_t(\bar{x}(t))\|^2 + (1 - \gamma(t)) \|x(t) - \bar{x}(t)\|^2$$
$$\quad - \gamma(t)(1 - \gamma(t)) \|T_t(x(t)) - x(t)\|^2 - \|\dot{x}(t)\|^2 - \|x(t) - \bar{x}(t)\|^2$$
$$\leq \gamma(t)\lambda(t)\varepsilon(t)(\lambda(t)\varepsilon(t) - 2) \|x(t) - \bar{x}(t)\|^2.$$

□

The next assumption strengthens the condition spelled out in Lemma 12, and makes sure that both inequalities formulated in eq. (21) are satisfied.

Assumption 4 The function $\lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ is continuous, with

$$\lambda(t) < \frac{1}{1/\eta + \varepsilon(t) + \beta(t)/\mu} \qquad \forall t \geq 0.$$

Theorem 13 *Let* $t \mapsto x(t)$ *be the strong solution of (FB). Let Assumptions 1-4 be in place. Moreover, consider the following scaling relations:*

$$\lim_{t \to \infty} \frac{\dot{\varepsilon}(t)}{\gamma(t)\lambda(t)\varepsilon^2(t)} = 0, \tag{22}$$

$$\lim_{t \to \infty} \frac{\dot{\beta}(t)}{\gamma(t)\lambda(t)\varepsilon^2(t)} = 0, \ and \tag{23}$$

$$\int_0^\infty \gamma(t)\lambda(t)\varepsilon(t)(2 - \lambda(t)\varepsilon(t)) \, \mathrm{d}t = \infty. \tag{24}$$

Then, $x(t) \to \Pi_{\mathrm{zer}(A+D+N_C)}(0)$ as $t \to \infty$.

Proof Set $\theta(t) \triangleq \frac{1}{2} \|x(t) - \bar{x}(t)\|^2$. We then have

$$\dot{\theta}(t) = \langle x(t) - \bar{x}(t), \dot{x}(t) - \frac{\mathrm{d}}{\mathrm{d}t}\bar{x}(t) \rangle$$

$$= \langle x(t) - \bar{x}(t), \dot{x}(t) \rangle - \langle x(t) - \bar{x}(t), \frac{\mathrm{d}}{\mathrm{d}t}\bar{x}(t) \rangle$$

$$\leq \langle x(t) - \bar{x}(t), \dot{x}(t) \rangle + \|x(t) - \bar{x}(t)\| \cdot \left\| \frac{\mathrm{d}}{\mathrm{d}t}\bar{x}(t) \right\|.$$

Setting $\delta(t) \triangleq -\gamma(t)\lambda(t)\varepsilon(t)(\frac{\lambda(t)\varepsilon(t)}{2} - 1)$ and $\Delta(t, 0) \triangleq \int_0^t \delta(s) \, \mathrm{d}s$, Lemma 12 gives

$$\dot{\theta}(t) \leq -2\delta(t)\theta(t) + \|x(t) - \bar{x}(t)\| \cdot \left\| \frac{\mathrm{d}}{\mathrm{d}t}\bar{x}(t) \right\|.$$

Following Lemma 9, we can bound the second addendum to get

$$\dot{\theta}(t) \leq -2\delta(t)\theta(t) + \sqrt{2\theta(t)} \left(\frac{\dot{\beta}(t)}{\varepsilon(t)} \|B(\bar{x}(t))\| + \frac{|\dot{\varepsilon}(t)|}{\varepsilon(t)} \|\bar{x}(t)\| \right).$$

By putting $\varphi(t) \triangleq \sqrt{2\theta(t)}$, we thus finally arrive at the inequality

$$\dot{\varphi}(t) \leq -\delta(t)\varphi(t) + w(t), \tag{25}$$

where (recall that $\dot{\varepsilon}(t) \leq 0$)

$$w(t) \triangleq \frac{\dot{\beta}(t)}{\varepsilon(t)} \|B(\bar{x}(t))\| - \frac{\dot{\varepsilon}(t)}{\varepsilon(t)} \|\bar{x}(t)\|. \tag{26}$$

Introducing the integration factor $\exp(-\Delta(t, 0))$, we thus obtain

$$\varphi(t) \leq \varphi(0) \exp(-\Delta(t, 0)) + \exp(-\Delta(t, t_0)) \int_0^t w(s) \exp(\Delta(s, t_0)) \, \mathrm{d}s.$$

If the integral on the right-hand side is bounded, we are done. Else, apply l'Hôspital's rule, the definition of $\delta(t)$, and conditions (22) and (23), to get

$$\lim_{t \to \infty} \exp(-\Delta(t,0)) \int_0^t w(s) \exp(\Delta(s,0)) \, ds = \lim_{t \to \infty} \frac{w(t) \exp(\Delta(t,0))}{\delta(t) \exp(\Delta(t,0))}$$

$$= \lim_{t \to \infty} \frac{\dot{\varepsilon}(t)}{\delta(t)\varepsilon(t)} \left(\frac{\dot{\beta}(t)}{\dot{\varepsilon}(t)} \|B(\bar{x}(t))\| - \|\bar{x}(t)\| \right) = 0.$$

This shows that $\lim_{t \to \infty} \|x(t) - \bar{x}(t)\| = 0$. Since

$$\left\| x(t) - \Pi_{\mathrm{zer}(A+D+N_C)}(0) \right\| \le \|x(t) - \bar{x}(t)\| + \left\| \bar{x}(t) - \Pi_{\mathrm{zer}(A+D+N_C)}(0) \right\|,$$

the strong convergence claim follows from Corollary 7.   □

Remark 2 Since $\varepsilon(t) \to 0$ and $\beta(t) \to \infty$ as $t \to \infty$, the hypothesis (21) implies that $\lim_{t\to\infty} \lambda(t) = 0$, as well as $\limsup_{t\to\infty} \lambda(t)\beta(t) \le \mu$. We also see that $\lambda(t) < \eta$ for all $t \ge 0$. Hence, the rate of the decay of the step size must be on par with the rate of divergence of the penalty parameter.

Remark 3 The assumptions formulated in Theorem 13 can be satisfied by the following set of functions: $\gamma(t) = \cos(1/t)$, $\lambda(t) = \frac{\lambda}{\beta(t)+\lambda\varepsilon(t)}$, where $\lambda < c \min\{\mu, \eta\}$ for some $c \in (0,1)$ (recommended to be close to 1), as well as $\varepsilon(t) = (t+b)^{-r}$ and $\beta(t) = (t+b)^s$ for $b \ge 1$ and $0 < r < s$. Then, $\delta(t) = \mathcal{O}\big(\frac{\varepsilon^2(t)}{\beta^2(t)}\big) = \mathcal{O}\big((t+b)^{-2(r+s)}\big)$, and consequently we need to impose the restriction $s + r < \frac{1}{2}$ to ensure that $\delta \notin L^1(\mathbb{R}_+)$. Additionally, we compute $\frac{\dot{\varepsilon}(t)}{\gamma(t)\lambda(t)\varepsilon^2(t)} = \mathcal{O}\big((t+b)^{r+s-1}\big)$. This yields $r + s < 1$. Finally, we have $\frac{\dot{\beta}(t)}{\gamma(t)\lambda(t)\varepsilon^2(t)} = \mathcal{O}\big((t+b)^{2(r+s)-1}\big)$, and to make this a bounded sequence, we have the restriction $r + s < \frac{1}{2}$. These conditions together span a region of feasible parameters (r, s) which is nonempty.

3.2 Penalty Regulated Forward-Backward-Forward Dynamics

A critical assumption underlying the forward-backward dynamical system (FB) is the cocoercivity (inverse strong monotonicity) of the single-valued operators D and B. Cocoercivity is guaranteed to hold when the monotone inclusion problem (P) models optimality conditions for constrained convex optimization problems. However, it generically fails in structured monotone splitting problems arising from primal-dual optimality conditions derived from the Fenchel-Rockafellar theorem. Section 5 describes a very general class of splittings illustrating this claim. Motivated by this observation, this section derives a new dynamical system formulation exhibiting multiscale aspects, respecting Tikhonov regularization and penalization. Specifically, the class of dynamical systems we are investigating in this section builds on [24], and extends it to the constrained setting.

We consider the following first-order dynamical system

$$p(t) = J_{\lambda(t)A}(x(t) - \lambda(t)V_t(x(t)),$$
$$\dot{x}(t) = p(t) - x(t) + \lambda(t)[V_t(x(t)) - V_t(p(t))]$$

To obtain a simpler expression of this dynamics, we define the reflection $R_t(x) \triangleq x - \lambda(t)V_t(x)$, and the vector field $f(t, x) : \mathbb{R}_{\geq 0} \times \mathcal{H} \to \mathcal{H}$ given by

$$f(t, x) \triangleq \left(R_t \circ \mathsf{J}_{\lambda(t)\mathsf{A}} \circ R_t \right)(x) - R_t(x) \tag{27}$$

The first-order dynamical system (FBF) is then exactly of the form (D). To prove existence and uniqueness of strong global solutions, we can use the same arguments as in Section 5.1 of [24], based on the Cauchy-Lipschitz theorem for absolutely continuous trajectories. We therefore omit these straightforward derivations.

Remark 4 A discrete-time formulation of the FBF dynamical system can be given along the same lines as in Remark 1. Adopting the notation used there, let $\pi_\delta = \{0 < t_1 < t_2 < \ldots < t_{n^\delta} = T\}$ be a partition of $[0, T]$; let $h_k^\delta \triangleq t_{k+1}^\delta - t_k^\delta$, and let δ be the mesh size given by $\delta \triangleq \max\{h_k^\delta : 0 \leq k \leq n^\delta - 1\}$. We define a pair process $(P_k^\delta, X_k^\delta)_{k \in \{0, 1, \ldots, n^\delta - 1\}}$ recursively by

$$P_k^\delta = \mathsf{J}_{\lambda_k^\delta \mathsf{A}}(X_k^\delta - \lambda_k^\delta V_k(X_k^\delta)),$$
$$X_{k+1}^\delta = P_k^\delta + \lambda_k^\delta(V_k(X_k^\delta) - V_k(P_k^\delta)).$$

This is just a dynamic version of Tseng's modified extragradient method, augmented with penalty terms (as studied in [16]), together with a Tikhonov term (a new ingredient).

3.2.1 Strong Convergence of the Trajectories to the Least-Norm Solution

This section contains the necessary Lyapunov analysis for understanding the long-run behavior of trajectories issued by (FBF). We begin with some technical lemmata.

Lemma 14 *For almost all $t \in \mathbb{R}_{\geq 0}$, we obtain*

$$0 \leq -\|x(t) - p(t)\|^2 + \|x(t) - \bar{x}(t)\|^2 - (1 + 2\lambda(t)\varepsilon(t))\|p(t) - \bar{x}(t)\|^2$$
$$+ 2\lambda(t)\Big\langle V_t(p(t)) - V_t(x(t)), p(t) - \bar{x}(t)\Big\rangle$$

Proof From the definition of the process $t \mapsto p(t)$, we deduce

$$(\mathrm{Id}_{\mathcal{H}} + \lambda(t)\mathsf{A})p(t) \ni x(t) - \lambda(t)V_t(x(t)),$$

it follows,

$$\Phi_t(p(t)) = \mathsf{A}(p(t)) + V(t, p(t)) \ni \frac{x(t) - p(t)}{\lambda(t)} - V_t(x(t)) + V_t(p(t)) = -\frac{\dot{x}(t)}{\lambda(t)}.$$

Recall that $\Phi_t \equiv \Phi_{\varepsilon(t), \beta(t)}$ is $\varepsilon(t)$-strongly monotone, so that $\{\bar{x}(t)\} = \mathrm{zer}(\Phi_t)$. Consequently,

$$\left\langle -\frac{\dot{x}(t)}{\lambda(t)} - 0, p(t) - \bar{x}(t) \right\rangle \geq \varepsilon(t)\|p(t) - \bar{x}(t)\|^2. \tag{28}$$

Using these properties, we obtain

$$
\begin{aligned}
2\lambda(t)\varepsilon(t)\,\|p(t)-\bar{x}(t)\|^2 &\leq 2\langle x(t)-p(t),\,p(t)-\bar{x}(t)\rangle \\
&\quad + 2\lambda(t)\Big\langle V_t(p(t))-V_t(x(t)),\,p(t)-\bar{x}(t)\Big\rangle \\
&= -\|x(t)-p(t)\|^2 + \|x(t)-\bar{x}(t)\|^2 - \|p(t)-\bar{x}(t)\|^2 \\
&\quad + 2\lambda(t)\Big\langle V_t(p(t))-V_t(x(t)),\,p(t)-\bar{x}(t)\Big\rangle,
\end{aligned}
\tag{29}
$$

which completes the proof. $\square$

Lemma 15 *Let $t\mapsto x(t)$ be the strong global solution of ([FBF]). Then*

$$
\langle x(t)-\bar{x}(t),\,\dot{x}(t)\rangle \leq (\lambda(t)L(t)-1)\,\|x(t)-p(t)\|^2 - \lambda(t)\varepsilon(t)\,\|p(t)-\bar{x}(t)\|^2
$$

for almost all $t\geq 0$.

Proof For almost all $t\geq 0$, we have

$$
\begin{aligned}
2\langle x(t)-\bar{x}(t),\,\dot{x}(t)\rangle &= 2\langle x(t)-\bar{x}(t),\,p(t)-x(t)\rangle \\
&\quad + 2\Big\langle x(t)-\bar{x}(t),\,\lambda(t)[V_t(x(t))-V_t(p(t))]\Big\rangle \\
&= \|\bar{x}(t)-p(t)\|^2 - \|\bar{x}(t)-x(t)\|^2 - \|x(t)-p(t)\|^2 \\
&\quad + 2\lambda(t)\Big\langle x(t)-\bar{x}(t),\,V_t(x(t))-V_t(p(t))\Big\rangle
\end{aligned}
$$

combining with Lemma 14, we get

$$
\begin{aligned}
\|\bar{x}(t)-p(t)\|^2 - \|\bar{x}(t)-x(t)\|^2 &\leq -\|x(t)-p(t)\|^2 - 2\lambda(t)\varepsilon(t)\|p(t)-\bar{x}(t)\|^2 \\
&\quad + 2\lambda(t)\Big\langle V_t(p(t))-V_t(x(t)),\,p(t)-\bar{x}(t)\Big\rangle
\end{aligned}
$$

then,

$$
\begin{aligned}
2\langle x(t)-\bar{x}(t),\,\dot{x}(t)\rangle &\leq -2\|x(t)-p(t)\|^2 - 2\lambda(t)\varepsilon(t)\|p(t)-\bar{x}(t)\|^2 \\
&\quad + 2\lambda(t)\Big\langle V_t(p(t))-V_t(x(t)),\,p(t)-\bar{x}(t)\Big\rangle \\
&\quad + 2\lambda(t)\Big\langle x(t)-\bar{x}(t),\,V_t(x(t))-V_t(p(t))\Big\rangle \\
&= -2\|x(t)-p(t)\|^2 - 2\lambda(t)\varepsilon(t)\|p(t)-\bar{x}(t)\|^2 \\
&\quad + 2\lambda(t)\Big\langle x(t)-p(t),\,V_t(x(t))-V_t(p(t))\Big\rangle \\
&\leq -2\|x(t)-p(t)\|^2 - 2\lambda(t)\varepsilon(t)\|p(t)-\bar{x}(t)\|^2 \\
&\quad + 2\lambda(t)\|x(t)-p(t)\|\cdot\|V_t(x(t))-V_t(p(t))\| \\
&\leq -2\|x(t)-p(t)\|^2 - 2\lambda(t)\varepsilon(t)\|p(t)-\bar{x}(t)\|^2
\end{aligned}
$$

$$+ 2\lambda(t)L(t)\|x(t) - p(t)\|^2$$
$$\leq -2(1 - \lambda(t)L(t))\|x(t) - p(t)\|^2 - 2\lambda(t)\varepsilon(t)\|p(t) - \bar{x}(t)\|^2$$

the proof is completed. $\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\qquad\square$

Theorem 16 *Let $t \mapsto x(t)$ be the strong global solution of (FBF). Let Assumptions 1-4 be in place. Furthermore, we impose the following conditions:*

(i) $\lim\limits_{t\to\infty} \int_0^t \delta(s)\,\mathrm{d}s = \infty$, *where* $\delta(t) = \frac{1-\lambda(t)L(t)}{a^2(t)}$, *and*

$$a(t) \triangleq 2 + \frac{1}{\lambda(t)\varepsilon(t)} + \frac{1}{\eta\varepsilon(t)} + \frac{\beta(t)}{\mu\varepsilon(t)}. \tag{30}$$

(ii) $\lim_{t\to\infty} \frac{\dot\varepsilon(t)}{\varepsilon(t)\delta(t)} = 0$ *and* $\lim_{t\to\infty} \frac{\dot\beta(t)}{\varepsilon(t)\delta(t)} = 0$.

Then $x(t) \to \Pi_{A+D+N_C}(0)$ *as* $t \to +\infty$.

Proof Define $\theta(t) = \frac{1}{2}\|x(t) - \bar{x}(t)\|^2$ where $t \geq 0$. From $\bar{x}(t) = \bar{x}(\varepsilon(t), \beta(t))$, we have

$$\dot\theta(t) = \left\langle x(t) - \bar{x}(t), \dot{x}(t) - \frac{\mathrm{d}}{\mathrm{d}t}\bar{x}(t) \right\rangle,$$

where

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{x}(t) = \frac{\partial}{\partial\varepsilon}\bar{x}(\varepsilon(t), \beta(t))\dot\varepsilon(t) + \frac{\partial}{\partial\beta}\bar{x}(\varepsilon(t), \beta(t))\dot\beta(t)$$

combining (14) with Lemma 15, we get

$$\dot\theta = \left\langle x(t) - \bar{x}(t), \dot{x}(t) - \frac{\partial}{\partial\varepsilon}\bar{x}(\varepsilon(t), \beta(t))\dot\varepsilon(t) - \frac{\partial}{\partial\beta}\bar{x}(\varepsilon(t), \beta(t))\dot\beta(t) \right\rangle$$

$$= \langle x(t) - \bar{x}(t), \dot{x}(t)\rangle - \dot\varepsilon(t)\left\langle x(t) - \bar{x}(t), \frac{\partial}{\partial\varepsilon}\bar{x}(\varepsilon(t), \beta(t)) \right\rangle$$

$$- \dot\beta(t)\left\langle x(t) - \bar{x}(t), \frac{\partial}{\partial\beta}\bar{x}(\varepsilon(t), \beta(t)) \right\rangle$$

$$\leq -(1 - \lambda(t)L(t))\|x(t) - p(t)\|^2 - \varepsilon(t)\lambda(t)\|p(t) - \bar{x}(t)\|^2$$

$$- \dot\varepsilon(t)\left\langle x(t) - \bar{x}(t), \frac{\partial}{\partial\varepsilon}\bar{x}(\varepsilon(t), \beta(t)) \right\rangle - \dot\beta(t)\left\langle x(t) - \bar{x}(t), \frac{\partial}{\partial\beta}\bar{x}(\varepsilon(t), \beta(t)) \right\rangle \tag{31}$$

Since $\Phi_t = A + V_t$ is $\varepsilon(t)$-strongly monotone, (28) shows that

$$\lambda(t)\varepsilon(t)\|p(t) - \bar{x}(t)\|^2 \leq \langle x(t) - p(t) + \lambda(t)[V_t(p(t)) - V_t(x(t))], p(t) - \bar{x}(t)\rangle,$$

Using Cauchy-Schwarz, and the $L(t)$-Lipschitz continuity of V_t, we obtain

$$\|p(t) - \bar{x}(t)\| \leq \left(\frac{1}{\lambda(t)\varepsilon(t)} + 1 + \frac{1}{\eta\varepsilon(t)} + \frac{\beta(t)}{\mu\varepsilon(t)} \right)\|x(t) - p(t)\|.$$

It follows that

$$\|x(t) - \bar{x}(t)\| \leq \|x(t) - p(t)\| + \|p(t) - \bar{x}(t)\|$$

$$\leq \left(2 + \frac{1}{\lambda(t)\varepsilon(t)} + \frac{1}{\eta\varepsilon(t)} + \frac{\beta(t)}{\mu\varepsilon(t)}\right)\|x(t) - p(t)\| = a(t)\|x(t) - p(t)\|.$$

For almost all $t \geq 0$, we thus get

$$-\|x(t) - p(t)\|^2 \leq -\frac{1}{a^2(t)}\|x(t) - \bar{x}(t)\|^2. \tag{32}$$

Defining $\varphi \triangleq \sqrt{2\theta}$, we obtain

$$\dot{\theta}(t) = \dot{\varphi}(t)\varphi(t) \leq -\frac{1 - \lambda(t)L(t)}{a^2(t)}\|x(t) - \bar{x}(t)\|^2$$

$$- \dot{\varepsilon}(t)\|x(t) - \bar{x}(t)\| \cdot \|\frac{\partial}{\partial\varepsilon}\bar{x}(\varepsilon(t), \beta(t))\| + \dot{\beta}(t)\|x(t) - \bar{x}(t)\| \cdot \|\frac{\partial}{\partial\beta}\bar{x}(\varepsilon(t), \beta(t))\|$$

$$\leq -\frac{1 - \lambda(t)L(t)}{a^2(t)}\varphi(t)^2 - \dot{\varepsilon}(t)\varphi(t)\frac{\|\bar{x}(\varepsilon(t), \beta(t))\|}{\varepsilon(t)} + \frac{\dot{\beta}(t)}{\varepsilon(t)}\varphi(t)\|\mathsf{B}(\bar{x}(\varepsilon(t), \beta(t)))\|.$$

We define $\delta(t) \triangleq \frac{1 - \lambda(t)L(t)}{a^2(t)}$, and the integrating factor $\Delta(t) \triangleq \int_0^t \delta(s)\mathrm{d}s$. Upon using the simplified notation $\bar{x}(t) \equiv \bar{x}(\varepsilon(t), \beta(t))$, we then continue from the previous display with

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\varphi(t)\exp(\Delta(t))\right) \leq -\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}\exp(\Delta(t))\left(\|\bar{x}(t)\| - \frac{\dot{\beta}(t)}{\dot{\varepsilon}(t)}\|B\bar{x}(t)\|\right)$$

We set $w(t) \triangleq \|\bar{x}(t)\| - \frac{\dot{\beta}(t)}{\dot{\varepsilon}(t)}\|\mathsf{B}(\bar{x}(t))\|$, and subsequently integrate both sides in the previous display from 0 to t, to conclude with

$$0 \leq \varphi(t) \leq \exp(-\Delta(t))\left[\varphi(0) - \int_0^t \left(\exp(\Delta(s))\frac{\dot{\varepsilon}(s)}{\varepsilon(s)}w(s)\right)\mathrm{d}s\right] \tag{33}$$

If $t \mapsto \int_0^t \exp(\Delta(s))\frac{\dot{\varepsilon}(s)}{\varepsilon(s)}w(s)\mathrm{d}s$ happens to be bounded, then we immediately obtain from hypothesis (i) that $\varphi(t) \to 0$. Otherwise, we apply l'Hôpital's rule to get

$$\lim_{t\to\infty}\exp(-\Delta(t))\int_0^t \exp(\Delta(s))\frac{\dot{\varepsilon}(s)}{\varepsilon(s)}w(s)\mathrm{d}s = \lim_{t\to\infty}\frac{\exp(\Delta(t))\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}w(t)}{\delta(t)\exp(\Delta(t))} = \lim_{t\to\infty}\frac{\frac{\dot{\varepsilon}(t)}{\varepsilon(t)}w(t)}{\delta(t)}$$

Additionally, we know from the proof of Proposition 4 that both $t \mapsto \|\mathsf{B}(\bar{x}(t))\|$ and $t \mapsto \|\bar{x}(t)\|$ are both bounded. Furthermore, since $\dot{\varepsilon}(t) \leq 0$ by Assumption 2, we observe that $w(t) \geq 0$. Using conditions (a) and (b), we deduce that $\varphi(t) \to 0$ and therefore $\|x(t) - \bar{x}(t)\| \to 0$. By the triangle inequality $\|x(t) - \Pi_{\mathrm{zer}(\Phi)}(0)\| \leq \|x(t) - \bar{x}(t)\| + \|\bar{x}(t) - \Pi_{\mathrm{zer}(\Phi)}(0)\|$. Using Corollary 7, we conclude $x(t) \to \Pi_{\mathrm{zer}(\Phi)}(0)$ as $t \to +\infty$. $\square$

Remark 5 We give some concrete specifications for functions $\varepsilon(t)$, $\lambda(t)$ and $\beta(t)$ satisfying all conditions for Theorem 16 to hold. Writing Assumption 4 in dynamical terms, we obtain the condition $\lambda(t)L(t) < 1$. We claim that

$$\liminf_{t\to\infty}(1 - \lambda(t)L(t)) = 1 - \limsup_{t\to\infty}\lambda(t)L(t) > 0.$$

Indeed, using the definition of the Lipschitz constant $L(t)$ in (13), we obtain

$$\lambda(t)L(t) = (1/\eta + \varepsilon(t))\lambda(t) + \lambda(t)\beta(t)/\mu,$$

so that $\limsup_{t\to\infty}\lambda(t)L(t) < 1$. Additionally,

$$a(t) = 2 + \frac{1}{\varepsilon(t)}\left(\frac{1}{\lambda(t)} + \frac{1}{\eta} + \frac{\beta(t)}{\mu}\right) = \frac{\lambda(t)(\varepsilon(t) + L(t)) + 1}{\lambda(t)\varepsilon(t)}$$
$$= \frac{L(t)}{\varepsilon(t)}(1 + \frac{\varepsilon(t)}{L(t)} + \frac{1}{L(t)\lambda(t)}) = \mathcal{O}(\beta(t)/\varepsilon(t))$$

using that $L(t) = \mathcal{O}(\beta(t))$. This in turn implies $\delta(t) = \frac{1 - \lambda(t)L(t)}{a^2(t)} = \mathcal{O}(\frac{\varepsilon^2(t)}{\beta^2(t)})$. Hence, $\lim_{t\to\infty}\delta(t) = 0$, and for obtaining $\delta \notin L^1(\mathbb{R}_{\geq 0})$ it suffices to guarantee that $\int_0^\infty \frac{\varepsilon^2(t)}{\beta^2(t)}dt = \infty$. Then,

$$\frac{\dot{\varepsilon}(t)}{\varepsilon(t)\delta(t)} = \frac{\dot{\varepsilon}(t)L^2(t)}{\varepsilon^3(t)}\frac{(1 + \frac{\varepsilon(t)}{L(t)} + \frac{1}{L(t)\lambda(t)})^2}{1 - \lambda(t)L(t)} = \frac{\dot{\varepsilon}(t)\beta^2(t)}{\varepsilon^3(t)}\mathcal{O}(1).$$

It therefore suffices to have $\lim_{t\to\infty}\frac{\dot{\varepsilon}(t)\beta^2(t)}{\varepsilon^3(t)} = 0$. By a similar argument, it is easy to see that $\frac{\dot{\beta}(t)}{\varepsilon(t)\delta(t)} = \frac{\dot{\beta}(t)\beta(t)^2}{\varepsilon(t)^3}\mathcal{O}(1)$. Therefore, it suffices to ensure that $\frac{\dot{\beta}(t)\beta(t)^2}{\varepsilon(t)^3}$ is bounded. Finding such functions is not too difficult.

Assume $\varepsilon(t) = (t + b)^{-r}$, $\beta(t) = (t + b)^s$, where $s, b > 0$ and r is chosen such that $r + s > 0$ and $r < s$. Then $\frac{\varepsilon^2(t)}{\beta^2(t)} = (t + b)^{-2(r+s)}$, and consequently we need to impose the restriction $s + r < \frac{1}{2}$ to ensure that $\delta \notin L^1(\mathbb{R}_+)$. Additionally, we compute $\frac{\dot{\varepsilon}(t)\beta(t)^2}{\varepsilon(t)^3} = -r(t + b)^{2(r+s)-1}$. This yields the same restriction $r + s < \frac{1}{2}$. Finally, $\frac{\dot{\beta}(t)\beta(t)^2}{\varepsilon^3(t)} = s(t + b)^{3(r+s)-1}$, and to make this a bounded sequence, we need to impose the condition $s + r < \frac{1}{3}$. These conditions together span a region of feasible parameters (r, s) which is nonempty.   ◊

4 Multiscale Forward-Backward Dynamics

In this section we extend the forward-backward penalty dynamics to the challenging case in which the feasible set $\mathcal{C}$ can be represented as the set of joint minimizers of two penalty terms. This structural assumption can be motivated by decomposition methods

in optimization and optimal control, in which different structural properties of the feasible set can be incorporated via different penalty functions. To make this concrete, we refine the general problem formulation (P) by assuming that the constrained domain $\mathcal{C}$ is of the form

$$\mathcal{C} = \mathcal{C}_1 \cap \mathcal{C}_2,$$

with $\mathcal{C}_1, \mathcal{C}_2$ closed convex and nonempty sets. This geometry could represent conic domains intersected with an affine subspace, leading to a very generic class of variational problems. To be consistent with our penalty formulation, we impose the following structural assumption on the geometry.

Assumption 5 The constrained domain $\mathcal{C} \subset \mathcal{H}$ admits the representation

$$\mathcal{C} = \operatorname{argmin} \Psi_1 \cap \operatorname{argmin} \Psi_2, \tag{34}$$

with convex potentials $\Psi_1, \Psi_2 : \mathcal{H} \to (-\infty, \infty)$ satisfying

1. $\Psi_1 : \mathcal{H} \to \mathbb{R}$ is convex and L_{Ψ_1}-smooth;
2. $\Psi_2 : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is proper, lower semicontinuous and convex with subdifferential $\partial \Psi_2$;
3. $\Psi_1 + \Psi_2 : \mathcal{H} \to \mathbb{R} \cup \{+\infty\}$ is coercive;
4. The operator $\mathsf{A} + \partial \Psi_2 : \mathcal{H} \to 2^{\mathcal{H}}$ is maximally monotone.

To align the notation with the previously studied penalty dynamics, we set $\mathsf{B}_1 \triangleq \nabla \Psi_1$ and $\mathsf{B}_2 \triangleq \partial \Psi_2$. Note that B_1 is $\frac{1}{L_{\Psi_1}}$-co-coercive and monotone, and B_2 is maximally monotone. Since our penalization framework only uses information on the gradient and subgradients of the penalty potentials, we can assume without loss of generality that $\operatorname{argmin} \Psi_i = \Psi_i^{-1}(0)$ for $i \in \{1, 2\}$. If this is not originally the case, we can always re-shift the graph of the function so that the problem formulation remains the same. To solve the constrained variational inequality (P) within this more structured setup, we propose a forward-backward based dynamical system via a full splitting of the resulting problem. Given positive functions $\lambda(t), \beta(t), \varepsilon(t)$, we assume that the time-varying operator

$$\mathsf{A}_t(x) \triangleq \mathsf{A}(x) + \beta(t) \mathsf{B}_2(x) \tag{35}$$

has an easy-to-compute resolvent mapping $\mathsf{J}_{\lambda(t)\mathsf{A}_t} = (\operatorname{Id}_{\mathcal{H}} + \lambda(t)\mathsf{A}_t)^{-1}$. This evaluation condition is the main reason why we assume that $\mathsf{A} + \partial \Psi_2$ is maximally monotone. We further assume that the resolvent is everywhere single-valued and nonexpansive. Proceeding then in the spirit of the dynamical system (FB), we design a vector field which exploits a full splitting of the operators involved by moving all single-valued operators into the backward step and all set-valued information into the forward step. Hence, we arrive at the following first-order dynamical system

$$\dot{x}(t) + x(t) = \mathsf{J}_{\lambda(t)\mathsf{A}_t}(x(t) - \lambda(t)V_t(x(t))) \tag{SFBP}$$

where

$$V_t(x) \triangleq \mathsf{D}(x) + \varepsilon(t)x + \beta(t)\mathsf{B}_1(x). \tag{36}$$

With the introduction of the time-varying operators A_t and V_t, we achieve a full splitting of the penalized auxiliary problems of the form (5). This is done on purpose to reduce computational costs in the implementation of the dynamics. To the best of our knowledge, the first full splitting dynamics of this kind has been studied in [30] in the potential case and without Tikhonov regularization. We extend their analysis to the monotone operator case and add Tikhonov regularization on top of the operators to induce strong convergence. Moreover, the dynamical system (SFBP) contains the penalty-regulated forward-backward dynamical system (FB) by setting $B_2 = 0$.

We shall prove the convergence of the trajectories associated with (SFBP) under the Attouch-Czarnecki condition [5, 8]:

Assumption 6 (Attouch-Czarnecki condition)

$$(\forall \xi \in \operatorname{ran}(N_C)) : \int_0^\infty \lambda(t)\beta(t)\left[(\Psi_1 + \Psi_2)^*(\frac{\xi}{\beta(t)}) - \sigma_C(\frac{\xi}{\beta(t)})\right] dt < \infty. \quad (37)$$

This condition involves the Fenchel conjugate of the convex function $\Psi_1 + \Psi_2$, defined as

$$(\Psi_1 + \Psi_2)^*(y) \triangleq \sup_{x \in \mathcal{H}}\{\langle y, x \rangle - (\Psi_1(x) + \Psi_2(x))\}$$

The integrability criterion (37) has been exploited heavily in the analysis of penalty regulated dynamical systems. Early references include [6, 16, 17, 38], among many others. It is implied by an Hölderian growth condition of the combined penalty function $\Psi = \Psi_1 + \Psi_2$. A proper convex and lower semi-continuous function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ with $\operatorname{argmin}(f) \neq \varnothing$ is said to satisfy a Hölderian growth condition with exponent $\rho \in (1, 2]$ if for some $\tau > 0$ we have

$$\frac{\tau}{\rho}\operatorname{dist}(x, \operatorname{argmin} f)^\rho \leq f(x) - \min f \quad \forall x \in \mathbb{R}^d.$$

Assuming that Ψ satisfies a Hölderian growth condition around $\operatorname{argmin} \Psi = C$, we have

$$0 \leq \Psi^*(z) - \sigma_C(z) \leq \tau^{1-\rho^*}\frac{1}{\rho^*}\|z\|^{\rho^*} \quad \forall z \in \mathbb{R}^d,$$

where ρ^* is he dual conjugate exponent given by $\frac{1}{\rho} + \frac{1}{\rho^*} = 1$. Therefore, the Hölderian growth condition implies (37) whenever $\beta(t)$ is a function satisfying the integrability condition

$$\int_0^\infty \tau^{1-\rho^*}\|p/\beta(r)\|^{\rho^*} dt = \|p\|^{\rho^*}\tau^{1-\rho^*}\int_0^\infty \lambda(t)\beta(t)^{1-\rho^*} dt < \infty.$$

In the important case where $\Psi(z) = \frac{1}{2}\operatorname{dist}_C(z)^2$, then $\Psi^*(z) - \sigma_C(z) = \frac{1}{2}\|z\|^2$, and thus the integrability condition (37) simplifies to $\int_0^\infty \frac{\lambda(t)}{\beta(t)} dt < \infty$.

General Weak Convergence Analysis

The asymptotic behavior of the multi-penalty system turned out to be more complex. However, we can establish the weak ergodic convergence of the trajectories, following the arguments in [5, 6]. The proof of the following result can be found in Appendix A.

Theorem 17 *Let $\varepsilon, \lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}_{>0}$ be absolutely continuous functions in $L^2(\mathbb{R}_{>0}) \setminus L^1(\mathbb{R}_{>0})$, and such that $\lim_{t\to\infty} \varepsilon(t) = \lim_{t\to\infty} \lambda(t) = 0$ and $\lim_{t\to\infty} \frac{\lambda(t)}{\varepsilon(t)} = \infty$. Suppose, moreover, that $\liminf_{t\to\infty} \lambda(t)\beta(t) > 0$ and, for every $\xi \in \operatorname{ran}(\mathsf{N}_{\mathcal{C}})$, we have*

$$\int_0^\infty \lambda(t)\beta(t) \left[(\Psi_1 + \Psi_2)^*(\frac{\xi}{\beta(t)}) - \sigma_{\mathcal{C}}(\frac{\xi}{\beta(t)}) \right] dt < \infty.$$

Then,

$$\lim_{t\to\infty} \|\mathsf{B}_1(x(t))\| = \lim_{t\to\infty} (\Psi_1 + \Psi_2)(x(t) + \dot{x}(t)) = 0.$$

Define the ergodic trajectory

$$\bar{x}(T) \triangleq \frac{\int_0^T \lambda(s)x(s)\,ds}{\int_0^T \lambda(s)\,ds} \quad (T > 0).$$

Then, $\bar{x}(t)$ converges weakly, as $t \to \infty$, to a point in $\operatorname{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$.

Remark 6 Assume that the function $\Psi \triangleq \Psi_1 + \Psi_2$ are boundedly inf-compact, which means that every set of the form

$$\{x \in \mathcal{H}| \ \|x\| \leq R \wedge (\Psi_1 + \Psi_2)(x) \leq M\},$$

with $R \geq 0$ and $M \in \mathbb{R}$, is relatively compact. Under the assumptions of Theorem 17, since $\lim_{t\to\infty}(\Psi_1 + \Psi_2)(x(t)) = 0$, the convergence of $\{x(t)\}$ to $\operatorname{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$ must be strong. ◊

5 Applications

In this section we describe some prototypical applications of our splitting framework.

5.1 Sparse Optimal Control of Linear Systems

Given $y_0 \in \mathbb{R}^n$ and matrix-valued functions $A : [0, T] \to \mathbb{R}^{n \times n}$, $B : [0, T] \to \mathbb{R}^{n \times m}$ as well as a vector-valued function $c : [0, T] \to \mathbb{R}^n$, consider the control system

$$\dot{y}(t) = Ay(t) + B(t)u(t) + c(t) \quad y(0) = y_0. \tag{38}$$

The process $u \in L^\infty(0, T; \mathbb{R}^m)$ is an open-loop control. We assume that A, B and c are bounded and sufficiently regular, so that the $u \in L^\infty(0, T; \mathbb{R}^m)$, and the system has

a absolutely continuous solution, denoted by $Y_\bullet^{u,y_0} : [0, T] \to \mathbb{R}^n$. We are interested in solving the optimal control problem

$$\min\{\frac{1}{2} \left\| Y_\bullet^{u,y_0} - \bar{y}(\bullet) \right\|_{L^2(0,T;\mathbb{R}^n)}^2 + \alpha_1 \|u\|_{L^2(0,T;R^m)}^2 + \alpha_2 \|u\|_{L^1(0,T;\mathbb{R}^m)}\} \quad (39)$$

where the minimum is taken over the set of admissible controls

$$\mathcal{U} := \{u \in L^\infty(0, T; \mathbb{R}^m) | u(\bullet) \text{ is measurable and } \|u(t)\|_\infty \le 1 \text{ a.e. } t \in [0, T]\}$$

Let $P : [0, T] \to \mathbb{R}^{n \times n}$ denote the resolvent of the matrix equation $\dot{X} = AX$, $X(0) = \text{Id}_{\mathbb{R}^n}$ satisfying $P(t) = \exp(tA)$. Then

$$Y_t^{y_0,u} = P(t)y_0 + P(t) \int_0^t P(s)^{-1}[B(s)u(s) + c(s)]\, ds$$

satisfies

$$\frac{d}{dt} Y_t^{y_0,u} = AY_t^{y_0,u}, \quad Y_0^{y_0,u} = y_0.$$

This in turn is equivalent to

$$S(u, y) + z_0 = 0,$$

where

$$S(u, y)(t) \triangleq -y(t) + P(t) \int_0^t P(s)^{-1} B(s)u(s)\, ds, \text{ and}$$

$$z_0(t) \triangleq P(t)y_0 + P(t) \int_0^t P(s)^{-1} c(s)\, ds$$

Set $\mathcal{H} \triangleq L^2(0, T; \mathbb{R}^m) \times L^2(0, T; \mathbb{R}^n)$. S is a bounded linear operator from $\mathcal{H}$ to $L^2(0, T; \mathbb{R}^n)$, and consequently the function $\Psi_1 : \mathcal{H} \to \mathbb{R}$ defined by

$$\Psi_1(u, y) \triangleq \frac{1}{2} \|S(u, y) + z_0\|_{L^2(0,T;\mathbb{R}^n)}^2$$

is convex and continuously differentiable. Next, define $\Psi_2(u, y) = \delta_{\mathcal{U}}(u)$ to obtain a convex, proper and lower semi-continuous function. Moreover, $(u, y) \in \mathcal{H}$ solves the control system if and only if $(u, y) \in \text{argmin}(\Psi_1 + \Psi_2)$. With this notation,

$$J_1(u, y) \triangleq \frac{1}{2} \|y - \bar{y}\|_{L^2(0,T;\mathbb{R}^n)}^2 + \alpha_2 \|u\|_{L^2(0,T;\mathbb{R}^m)}^2,$$

$$J_2(u, y) \triangleq \alpha_1 \|u\|_{L^1(0,T;\mathbb{R}^m)}$$

so that the optimal control problem becomes

$$\min\{J_1(u, y) + J_2(u, y) : (u, y) \in \text{argmin}(\Psi_1 + \Psi_2)\}$$

With $\mathsf{D} \triangleq \nabla J_1$ and $\mathsf{A} \triangleq \partial J_2$, so that $\mathcal{C} = \operatorname{argmin}(\Psi_1 + \Psi_2)$, we arrive at a constrained variational inequality problem of the form (P).

5.2 Monotone Inclusions Involving Compositions with Linear Continuous Operators

We next show how our method can be applied to solve monotone inclusion problems involving compositions of operators, as proposed by [21, 28] Let $\mathcal{H}$ and $\mathcal{G}$ be real Hilbert spaces. We introduce operators $\mathsf{A}_1 : \mathcal{H} \to 2^{\mathcal{H}}$ and $\mathsf{A}_2 : \mathcal{G} \to 2^{\mathcal{G}}$ which we assume to be maximally monotone. Additionally, we let $L : \mathcal{H} \to \mathcal{G}$ represent a linear continuous operator. Lastly, we consider $\mathsf{D} : \mathcal{H} \to \mathcal{H}$ monotone and and $\frac{1}{\eta}$-Lipschitz continuous operator with $\eta > 0$, and a monotone operator $\mathsf{B} : \mathcal{H} \to \mathcal{H}$ a monotone and $\frac{1}{\mu}$-Lipschitz continuous operator with $\mu > 0$ satisfying $\mathcal{C} = \operatorname{zer}(\mathsf{B}) \neq \varnothing$. The monotone inclusion problem to solve is

$$0 \in \mathsf{A}_1(x) + L^* \circ \mathsf{A}_2 \circ Lx + \mathsf{D}(x) + \mathsf{N}_{\mathcal{C}}(x). \tag{40}$$

This splitting gains relevance in the generic convex optimization model

$$\min_{x \in \mathcal{C}} \{ f(x) + h(x) + g(Lx) \}$$

where $f \in \Gamma_0(\mathcal{H}), g \in \Gamma_0(\mathcal{G})$, $K : \mathcal{H} \to \mathcal{G}$ is a bounded linear operator and $h \in \mathbf{C}_{L_h}^{1,1}(\mathcal{H}; \mathbb{R})$. Via the classical Fenchel-Rockafellar duality, we can transform this problem into the constrained saddle point problem

$$\min_{x \in \mathcal{C}} \max_{y} \{ f(x) + h(x) + \langle Lx, y \rangle - g^*(y) \}$$

which amounts to solving the monotone inclusion problem consisting in finding a pair (x^*, y^*) such that

$$0 \in \partial f(x^*) + \nabla h(x^*) + L^* y + \mathsf{N}_{\mathcal{C}}(x^*)$$
$$0 \in Lx^* - \partial g^*(y^*)$$

Since $\partial g^* = (\partial g)^{-1}$, we can combine these two inclusions to a single one reading as

$$0 \in \partial f(x^*) + \nabla h(x^*) + L^* \partial g(Lx^*) + \mathsf{N}_{\mathcal{C}}(x^*).$$

We thus arrive at an instantiation of problem (40), by identifying $\mathsf{A}_1 = \partial f, \mathsf{A}_2 = \partial g, \mathsf{D} = \nabla h$.

We use the product space approach in order to show that the general problem (40) can be formulated as the monotone inclusion problem (P). To this end, we consider the product space $\mathcal{H} \times \mathcal{G}$ endowed with the inner product

$$\langle (x, y), (x', y') \rangle_{\mathcal{H} \times \mathcal{G}} = \langle x, x' \rangle_{\mathcal{H}} + \langle y, y' \rangle_{\mathcal{G}}$$

and corresponding norm. We define the operators

$$\tilde{A}(x, y) \triangleq A_1(x) \times A_2^{-1}(y), \; \tilde{D}(x, y) \triangleq \begin{pmatrix} D(x) + L^* y \\ -Lx \end{pmatrix}, \; \tilde{B}(x, y) \triangleq \begin{pmatrix} B(x) \\ 0 \end{pmatrix},$$

and for $\tilde{C} \triangleq C \times G = \mathrm{zer}(\tilde{B})$,

$$N_{\tilde{C}}(x, v) = N_C(x) \times \{0\}.$$

One can easily show that if $(x, v) \in \mathrm{zer}(\tilde{A} + \tilde{D} + N_{\tilde{C}})$, then $x \in \mathrm{zer}(A_1 + L^* A_2 L + D + N_C)$. Conversely, when $x \in \mathrm{zer}(A_1 + L^* A_2 L + D + N_C)$, then there exists $v \in A_2(Lx)$ such that $(x, v) \in \mathrm{zer}(\tilde{A} + \tilde{D} + N_{\tilde{C}})$. Thus, determining the zeros of operator $\tilde{A} + \tilde{D} + N_{\tilde{C}}$ will provide a solution for the monotone inclusion problem (40).

$\tilde{A}$ is maximally monotone [12, Proposition 20.23], $\tilde{D}$ is monotone and $\tilde{\eta}$-Lipschitz continuous, where $\tilde{\eta} = \sqrt{2(1/\eta^2 + \|K\|^2)}$, and $\tilde{B}$ is monotone and $(1/\mu)$-Lipschitz continuous. We can thus directly use our dynamical system to determine zeros of $\tilde{A} + \tilde{D} + N_{\tilde{C}}$. We write the trajectory in terms of pairs $t \mapsto (p(t), q(t))$ and $t \mapsto (x(t), y(t))$ given by

$$p(t) = J_{\lambda(t)A_1}(x(t) - \lambda(t)(D(x(t)) + \varepsilon(t)x(t) + L^* y(t) + \beta(t)B(x(t))))$$

$$q(t) = J_{\lambda(t)A_2^{-1}}(y(t) + \lambda(t)Lx(t) - \lambda(t)\varepsilon(t)y(t))$$

$$\dot{x}(t) = (1 - \lambda(t)\varepsilon(t))(p(t) - x(t))$$
$$\qquad + \lambda(t)[D(x(t)) - D(p(t)) + \beta(t)(B(x(t)) - B(p(t))) + L^*(y(t) - q(t))]$$

$$\dot{y}(t) = (1 - \lambda(t)\varepsilon(t))(q(t) - y(t)) + \lambda(t)L(p(t) - x(t)).$$

Remark 7 Let us underline the fact that, even in the situation when B is cocoercive and, hence, $\tilde{B}$ is cocoercive, the forward-backward penalty scheme studied in [17] cannot be applied in this context, because the operator $\tilde{D}$ is definitely not cocoercive. This is due to the presence of the skew operator $(x, y) \mapsto (K^* y, Kx)$ in its definition. This fact provides a good motivation for formulating, along the forward-backward penalty scheme, a forward-backward-forward penalty scheme for the monotone inclusion problem investigated in this paper.

5.2.1 Application to Linear Inverse Problems

Building on the primal-dual splitting approach of [18, 21], we consider a linear inverse problem with forward operator $K : \mathbb{R}^n \to \mathbb{R}^m$ which is the problem of finding $\theta \in \mathbb{R}^n$ that solves the linear system

$$K\theta = b$$

Typically, this linear system is ill-posed, and therefore a regularization framework is adopted. A popular formulation is to consider flattened gradient via an isotropic total

variation regularization, which reads as the simple bilevel optimization problem

$$\min_{\theta \in [0,1]^n} \mathsf{TV}(\theta) \quad \text{s.t.:} \ \theta \in S := \operatorname*{argmin}_{\theta' \in \mathbb{R}^n} \left\{ \frac{1}{2} \left\| K\theta' - b \right\|^2 \right\} \tag{41}$$

where the mapping $\mathsf{TV} : \mathbb{R}^n \to \mathbb{R}$ is defined as

$$\mathsf{TV}(\theta) = \sum_{i=1}^{M-1} \sum_{j=1}^{N-1} \sqrt{(\theta_{i+1,j} - \theta_{i,j})^2 + (\theta_{i,j+1} - \theta_{i,j})^2} + \sum_{i=1}^{M-1} |\theta_{i+1,N} - \theta_{i,N}|$$
$$+ \sum_{j=1}^{N-1} |\theta_{M,j+1} - \theta_{M,j}|$$

and $\theta_{i,j}$ denotes the normalized value of the pixel located in the i-th row and the j-th column, for $i \in \{1, \ldots, M\}$ and $j \in \{1, \ldots, N\}$. Let $\mathcal{H} = \mathbb{R}^n$ and $\mathcal{Y} = \mathbb{R}^n \times \mathbb{R}^m$. Define the linear operator $L : \mathbb{R}^n \to \mathcal{Y}$ by $\theta \mapsto (L_1\theta, L_2\theta) \in \mathcal{Y}$ defined coordinate-wise by

$$L_1\theta_{i,j} = \begin{cases} \theta_{i+1,j} - \theta_{i,j} & \text{if } i < M, \\ 0 & \text{else.} \end{cases} \quad , \quad L_2\theta_{i,j} = \begin{cases} \theta_{i,j+1} - \theta_{i,j} & \text{if } j < N, \\ 0 & \text{else.} \end{cases}$$

L represents a discretization of the gradient using Neumann boundary conditions. We note that $\|L\|^2 \leq 8$.

For $(y, z), (u, v) \in \mathcal{Y}$, we introduce the inner product

$$\langle (y, z), (u, v) \rangle := \sum_{i=1}^{M} \sum_{j=1}^{N} (y_{i,j} u_{i,j} + z_{i,j} v_{i,j}),$$

with the corresponding norm $\|(y, z)\|_{\mathcal{Y}} = \sqrt{\langle (y, z), (y, z) \rangle}$. It then follows $\mathsf{TV}(\theta) = \|L\theta\|_{\mathcal{Y}}$. The dual norm to $\|\cdot\|_{\mathcal{Y}}$ is defined as

$$\|(u, v)\|_{\mathcal{Y},*} := \sup_{\|(u,v)\|_{\mathcal{Y}} \leq 1} \langle (y, z), (u, v) \rangle.$$

Accordingly, we define the dual space $\mathcal{Y}^*$, as the Euclidean space $\mathcal{Y}$ endowed with the norm $\|\cdot\|_{\mathcal{Y}}$.

Define the function $f(\theta) = \delta_{[0,1]^n}(\theta)$ and $g(u, v) = \|(u, v)\|_{\mathcal{Y}}$. It follows that (41) is representable as

$$\min_{\theta \in \mathcal{H}} \{ f(\theta) + g(L\theta) \} \quad \text{s.t.:} \ \theta \in S := \operatorname*{argmin}_{\theta' \in \mathcal{H}} \{ \frac{1}{2} \left\| K\theta' - b \right\|^2 \}.$$

The Fenchel-Rockafellar dual approach gives us the saddle point bilevel problem

$$\min_{\theta \in \mathcal{H}} \max_{(u,v) \in \mathcal{Y}^*} \{f(\theta) + \langle L\theta, (u,v) \rangle - g^*(u,v)\} \quad \text{s.t.: } \theta \in S := \operatorname*{argmin}_{\theta' \in \mathcal{H}} \{\frac{1}{2} \left\| K\theta' - b \right\|^2\}.$$

(42)

where

$$g^*(u,v) = \delta_M(u,v), \ M = \{(u,v) \in \mathcal{Y} | \ \|(u,v)\|_{\mathcal{Y},*} \le 1\}.$$

This yields the optimality conditions

$$0 \in \partial f(\bar{\theta}) + L^*(\bar{u}, \bar{v}) + \mathsf{N}_S(\bar{\theta})$$
$$0 \in \partial g^*(\bar{u}, \bar{v}) - L\bar{\theta}$$

Define $\mathcal{C} = S \times \mathcal{Y}$, so that $\mathsf{N}_{\mathcal{C}}(\theta, u, v) = \mathsf{N}_S(\theta) \times \{\mathbf{0}_{\mathcal{Y}}\}$, to obtain the monotone inclusion

$$0 \in \mathsf{A}(\bar{\theta}, \bar{u}, \bar{v}) + \mathsf{D}(\bar{\theta}, \bar{u}, \bar{v}) + \mathsf{N}_{\mathcal{C}}(\theta, u, v).$$

where D is the skew symmetric linear operator $\mathsf{D}(\theta, u, v) = [L^*(u, v), -L\theta]$. To solve this problem with our penalty regularized dynamical system, we relax the variational problem to arrive the unconstrained min-max optimization formulation

$$\min_{\theta \in \mathcal{H}} \max_{y \in \mathcal{Y}} \{f(\theta) + \langle L\theta, (u,v) \rangle - g^*(u,v) + \frac{\varepsilon}{2} \|\theta\|^2 - \frac{\varepsilon}{2} \|(u,v)\|_{\mathcal{Y}}^2 + \beta\Psi(\theta, u, v)\},$$

(43)

where $\Psi(\theta, u, v) \triangleq \frac{1}{2} \|K\theta - b\|^2$. Define the monotone and co-coercive operator

$$\mathsf{B}(\theta, u, v) \triangleq \nabla\Psi(\theta, u, v) = [K^*(K\theta - b); \mathbf{0}_{\mathcal{Y}}] \in \mathcal{H} \times \mathcal{Y},$$

so that $\mathcal{C} = \operatorname{zer}(B)$. We thus can approach the solution of our linear inverse problem with the outer penalization scheme using the monotone operator

$$\Phi_{\varepsilon,\beta}(\theta, u, v) = \mathsf{A}(\theta, u, v) + \mathsf{D}(\theta, u, v) + \varepsilon[\theta, u, v] + \beta\mathsf{B}(\theta, u, v).$$

For the implementation of the algorithm, we use the formulas

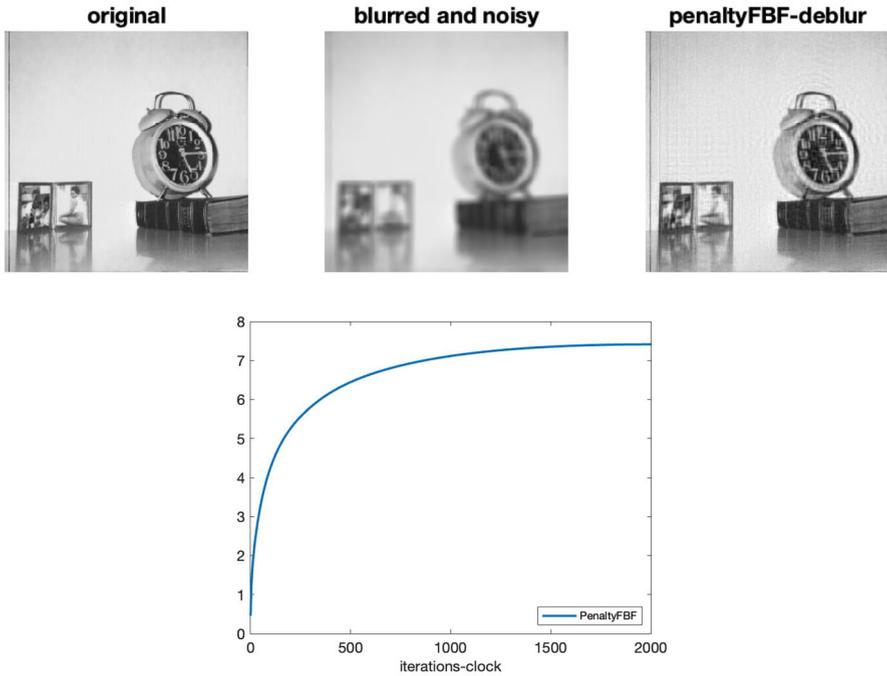$$\mathsf{J}_{\lambda\partial f} = \Pi_{[0,1]^n}, \ \mathsf{J}_{\lambda\partial g^*} = \Pi_M,$$

where $\Pi_S : \mathcal{Y} \to M$ is defined componentwise as [21]

$$(u_{i,j}, v_{i,j}) \mapsto \frac{(p_{i,j}, q_{i,j})}{\max\{1, \sqrt{p_{i,j}^2 + q_{i,j}^2}\}} \quad \forall 1 \le i \le M, 1 \le j \le N.$$

Writing out the iterations of the FBF penalty system, we construct two absolutely continuous functions $p(t) = [\tilde{\theta}(t), \tilde{u}(t), \tilde{v}(t)]$ and $x(t) = [\theta(t), u(t), v(t)]$ solving the following system of ODEs

$$\tilde{\theta}(t) = \Pi_{[0,1]^n}[\theta(t) - \lambda(t)L^*(u(t), v(t)) - \lambda(t)\varepsilon(t)\theta(t) - \lambda(t)\beta(t)K^*(K\theta(t) - b)],$$

original   blurred and noisy   penaltyFBF-deblur

Fig. 1 The figure shows that original image, the blurred image, and the reconstructed image, as well as the evolution of the ISNR for the penalty-regulated FBF dynamical system (FBF)

$$\begin{pmatrix} \tilde{u}(t) \\ \tilde{v}(t) \end{pmatrix} = \Pi_M \left[\begin{pmatrix} u(t) \\ v(t) \end{pmatrix} + \lambda(t) \begin{pmatrix} L_1\theta(t) \\ L_2\theta(t) \end{pmatrix} - \lambda(t)\varepsilon(t) \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} \right],$$

$$\dot{\theta}(t) + \theta(t) = \tilde{\theta}(t) + \lambda(t) \Big[L^*(u(t) - \tilde{u}(t), v(t) - \tilde{v}(t))$$

$$+ \varepsilon(t)(\theta(t) - \tilde{\theta}(t)) + \beta(t)K^*K(\theta(t) - \tilde{\theta}(t)) \Big],$$

$$\begin{pmatrix} \dot{u}(t) \\ \dot{v}(t) \end{pmatrix} + \begin{pmatrix} u(t) \\ v(t) \end{pmatrix} = \begin{pmatrix} \tilde{u}(t) \\ \tilde{v}(t) \end{pmatrix} - \lambda(t) \begin{pmatrix} L_1(\theta(t) - \tilde{\theta}(t)) \\ L_2(\theta(t) - \tilde{\theta}(t)) \end{pmatrix} + \lambda(t)\varepsilon(t) \begin{pmatrix} u(t) - \tilde{u}(t) \\ v(t) - \tilde{v}(t) \end{pmatrix}.$$

For the numerical experiments, we considered two different test images, discretised on a grid of size 256×256, and constructed a blurred and noisy image by making first use of a Gaussian blur operator of size 9×9 and standard deviation 4. Afterwards, we've added a zero mean white Gaussian noise with standard deviation 10^{-3}. The obtained numerical results are illustrated in Figure 1 and 2.
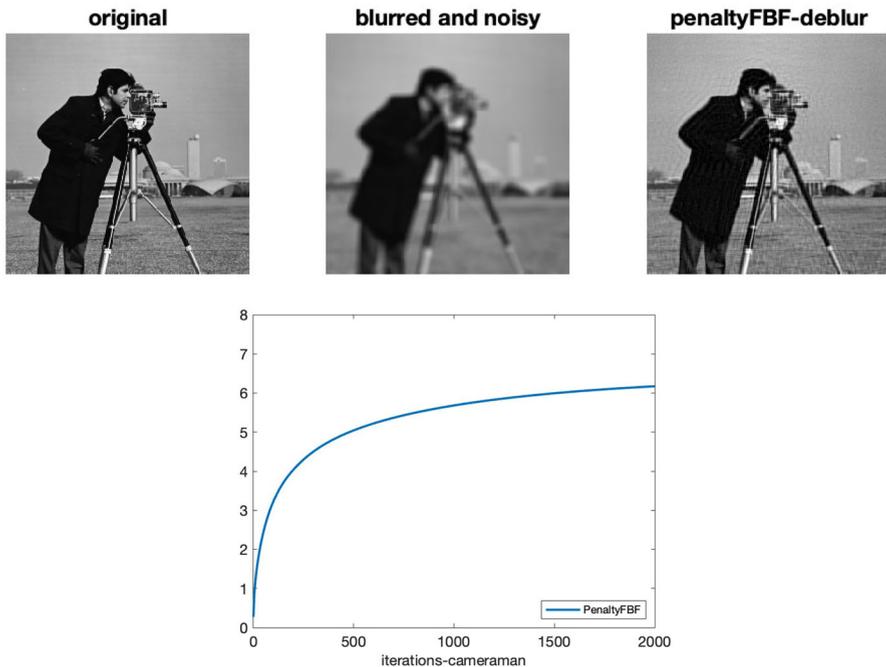
Fig. 2 The figure shows that original image, the blurred image, and the reconstructed image, as well as the evolution of the ISNR for the penalty-regulated FBF dynamical system (FBF)

6 Conclusion

The asymptotic analysis of dynamical systems derived from operator splitting problems has been a very productive line of research over the past 20 years. In this paper we develop a family of dynamical systems featuring Tikhonov regularization and penalty effects. Tikhonov regularization induces strong convergence towards the minimum norm solution, whereas the penalty term steers the system to satisfy constraints subjected to the variational problem we aim to solve. We prove the asymptotic convergence in three paradigmatic settings: (i) Operator splitting with cocoercive data, (ii) Operator splitting with monotone and Lipschitz data, and (iii) Operator splitting with multiple penalty terms. Future directions of research include the extension to stochastic operator equations, such as [34]. Other potentially interesting directions would be the inclusion of inertia or momentum effects into the dynamical system in second order investigate the potential for acceleration. We leave these interesting directions for future research.

Declarations

Competing interests The authors have no relevant financial or non-financial interests to disclose.

Appendix A   Proof of Theorem 17

Pick $(u, w) \in \mathrm{gr}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$ so that $w = a + \mathsf{D}(u) + \xi$, for $a \in \mathsf{A}(u)$ and $\xi \in \mathsf{N}_{\mathcal{C}}(u)$. From the definition of the forward-backward dynamical system, we have

$$-\frac{1}{\lambda(t)}\dot{x}(t) - V_t(x(t)) - \beta(t)b_2(t) \in \mathsf{A}(x(t) + \dot{x}(t))$$

for some $b_2(t) \in \mathsf{B}_2(x(t) + \dot{x}(t)) = \partial \Psi_2(x(t) + \dot{x}(t))$. Combined with $a \in \mathsf{A}(u)$, we obtain

$$\langle a + \frac{1}{\lambda(t)}\dot{x}(t) + V_t(x(t)) + \beta(t)b_2(t), u - x(t) - \dot{x}(t)\rangle \geq 0$$

Rearranging this, we arrive at

$$\langle \dot{x}(t), x(t) + \dot{x}(t) - u\rangle \leq \lambda(t)\langle a + \mathsf{D}(x(t)) + \varepsilon(t)x(t) + \beta(t)\mathsf{B}_1(x(t)), u - x(t) - \dot{x}(t)\rangle \\ + \lambda(t)\beta(t)\langle b_2(t), u - \dot{x}(t) - x(t)\rangle$$

The subgradient inequality yields

$$0 = \Psi_2(u) \geq \Psi_2(x(t) + \dot{x}(t)) + \langle b_2(t), u - x(t) - \dot{x}(t)\rangle \\ \Leftrightarrow -\Psi_2(x(t) + \dot{x}(t)) \geq \langle b_2(t), u - x(t) - \dot{x}(t)\rangle$$

Hence, we can continue the previous display as

$$\langle \dot{x}(t), x(t) + \dot{x}(t) - u\rangle \leq \lambda(t)\langle a + \mathsf{D}(x(t)) + \beta(t)\mathsf{B}_1(x(t)) + \varepsilon(t)x(t), u - \dot{x}(t) - x(t)\rangle \\ - \lambda(t)\beta(t)\Psi_2(x(t) + \dot{x}(t)).$$

Therefore,

$$\frac{\mathrm{d}}{\mathrm{d}t}\|x(t) - u\|^2 = 2\langle \dot{x}(t), x(t) - u \rangle$$

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

$$\leq 2\lambda(t)\langle a + \mathsf{D}(x(t)) + \varepsilon(t)x(t), u - \dot{x}(t) - x(t) \rangle \qquad \text{(A1)}$$

$$- 2\lambda(t)\beta(t)\Psi_2(x(t) + \dot{x}(t)) - 2\|\dot{x}(t)\|^2$$

$$+ 2\lambda(t)\beta(t)\langle \mathsf{B}_1(x(t)), u - \dot{x}(t) - x(t) \rangle$$

Since B_1 is $\frac{1}{L_{\Psi_1}}$-cocoercive, we have

$$\langle \mathsf{B}_1(x(t)), x(t) - u \rangle \geq \frac{1}{L_{\Psi_1}}\|\mathsf{B}_1(x(t))\|^2 \qquad \forall u \in \mathcal{C}. \qquad \text{(A2)}$$

Additionally, the convex gradient inequality yields

$$0 = \Psi_1(u) \geq \Psi_1(x(t)) + \langle \mathsf{B}_1(x(t)), u - x(t) \rangle \qquad \text{(A3)}$$

Performing a convex combination of these two inequalities, we obtain for all $c_1 > 0$,

$$\langle \mathsf{B}_1(x(t)), x(t) - u \rangle \geq \frac{1}{(1 + c_1)L_{\Psi_1}}\|\mathsf{B}_1(x(t))\|^2 + \frac{c_1}{1 + c_1}\Psi_1(x(t)). \qquad \text{(A4)}$$

Next, take $c_0 > 0$, and observe

$$0 \leq \frac{1}{1 + c_0}\|\dot{x}(t) + (1 + c_0)\lambda(t)\beta(t)\mathsf{B}_1(x(t))\|^2$$

$$= \frac{1}{1 + c_0}\|\dot{x}(t)\|^2 + (1 + c_0)\lambda(t)^2\beta(t)^2\|\mathsf{B}_1(x(t))\|^2 + 2\lambda(t)\beta(t)\langle \dot{x}(t), \mathsf{B}_1(x(t)) \rangle.$$

Hence,

$$2\lambda(t)\beta(t)\langle \dot{x}(t), \mathsf{B}_1(x(t)) \rangle \geq -\frac{1}{1 + c_0}\|\dot{x}(t)\|^2 - (1 + c_0)\lambda(t)^2\beta(t)^2\|\mathsf{B}_1(x(t))\|^2.$$
$$\text{(A5)}$$

On the other hand, the descent lemma for functions with a Lipschitz continuous gradient yields

$$\Psi_1(x(t) + \dot{x}(t)) \leq \Psi_1(x(t)) + \langle \mathsf{B}_1(x(t)), \dot{x}(t) \rangle + \frac{L_{\Psi_1}}{2}\|\dot{x}(t)\|^2,$$

so that

$$2\lambda(t)\beta(t)\langle \mathsf{B}_1(x(t)), \dot{x}(t) \rangle \geq 2\lambda(t)\beta(t)[\Psi_1(x(t) + \dot{x}(t)) - \Psi_1(x(t))] \qquad \text{(A6)}$$
$$- L_{\Psi_1}\lambda(t)\beta(t)\|\dot{x}(t)\|^2$$

A convex combination of (A5) with (A6) shows that

$$2\lambda(t)\beta(t)\langle \mathsf{B}_1(x(t)), \dot{x}(t)\rangle \geq -\frac{1+c_0}{1+c_1}\lambda(t)^2\beta(t)^2 \|\mathsf{B}_1(x(t))\|^2$$
$$- \left(\frac{1}{(1+c_1)(1+c_0)} + \frac{c_1 L_{\Psi_1}\lambda(t)\beta(t)}{1+c_1}\right)\|\dot{x}(t)\|^2 \quad (A7)$$
$$+ \frac{2c_1\lambda(t)\beta(t)}{1+c_1}\left(\Psi_1(x(t)+\dot{x}(t)) - \Psi_1(x(t))\right).$$

Therefore, adding (A4) to (A7), we arrive at

$$2\lambda(t)\beta(t)\langle \mathsf{B}_1(x(t)), x(t)+\dot{x}(t) - u\rangle$$
$$\geq \lambda(t)\beta(t)\left(\frac{2}{(1+c_1)L_{\Psi_1}} - \frac{1+c_0}{1+c_1}\lambda(t)\beta(t)\right)\|\mathsf{B}_1(x(t))\|^2$$
$$+ \frac{2c_1\lambda(t)\beta(t)}{1+c_1}\Psi_1(x(t)+\dot{x}(t)) - \left(\frac{1}{(1+c_1)(1+c_0)} + \frac{c_1 L_{\Psi_1}\lambda(t)\beta(t)}{1+c_1}\right)\|\dot{x}(t)\|^2$$
$$\tag{A8}$$

Plugging this into (A1), we can continue this thread as

$$\frac{\mathrm{d}}{\mathrm{d}t}\|x(t)-u\|^2 \leq 2\lambda(t)\langle a + \mathsf{D}(x(t)) + \varepsilon(t)x(t), u - \dot{x}(t) - x(t)\rangle$$
$$- 2\lambda(t)\beta(t)\Psi_2(x(t)+\dot{x}(t)) - \frac{2c_1\lambda(t)\beta(t)}{1+c_1}\Psi_1(x(t)+\dot{x}(t))$$
$$- \lambda(t)\beta(t)\left(\frac{2}{(1+c_1)L_{\Psi_1}} - \frac{1+c_0}{1+c_1}\lambda(t)\beta(t)\right)\|\mathsf{B}_1(x(t))\|^2$$
$$+ \left(\frac{1}{(1+c_1)(1+c_0)} + \frac{c_1 L_{\Psi_1}\lambda(t)\beta(t)}{1+c_1} - 2\right)\|\dot{x}(t)\|^2$$

We have
$$2\lambda(t)\beta(t)\Psi_2(x(t)+\dot{x}(t)) \geq \frac{2c_1\lambda(t)\beta(t)}{1+c_1}\Psi_2(x(t)+\dot{x}(t)).$$

Collecting terms, we therefore arrive at the expression

$$\frac{\mathrm{d}}{\mathrm{d}t}\|x(t)-u\|^2 + \lambda(t)\beta(t)\left(\frac{2}{(1+c_1)L_{\Psi_1}} - \frac{1+c_0}{1+c_1}\lambda(t)\beta(t)\right)\|\mathsf{B}_1(x(t))\|^2$$
$$+ \left(2 - \frac{1}{(1+c_1)(1+c_0)} - \frac{c_1 L_{\Psi_1}\lambda(t)\beta(t)}{1+c_1}\right)\|\dot{x}(t)\|^2$$
$$\leq 2\lambda(t)\langle a + \mathsf{D}(x(t)) + \varepsilon(t)x(t), u - x(t) - \dot{x}(t)\rangle$$
$$- \frac{2c_1\lambda(t)\beta(t)}{1+c_1}(\Psi_1 + \Psi_2)(x(t)+\dot{x}(t)).$$

For ease of notation, let us set $\mathsf{D}_{\varepsilon(t)} \triangleq \mathsf{D} + \varepsilon(t)\,\mathrm{Id}_{\mathcal{H}}$. We next observe that

$$2\lambda(t)\langle a + \mathsf{D}_{\varepsilon(t)}(x(t)), u - x(t) - \dot{x}(t)\rangle$$

$$= 2\lambda(t)\langle a + D_{\varepsilon(t)}(x(t)), u - x(t)\rangle + 2\lambda(t)\langle a + D_{\varepsilon(t)}(x(t)), -\dot{x}(t)\rangle$$

$$= 2\lambda(t)\langle a + D_{\varepsilon(t)}(u), u - x(t)\rangle + 2\lambda(t)\langle D_{\varepsilon(t)}(x(t)) - D_{\varepsilon(t)}(u), u - x(t)\rangle$$
$$+ 2\lambda(t)\langle a + D_{\varepsilon(t)}(x(t)), -\dot{x}(t)\rangle$$

$$\leq -2\lambda(t)\varepsilon(t)\|x(t) - u\|^2 + 2\lambda(t)\langle a + D_{\varepsilon(t)}(u), u - x(t)\rangle$$
$$+ 2\lambda(t)\langle a + D_{\varepsilon(t)}(x(t)), -\dot{x}(t)\rangle$$

$$\leq -2\lambda(t)\varepsilon(t)\|x(t) - u\|^2 + 2\lambda(t)\langle a + D_{\varepsilon(t)}(u), u - x(t)\rangle$$
$$+ \frac{c_2}{2}\|\dot{x}\|^2 + \frac{2\lambda^2(t)}{c_2}\left\|a + D_{\varepsilon(t)}(x(t))\right\|^2$$

$$\leq -2\lambda(t)\varepsilon(t)\|x(t) - u\|^2 + 2\lambda(t)\langle a + D_{\varepsilon(t)}(u), u - x(t)\rangle$$
$$+ \frac{c_2}{2}\|\dot{x}\|^2 + \frac{4\lambda^2(t)}{c_2}\left\|a + D_{\varepsilon(t)}(u)\right\|^2 + + \frac{4\lambda^2(t)}{c_2}\left\|D_{\varepsilon(t)}(x(t)) - D_{\varepsilon(t)}(u)\right\|^2$$

$$\leq -2\lambda(t)\varepsilon(t)\|x(t) - u\|^2 + 2\lambda(t)\langle a + D_{\varepsilon(t)}(u), u - x(t)\rangle$$
$$+ \frac{c_2}{2}\|\dot{x}\|^2 + \frac{4\lambda^2(t)}{c_2}\left\|a + D_{\varepsilon(t)}(u)\right\|^2 + \frac{4\lambda^2(t)}{c_2}\left(\frac{2}{\eta^2} + 2\varepsilon(t)^2\right)\|x(t) - u\|^2$$

from the definition, we have $a + D(u) = w - \xi$, which means

$$2\lambda(t)\langle a + D_{\varepsilon(t)}(x(t)), u - x(t) - \dot{x}(t)\rangle$$
$$\leq \left(\frac{8\lambda^2(t)}{c_2\eta^2} + \frac{8\lambda^2(t)\varepsilon^2(t)}{c_2} - 2\lambda(t)\varepsilon(t)\right)\|x(t) - u\|^2 + 2\lambda(t)\langle a + D_{\varepsilon(t)}(u), u - x(t)\rangle$$
$$+ \frac{c_2}{2}\|\dot{x}\|^2 + \frac{4\lambda^2(t)}{c_2}\left\|a + D_{\varepsilon(t)}(u)\right\|^2$$
$$\leq \left(\frac{8\lambda^2(t)}{c_2\eta^2} + \frac{8\lambda^2(t)\varepsilon^2(t)}{c_2} - 2\lambda(t)\varepsilon(t)\right)\|x(t) - u\|^2 + 2\lambda(t)\langle w - \xi, u - x(t)\rangle$$
$$+ 2\lambda(t)\varepsilon(t)\langle u, u - x(t)\rangle + \frac{c_2}{2}\|\dot{x}(t)\|^2 + \frac{4\lambda^2(t)}{c_2}\left\|a + D_{\varepsilon(t)}(u)\right\|^2$$

By Young's inequality,

$$2\lambda(t)\varepsilon(t)\langle u, u - x(t)\rangle = 2\langle \varepsilon(t)u, \lambda(t)(u - x(t))\rangle$$
$$\leq c_3\varepsilon^2(t)\|u\|^2 + \frac{1}{c_3}\lambda^2(t)\|u - x(t)\|^2$$

hence,

$$2\lambda(t)\langle a + D_{\varepsilon(t)}(x(t)), u - x(t) - \dot{x}(t)\rangle$$
$$\leq \left(\frac{8\lambda^2(t)}{c_2\eta^2} + \frac{8\lambda^2(t)\varepsilon^2(t)}{c_2} + \frac{1}{c_3}\lambda^2(t) - 2\lambda(t)\varepsilon(t)\right)\|x(t) - u\|^2$$
$$+ 2\lambda(t)\langle w - \xi, u - x(t)\rangle + c_3\varepsilon^2(t)\|u\|^2 + \frac{c_2}{2}\|\dot{x}(t)\|^2$$

$$+ \frac{4\lambda^2(t)}{c_2} \left\| a + \mathsf{D}_{\varepsilon(t)}(u) \right\|^2 .$$

Plugging this bound into the above display, we can continue with

$$\frac{\mathrm{d}}{\mathrm{d}t} \|x(t) - u\|^2 + \lambda(t)\beta(t) \left(\frac{2}{(1+c_1)L_{\Psi_1}} - \frac{1+c_0}{1+c_1}\lambda(t)\beta(t) \right) \|\mathsf{B}_1(x(t))\|^2$$

$$+ \left(2 - \frac{1}{(1+c)(1+c_0)} - \frac{c_1 L_{\Psi_1}\lambda(t)\beta(t)}{1+c_1} \right) \|\dot{x}(t)\|^2$$

$$\leq 2\lambda(t)\langle w, u - x(t)\rangle - \frac{2c_1\lambda(t)\beta(t)}{1+c_1}(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))$$

$$+ \left(\frac{8\lambda^2(t)}{c_2\eta^2} + \frac{8\lambda^2(t)\varepsilon^2(t)}{c_2} + \frac{1}{c_3}\lambda^2(t) - 2\lambda(t)\varepsilon(t) \right) \|x(t) - u\|^2$$

$$+ c_3\varepsilon^2(t) \|u\|^2 + \frac{c_2}{2} \|\dot{x}(t)\|^2 + \frac{4\lambda^2(t)}{c_2} \left\| a + \mathsf{D}_{\varepsilon(t)}(u) \right\|^2$$

$$- 2\lambda(t)\langle \xi, u - x(t) - \dot{x}(t)\rangle - 2\lambda(t)\langle \xi, \dot{x}(t)\rangle.$$

Applying Young's inequality again to the last term, we obtain

$$2\lambda(t)\langle \xi, -\dot{x}(t)\rangle \leq \frac{c_2}{2} \|\dot{x}(t)\|^2 + \frac{2\lambda^2(t)}{c_2} \|\xi\|^2 .$$

Consequently,

$$\frac{\mathrm{d}}{\mathrm{d}t} \|x(t) - u\|^2 + \lambda(t)\beta(t) \|\mathsf{B}_1(x(t))\|^2 \left(\frac{2}{(1+c_1)L_{\Psi_1}} - \frac{1+c_0}{1+c_1}\lambda(t)\beta(t) \right)$$

$$+ \left(2 - \frac{1}{(1+c_1)(1+c_0)} - \frac{c_1 L_{\Psi_1}\lambda(t)\beta(t)}{1+c_1} - c_2 \right) \|\dot{x}(t)\|^2$$

$$\leq 2\lambda(t)\langle w, u - x(t)\rangle - \frac{2c_1\lambda(t)\beta(t)}{1+c_1}(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))$$

$$+ \left(\frac{8\lambda^2(t)}{c_2\eta^2} + \frac{8\lambda^2(t)\varepsilon^2(t)}{c_2} + \frac{1}{c_3}\lambda^2(t) - 2\lambda(t)\varepsilon(t) \right) \|x(t) - u\|^2$$

$$+ c_3\varepsilon^2(t) \|u\|^2 + \frac{4\lambda^2(t)}{c_2} \left\| a + \mathsf{D}_{\varepsilon(t)}(u) \right\|^2$$

$$+ \frac{2\lambda^2(t)}{c_2} \|\xi\|^2 - 2\lambda(t)\langle \xi, u - x(t) - \dot{x}(t)\rangle$$

Let us set $\delta := \limsup_{t\to\infty} L_{\Psi_1}\lambda(t)\beta(t) < 2$. Then, we obtain

$$2 - \frac{1}{(1+c_1)(1+c_0)} - \frac{c_1 L_{\Psi_1}\lambda(t)\beta(t)}{1+c_1} - c_2 \geq 2$$

$$- \frac{1}{(1+c_1)(1+c_0)} - \frac{c_1\delta}{1+c_1} - c_2.$$

Letting $c_1 \to 0^+$, we obtain

$$2 - \frac{1}{1 + c_0} - c_2 > 0.$$

For $(c_0, c_2) > 0$ sufficiently small this can be achieved. Henceforth, by continuity, there exists a set of parameters (c_0, c_1, c_2) sufficiently small so that

$$b \triangleq 2 - \frac{1}{(1 + c_1)(1 + c_0)} - \frac{c_1 \delta}{1 + c_1} - c_2 > 0,$$

meanwhile

$$e \triangleq \frac{2}{(1 + c_1)L_{\Psi_1}} - \frac{1 + c_0}{(1 + c_1)L_{\Psi_1}}\delta = \frac{2 - (1 + c_0)\delta}{(1 + c_1)L_{\Psi_1}} > 0.$$

Let us further call

$$d \triangleq \frac{2c_1}{1 + c_1}, \quad \text{and} \quad \delta(t) \triangleq -\frac{8\lambda^2(t)}{c_2\eta^2} - \frac{8\lambda^2(t)\varepsilon^2(t)}{c_2} - \frac{1}{c_3}\lambda^2(t) + 2\lambda(t)\varepsilon(t),$$

and to arrive at the expression

$$\frac{d}{dt}\|x(t) - u\|^2 + e\lambda(t)\beta(t)\|B_1(x(t))\|^2 + b\|\dot{x}(t)\|^2$$
$$+ \frac{d}{2}\lambda(t)\beta(t)(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))$$
$$\leq -\frac{d}{2}\lambda(t)\beta(t)(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))$$
$$- \frac{d}{2}\lambda(t)\beta(t)\langle \frac{4\xi}{d\beta(t)}, u\rangle + \frac{d}{2}\lambda(t)\beta(t)\langle \frac{4\xi}{d\beta(t)}, \dot{x}(t) + x(t)\rangle + \frac{2\lambda^2(t)}{c_2}\|\xi\|^2$$
$$+ 2\lambda(t)\langle w, u - x(t)\rangle - \delta(t)\|x(t) - u\|^2 + c_3\varepsilon^2(t)\|u\|^2$$
$$+ \frac{4\lambda^2(t)}{c_2}\|a + D_{\varepsilon(t)}(u)\|^2$$

Since $\sigma_C(\frac{4\xi}{d\beta(t)}) = \langle \frac{4\xi}{d\beta(t)}, u\rangle$ and

$$\langle \frac{4\xi}{d\beta(t)}, \dot{x} + x(t)\rangle - (\Psi_1 + \Psi_2)(x(t) + \dot{x}(t)) \leq (\Psi_1 + \Psi_2)^*(\frac{4\xi}{d\beta(t)}),$$

we can continue with

$$\frac{d}{dt}\|x(t) - u\|^2 + e\lambda(t)\beta(t)\|B_1(x(t))\|^2 + b\|\dot{x}(t)\|^2$$
$$+ \frac{d}{2}\lambda(t)\beta(t)(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))$$

$$\leq \frac{d}{2}\lambda(t)\beta(t)\left[(\Psi_1 + \Psi_2)^*(\frac{4\xi}{d\beta(t)}) - \sigma_\mathcal{C}(\frac{4\xi}{d\beta(t)})\right] + \frac{2\lambda^2(t)}{c_2}\|\xi\|^2$$

$$+ 2\lambda(t)\langle w, u - x(t)\rangle - \delta(t)\|x(t) - u\|^2 + c_3\varepsilon^2(t)\|u\|^2 + \frac{4\lambda^2(t)}{c_2}\|a + \mathsf{D}_{\varepsilon(t)}(u)\|^2$$

Now, consider the case where $w = 0$, i.e. where $u \in \mathrm{zer}(\mathsf{A}+\mathsf{D}+\mathsf{N}_\mathcal{C})$. Our assumptions on ε and λ imply that there exists $T_0 > 0$ sufficiently large so that $\delta(t) > 0$ for all $t \geq T_0$. Fix such a T_0, so that for all $t \geq T_0$, we can further simplify the above expression to

$$\frac{\mathrm{d}}{\mathrm{d}t}\|x(t) - u\|^2 + e\lambda(t)\beta(t)\|\mathsf{B}_1(x(t))\|^2 + b\|\dot{x}(t)\|^2$$

$$+ \frac{d}{2}\lambda(t)\beta(t)(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))$$

$$\leq \frac{d}{2}\lambda(t)\beta(t)\left[(\Psi_1 + \Psi_2)^*(\frac{4\xi}{d\beta(t)}) - \sigma_\mathcal{C}(\frac{4\xi}{d\beta(t)})\right] + \frac{2\lambda^2(t)}{c_2}\|\xi\|^2$$

$$+ c_3\varepsilon^2(t)\|u\|^2 + \frac{4\lambda^2(t)}{c_2}\|a + \mathsf{D}_{\varepsilon(t)}(u)\|^2$$

$$\leq \frac{d}{2}\lambda(t)\beta(t)\left[(\Psi_1 + \Psi_2)^*(\frac{4\xi}{d\beta(t)}) - \sigma_\mathcal{C}(\frac{4\xi}{d\beta(t)})\right] + \frac{2\lambda^2(t)}{c_2}\|\xi\|^2$$

$$+ (c_3 + 8\lambda^2(t)/c_2)\varepsilon^2(t)\|u\|^2 + \frac{8\lambda^2(t)}{c_2}\|a + \mathsf{D}(u)\|^2. \tag{A9}$$

Integrating the penultimate display from T_1 to T_0 with $T_1 \geq T_0$, we obtain

$$\|x(T_1) - u\|^2 - \|x(T_0) - u\|^2 + \int_{T_0}^{T_1} e\lambda(t)\beta(t)\|\mathsf{B}_1(x(t))\|^2\,\mathrm{d}t$$

$$+ \int_{T_0}^{T_1} b\|\dot{x}(t)\|^2\,\mathrm{d}t + \int_{T_0}^{T_1} \frac{d}{2}\lambda(t)\beta(t)(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))\,\mathrm{d}t \leq G,$$

where G is the integral of the right-hand side of (A9) on $\mathbb{R}_{>0}$, which is finite in view of our assumptions. Using Lemma 2, we see that $\lim_{t\to\infty}\|x(t) - u\|$ exists, as well as

$$\int_T^\infty \lambda(t)\beta(t)\|\mathsf{B}_1(x(t))\|^2\,\mathrm{d}t < \infty,$$

$$\int_T^\infty \lambda(t)\beta(t)(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))\,\mathrm{d}t < \infty \text{ and}$$

$$\int_T^\infty \|\dot{x}(t)\|^2\,\mathrm{d}t < \infty.$$

Assuming that $\liminf_{t\to\infty}\lambda(t)\beta(t) > 0$, we thus see that

$$\lim_{t\to\infty}\|\mathsf{B}_1(x(t))\| = 0, \text{ and } \lim_{t\to\infty}(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t)) = 0.$$

If follows that every weak limit point of $x(t)$ lies in $\mathcal{C}$. We claim that every weak limit point of a suitably defined ergodic trajectory must belong to $\text{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$. To derive this conclusion, we fix T_0, and define the ergodic trajectory

$$\bar{x}(T) \triangleq \frac{\int_{T_0}^{T} \lambda(s) x(s)\, \mathrm{d}s}{\int_{T_0}^{T} \lambda(s)\, \mathrm{d}s}$$

as a function of the upper boundary of the integral. Let $T_n \uparrow \infty$, so that $\bar{x}(T_n) \to \bar{x}_\infty$. We claim that $\bar{x}_\infty \in \text{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$. To show this, we integrate again for a given pair $(u, w) \in \text{gr}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$, so that

$$\|x(T_1) - u\|^2 - \|x(T_0) - u\|^2 + \int_{T_0}^{T_1} e\lambda(t)\beta(t)\, \|\mathsf{B}_1(x(t))\|^2\, \mathrm{d}t$$

$$+ \int_{T_0}^{T_1} b\, \|\dot{x}(t)\|^2\, \mathrm{d}t + \int_{T_0}^{T_1} \frac{d}{2}\lambda(t)\beta(t)(\Psi_1 + \Psi_2)(x(t) + \dot{x}(t))\, \mathrm{d}t$$

$$\leq G + \int_{T_0}^{T_1} 2\lambda(t)\langle w, u - x(t)\rangle\, \mathrm{d}t.$$

Dividing both sides by $2\Lambda(T_0, T_1) := 2\int_{T_0}^{T_1} \lambda(t)\, \mathrm{d}t$, we arrive at

$$-\frac{\|x(T_0) - u\|^2}{2\Lambda(T_0, T_1)} \leq \frac{G}{\Lambda(T_0, T_1)} + \langle w, u - \frac{\int_{T_0}^{T_1} \lambda(t) x(t)\, \mathrm{d}t}{\Lambda(T_0, T_1)}\rangle.$$

Letting $T_1 \to \infty$, we obtain

$$0 \leq \langle u, w - \bar{x}_\infty\rangle.$$

Fact 2 allows us to conclude $\bar{x}_\infty \in \text{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$ and so every weak accumulation point of the trajectory $x(\cdot)$ must lie in this set. Since $\lim_{t \to \infty} \|x(t) - u\|$ exists for every $u \in \text{zer}(\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}})$, we conclude that $\bar{x}(t)$ converges weakly, as $t \to \infty$ to a zero of $\mathsf{A} + \mathsf{D} + \mathsf{N}_{\mathcal{C}}$, as claimed.

References

1. Abbas, B., Attouch, H., Svaiter, B.F.: Newton-like dynamics and forward-backward methods for structured monotone inclusions in Hilbert spaces. J. Optim. Theory Appl. **161**, 331–360 (2014)
2. Alvarez, F., Cabot, A.: Asymptotic selection of viscosity equilibria of semilinear evolution equations by the introduction of a slowly vanishing term. Discrete Contin. Dynam. Systems **15**(3), 921 (2006)
3. Attouch, H.: Viscosity solutions of minimization problems. SIAM J. Optim. **6**(3), 769–806 (1996)
4. Attouch, H., Cominetti, R.: A dynamical approach to convex minimization coupling approximation with the steepest descent method. J. Differential Equations **128**(2), 519–540 (1996)
5. Attouch, H., Czarnecki, M.-O.: Asymptotic behavior of coupled dynamical systems with multiscale aspects. J. Differential Equations **248**(6), 1315–1344 (2010)
6. Attouch, H., Czarnecki, M.-O., Peypouquet, J.: Coupling forward-backward with penalty schemes and parallel splitting for constrained variational inequalities. SIAM J. Optim. **21**(4), 1251–1274 (2011). https://doi.org/10.1137/110820300

7. Attouch, H., Chbani, Z., Riahi, H.: Combining fast inertial dynamics for convex optimization with Tikhonov regularization. J. Math. Anal. Appl. **457**(2), 1065–1094 (2018)
8. Attouch, H., Cabot, A., Czarnecki, M.-O.: Asymptotic behavior of nonautonomous monotone and subgradient evolution equations. Trans. Am. Math. Soc. **370**(2), 755–790 (2018)
9. Baillon, J.-B., Cominetti, R.: A convergence result for nonautonomous subgradient evolution equations and its application to the steepest descent exponential penalty trajectory in linear programming. J. Funct. Anal. **187**(2), 263–273 (2001)
10. Banert, S., Boţ, R.I.: Backward penalty schemes for monotone inclusion problems. J. Optim. Theory Appl. **166**(3), 930–948 (2015)
11. Barbu, V.: Optimal Control of Variational Inequalities, vol. 100. Pitman Advanced Publishing Program, Research Notes in Mathematics (1984)
12. Bauschke, H.H., Combettes, P.L.: Convex Analysis and Monotone Operator Theory in Hilbert Spaces. Springer, CMS Books in Mathematics (2016)
13. Borwein, J.M., Vanderwerff, J.D.: Convex Functions: Constructions, Characterizations and Counterexamples, vol. 109. Cambridge University Press, Cambridge (2010)
14. Boţ, R.I., Chenchene, E., Csetnek, E.R., Hulett, D.A.: Accelerating diagonal methods for bilevel optimization: Unified convergence via continuous-time dynamics. arXiv preprint arXiv:2505.14389 (2025)
15. Bot, R.I., Nguyen, D.-K.: Tikhonov regularization of monotone operator flows not only ensures strong convergence of the trajectories but also speeds up the vanishing of the residuals. arXiv preprint arXiv:2406.00852 (2024)
16. Boţ, R.I., Csetnek, E.: Forward-backward and Tseng's type penalty schemes for monotone inclusion problems. Set-Valued and Variational Analysis **22**(2), 313–331 (2014). https://doi.org/10.1007/s11228-014-0274-7
17. Boţ, R.I., Csetnek, E.: Approaching the solving of constrained variational inequalities via penalty term-based dynamical systems. J. Math. Anal. Appl. **435**(2), 1688–1700 (2016). https://doi.org/10.1016/j.jmaa.2015.11.032
18. Boţ, R.I., Csetnek, E.: An inertial forward-backward-forward primal-dual splitting algorithm for solving monotone inclusion problems. Numerical Algorithms **71**(3), 519–540 (2016). https://doi.org/10.1007/s11075-015-0007-5
19. Boţ, R.I., Csetnek, E.: A dynamical system associated with the fixed points set of a nonexpansive operator. J. Dyn. Diff. Equat. **29**(1), 155–168 (2017)
20. Boţ, R.I., Csetnek, E.R.: Penalty schemes with inertial effects for monotone inclusion problems. Optimization **66**(6), 965–982 (2017)
21. Boţ, R.I., Hendrich, C.: Convergence analysis for a primal-dual monotone + skew splitting algorithm with applications to total variation minimization. Journal of Mathematical Imaging and Vision **49**(3), 551–568 (2014). https://doi.org/10.1007/s10851-013-0486-8
22. Boţ, R.I., Nguyen, D.-K.: A forward-backward penalty scheme with inertial effects for monotone inclusions. applications to convex bilevel programming. Optimization 68(10), 1855–1880 (2019). https://doi.org/10.1080/02331934.2018.1556662
23. Boţ, R.I., Csetnek, E.R., László, S.C.: Second-order dynamical systems with penalty terms associated to monotone inclusions. Anal. Appl. **16**(05), 601–622 (2018)
24. Boţ, R.I., Grad, S.-M., Meier, D., Staudigl, M.: Inducing strong convergence of trajectories in dynamical systems associated to monotone inclusions with composite structure. Adv. Nonlinear Anal. **10**(1), 450–476 (2020)
25. Boţ, R.I., Csetnek, E.R., László, S.C.: On the strong convergence of continuous Newton-like inertial dynamics with Tikhonov regularization for monotone inclusions. J. Math. Anal. Appl. **530**(2), 127689 (2024). https://doi.org/10.1016/j.jmaa.2023.127689
26. Cabot, A.: Proximal point algorithm controlled by a slowly vanishing term: applications to hierarchical minimization. SIAM J. Optim. **15**(2), 555–572 (2005)
27. Clarke, F.H., Ledyaev, Y.S., Stern, R.J., Wolenski, P.R.: Nonsmooth Analysis and Control Theory, vol. 178. Springer, Graduate Texts in Mathematics (2008)
28. Combettes, P.L., Pesquet, J.-C.: Primal-dual splitting algorithm for solving inclusions with mixtures of composite, Lipschitzian, and parallel-sum type monotone operators. Set-Valued and variational analysis **20**(2), 307–330 (2012)

29. Cominetti, R., Peypouquet, J., Sorin, S.: Strong asymptotic convergence of evolution equations governed by maximal monotone operators with Tikhonov regularization. J. Differential Equations **245**(12), 3753–3763 (2008)

30. Czarnecki, M.O., Noun, N., Peypouquet, J.: Splitting forward-backward penalty scheme for constrained variational problems. J. Convex Anal. **23**(2), 531–565 (2016)

31. Dai, Y.-H., Wang, J., Zhang, L.: Optimality conditions and numerical algorithms for a class of linearly constrained minimax optimization problems. SIAM J. Optim. **34**(3), 2883–2916 (2024)

32. Ding, K.-W., Fliege, J., Vuong, P.T.: Fast convergence of the primal-dual dynamical system and corresponding algorithms for a nonsmooth bilinearly coupled saddle point problem. Comput. Optim. Appl. **90**(1), 151–192 (2025)

33. Guth, P.A., Schillings, C., Weissmann, S.: One-shot learning of surrogates in PDE-constrained optimization under uncertainty. SIAM/ASA Journal on Uncertainty Quantification **12**(2), 614–645 (2024). https://doi.org/10.1137/23M1553170

34. Maulen Soto, R., Fadili, J., Attouch, H.: An stochastic differential equation perspective on stochastic convex optimization. Mathematics of Operations Research (2024)

35. Noun, N., Peypouquet, J.: Forward-backward penalty scheme for constrained convex minimization without inf-compactness. J. Optim. Theory Appl. **158**, 787–795 (2013)

36. Ogura, N., Yamada, I.: Non-strictly convex minimization over the fixed point set of an asymptotically shrinking nonexpansive mapping (2002)

37. Peypouquet, J.: Convex Optimization in Normed Spaces: Theory, Methods and Examples. Springer, Springer Briefs in Optimization (2015)

38. Peypouquet, J.: Coupling the gradient method with a general exterior penalization scheme for convex minimization. J. Optim. Theory Appl. **153**(1), 123–138 (2012). https://doi.org/10.1007/s10957-011-9936-x

39. Reyes, J.C.: Optimal control of a class of variational inequalities of the second kind. SIAM J. Control. Optim. **49**(4), 1629–1658 (2011)

40. Reyes, J.C., Meyer, C.: Strong stationarity conditions for a class of optimization problems governed by variational inequalities of the second kind. J. Optim. Theory Appl. **168**, 375–409 (2016)

41. Sabach, S., Shtern, S.: A first order method for solving convex bilevel optimization problems. SIAM J. Optim. **27**(2), 640–660 (2017)

42. Sontag, E.D.: Mathematical Control Theory: Deterministic Finite Dimensional Systems, vol. 6. Springer, Texts in Applied Mathematics (TAM) (2013)

43. Staudigl, M., Qu, S.: Tikhonov regularized exterior penalty dynamics for constrained variational inequalities. IEEE Control Systems Letters **8**, 622–627 (2024). https://doi.org/10.1109/LCSYS.2024.3401019

44. Tseng, P.: A modified forward-backward splitting method for maximal monotone mappings. SIAM J. Control. Optim. **38**(2), 431–446 (2000)

45. Xu, H.K., Kim, T.H.: Convergence of hybrid steepest-descent methods for variational inequalities. J. Optim. Theory Appl. **119**(1), 185–201 (2003). https://doi.org/10.1023/B:JOTA.0000005048.79379.b6