Journal of Convex Analysis
Volume 33 (2026), No. 3&4, 597–614

Approximation of Bilevel Optimization Problems Involving an Optimal Control Problem at the Lower Level

Térence Bayen
Avignon Université, Laboratoire de Mathématiques d'Avignon (EA 2151), France
terence.bayen@univ-avignon.fr

Francis Mairet
Ifremer, Phytox, Physalg Laboratory, Nantes, France
francis.mairet@ifremer.fr

Dedicated to the memory of Hédy Attouch.

Received: May 13, 2025
Accepted: November 3, 2025

This work studies a regularization procedure whose goal is the approximation of a solution to a bilevel optimization problem where the lower-level problem is formulated as an optimal control problem. This is done by regularizing the lower-level problem in such a way that an optimal control to the regularized problem can be explicitly computed as a feedback of state and costate functions. Convergence between the regularized problems and the original one is addressed: in particular, it is shown that as the regularization parameter tends to zero, the upper-level objective function computed on the optimal control to the regularized problem converges to the value function associated with the original program. The analysis that is carried out partly relies on the hypothesis that the lower-level problem admits a unique solution.

Keywords: Bilevel optimization, optimal control, Pontryagin maximum principle.

2020 Mathematics Subject Classification: 34A38, 49A15, 49J53.

1. Introduction

In many application models, specific parameters may be unknown, and estimating them based on available measurements can be of significant interest, particularly in industrial, biotechnological, and related fields. Such questions arise, for example, in microbial ecology or resource allocation models, see, *e.g.*, [20, 21]. The objective in these cases is to infer parameters describing microorganisms growth (such as microalgae), under the key assumption that microorganisms maximise their growth rate by using an optimal strategy for allocating resources (acquired through evolution). Such questions naturally lead to the formulation of bilevel optimization problems of the following form: at the upper level, the objective is to minimize the discrepancy between observed and predicted values for parameter estimation, while the lower-level involves solving an optimal control problem (OCP) [11, 20, 21, 22]. This paper focuses on bilevel optimization problems that exhibit precisely this structure. While such a framework may also emerge in the context of models subject to uncertainties, our primary motivation here lies in optimization problems involving unknown parameters to be identified, as in [21].

ISSN 0944-6532 / $ 2.50

The framework under consideration is as follows. We consider a bilevel optimization problem in which the upper level is a classical nonlinear programming problem, while the lower-level consists of an optimal control problem. Bilevel optimization problems have been extensively studied in the literature (see, *e.g.*, [12, 13, 14, 34] among many other related works) and are generally known to be quite involved. Notably, even when both the upper and lower problems are linear, the bilevel formulation becomes non-linear. This is because replacing the lower-level problem with its Karush-Kuhn-Tucker conditions introduces complementarity constraints, which are inherently non-linear [16, 15]. Additionally, when the lower-level problem admits multiple solutions, bilevel formulations give rise to the distinction between optimistic and pessimistic solution concepts [12]. In the context of bilevel problems involving an OCP at the lower-level, several studies addressed the resulting issues both from theoretical and practical perspectives (see, *e.g.*, [11, 17, 21, 22, 32, 33]). In [21], the Levenberg–Marquardt algorithm is implemented at the upper level and, at each iteration, a direct method is used to numerically solve the lower-level OCP. A similar approach is employed in [22]. In [11], bilevel optimization is studied by means of the dynamic programming principle. Necessary optimality conditions in the form of a Pontryagin Maximum Principle (PMP) [24] have also been developed in [32, 33] and, based on the theory of multiprocesses [10], the articles [1, 2] also studies related issues.

In this work, we address this framework by introducing a regularization of the lower OCP problem, yielding a bilevel optimization problem that approaches the original one. First, we suppose that the lower-level OCP is written in the Mayer form and that it is governed by a mono-input affine control system. Next, we consider an augmentation of the dynamics which consists in the introduction of linearly independent vector fields at the lower-level (see, *e.g.*, [19, 27, 29, 30, 31]). Notably, this approach provides an efficient way for solving approximately optimal control problems that involve singular arcs and for which the optimal synthesis can be involved. For instance [27] presents an elegant application of this method to a problem featuring two competing turnpike-type singular arcs. With regard to the regularized OCP, an optimal control can be explicitly expressed in terms of state and costate functions thanks to the PMP. Next, we can evaluate the upper-level objective function at a solution to this new problem. Our main result establishes that, as the regularization parameter tends to zero, this value converges (up to a sub-sequence) to the value function of the original bilevel problem. We also prove that optimal solutions of the (regularized) lower-level OCP converge to an optimal solution of the original lower-level problem. A key point is that we restrict ourselves to optimistic solutions. Moreover, existence of an optimal solution (to the bilevel problem) depends on either an assumption of uniqueness of the solutions to the lower-level problem or on an assumption of stability of these solutions. The proof of the main convergence result requires analogous hypotheses.

The paper is structured as follows. In Section 2, we introduce the bilevel problem and we prove the existence of an optimistic solution. In Section 3, the regularization of the lower OCP is introduced. Convergence results are provided in Section 4. We also illustrate the effectiveness of this regularization technique on a bilevel optimization problem arising in microbial resource allocation models [21, 20], where the upper-level problem consists in recovering uncertain parameters. Finally, in the last section, some perspectives are discussed.

2. The bilevel optimization problem

In the sequel, $\|\cdot\|_{\mathbb{R}^d}$ and $\cdot$ denote respectively the Euclidean norm and the inner product in $\mathbb{R}^d$ for some $d \in \mathbb{N}^*$. The open ball, resp. closed ball in $\mathbb{R}^d$ of radius $r > 0$ and of center $x \in \mathbb{R}^d$ is written $B(x,r)$, resp. $B[x,r]$. Throughout the paper, $T > 0$ is fixed. We denote by $\mathrm{AC}([0,T],\mathbb{R}^d)$ the set of absolutely continuous functions from $[0,T]$ with values in $\mathbb{R}^d$. Moreover, for a smooth function $f : \mathbb{R}^d \to \mathbb{R}^{d'}$ where $d, d' \in \mathbb{N}^*$, $Df(x)$ and $\nabla f(x)$ stand respectively for the Jacobian and the gradient of f at some point $x \in \mathbb{R}^d$. Given a sequence of absolutely continuous functions $x_n : [0,T] \to \mathbb{R}^d$, we say that (x_n) strongly-weakly converges to x^* over $[0,T]$ if (x_n) uniformly converges to x^* over $[0,T]$ and $(\dot{x}_n)$ weakly converges to $\dot{x}^*$ in $L^2([0,T],\mathbb{R}^d)$.

Throughout this paper, m, n, and N are fixed positive integers, $\Theta \subset \mathbb{R}^m$ denotes a non-empty compact subset, $J : (\mathbb{R}^n)^N \to \mathbb{R}$ is a continuous function bounded below and for each $\theta \in \Theta$, $f_\theta : \mathbb{R}^n \to \mathbb{R}^n$ and $g_\theta^1 : \mathbb{R}^n \to \mathbb{R}^n$ are vector fields of class C^1. The function $\phi : \mathbb{R}^n \to \mathbb{R}$ is the terminal payoff and it is supposed to be of class C^1. We are also given N time instants $(t_i)_{1\le i\le N}$ of $[0,T]$ such that $0 < t_1 < t_2 < \cdots < t_N < T$. The bilevel optimization problem under consideration is as follows : the upper program is a classical non-linear programming problem involving J and the lower one is a Mayer OCP governed by a mono-input affine control system for which the set of admissible controls is $\mathcal{U} := \{u : [0,T] \to [-1,1] \ ; \ u \text{ meas.}\}$ and the initial condition $\xi \in \mathbb{R}^n$ is fixed. The resulting problem can be written as follows:

$$\underset{\theta\in\Theta}{\text{minimize}} \ J(x_\theta(t_1), ..., x_\theta(t_N)) \tag{1}$$

$$\text{subject to} \begin{cases} x_\theta \in \underset{x\in \mathrm{AC}([0,T],\mathbb{R}^n)}{\text{argmin}} \quad \phi(x(T)) \quad \text{s.t.} \\ \dot{x}(t) = f_\theta(x(t)) + u(t) g_\theta^1(x(t)) \ \text{ a.e. } t \in [0,T], \\ x(0) = \xi \text{ and } u \in \mathcal{U}. \end{cases} \tag{2}$$

Remark 2.1. An example of a problem like (1)–(2) as in [21] consists in minimizing a functional like

$$J(x_\theta(t_1), ..., x_\theta(t_N)) := \sum_{i=1}^N \|x_\theta(t_i) - y_i\|_{\mathbb{R}^n}$$

at the upper level where $(y_i)_{1\le j\le N}$ are given data and x_θ is a solution to a Mayer OCP governed by a mono-input affine control system in $\mathbb{R}^3$ describing a resource allocation model, see Section 4. □

Before going any further, let us underline the specificity of a bilevel optimization problem like (1)–(2) and in which sense a solution has to be understood:

- The variable θ is called the *leader* and the variable $x_\theta \in \mathrm{AC}([0,T],\mathbb{R}^n)$ (the unique solution to the Cauchy problem $\dot{x} = f_\theta(x) + u(t) g_\theta^1(x)$, $x(0) = \xi$ associated with a given $u \in \mathcal{U}$) is called the *follower*.
- If for a given value of θ, the lower-level problem does not admit a unique solution, then, taking the minimum in the upper level problem is not well defined, as there may be multiple solutions to (2).

For this reason, we shall consider in this paper the *optimistic position* meaning that the leader expects the follower to choose a solution which has the best objective function at the upper level. For this purpose, let us define the multifunction

$$\Psi^o : \Theta \rightrightarrows \mathrm{AC}([0,T],\mathbb{R}^n)$$

as $$\Psi^o(\theta) := \underset{x_\theta \in \mathrm{AC}([0,T],\mathbb{R}^n)}{\operatorname{argmin}} \{J(x_\theta(t_1),...,x_\theta(t_N)) \ ; \ x_\theta \text{ is a solution to (2)}\},$$

where $\theta \in \Theta$. A solution to the bilevel optimization problem has to be understood as follows.

Definition 2.2. A pair $(\theta^*, x_{\theta^*}) \in \Theta \times \mathrm{AC}([0,T],\mathbb{R}^n)$ is said to be a *solution* to (1)–(2) if:

$$\forall \theta \in \Theta, \ J(x_{\theta^*}(t_1),...,x_{\theta^*}(t_N)) \leq J(x_\theta(t_1),...,x_\theta(t_N)), \tag{3}$$

where for all $\theta \in \Theta$, $x_\theta \in \Psi^o(\theta)$. □

Note that if (θ^*, x_{θ^*}) is an optimal solution to (1)–(2), then, $(x_{\theta^*}, u_{\theta^*})$ is a solution to the OCP (2) with $\theta := \theta^*$. Throughout the paper, we suppose that the following basic hypotheses are satisfied.

(H1) For every compact set $K \subset \mathbb{R}^n$, there is $L_K \geq 0$ such that for all $x \in K$ and for all $(\theta, \theta') \in \Theta^2$, one has
$\|f_\theta(x) - f_{\theta'}(x)\|_{\mathbb{R}^n} \leq L_K \|\theta - \theta'\|_{\mathbb{R}^n}$; $\|g^1_\theta(x) - g^1_{\theta'}(x)\|_{\mathbb{R}^n} \leq L_K \|\theta - \theta'\|_{\mathbb{R}^m}$
and $(\theta, x) \mapsto f_\theta(x)$, $(\theta, x) \mapsto g^1_\theta(x)$ are continuous w.r.t. (x, θ).

(H2) There is a constant $c \geq 0$ such that:
$\forall \theta \in \Theta, \ \forall x \in \mathbb{R}^n, \ \ \|f_\theta(x)\|_{\mathbb{R}^n} \leq c(1 + \|x\|_{\mathbb{R}^n})$ and $\|g_\theta(x)\|_{\mathbb{R}^n} \leq c(1 + \|x\|_{\mathbb{R}^n})$.

Since the control system in (2) is affine w.r.t. the input u and the admissible control set $[-1, 1]$ is compact, for every $\theta \in \Theta$, the OCP (2) has an optimal solution $u_\theta \in \mathcal{U}$ and x_θ stands for the unique solution to the Cauchy problem

$$\begin{cases} \dot{x}(t) &= f_\theta(x(t)) + u(t) g^1_\theta(x(t)) \quad \text{a.e. } t \in [0,T], \\ x(0) &= \xi, \end{cases} \tag{4}$$

with $u = u_\theta$. Moreover, the set $\mathcal{A}_\theta$, defined as the set of admissible solutions to (4) for arbitrary $u \in \mathcal{U}$, is compact in $\mathrm{AC}([0,T],\mathbb{R}^n)$ under the strong-weak topology by standard arguments. We refer for instance to [9, Theorem 1.11].

Lemma 2.3. (i) *There is $R \geq 0$ such that for every $\theta \in \Theta$ and for every $x \in \mathcal{A}_\theta$, one has:*

$$\forall t \in [0,T], \ \|x(t)\|_{\mathbb{R}^n} \leq R.$$

(ii) *For every $\theta \in \Theta$, $\Psi^o(\theta)$ is non-empty.*

Proof. The first point follows from the linear growth hypothesis (H2). To prove (ii), let $(x^n_\theta) \in \mathcal{A}^{\mathbb{N}}_\theta$ be such that for every $n \in \mathbb{N}$, x^n_θ is an optimal solution to (2), and

$$J(x^n_\theta(t_1),...,x^n_\theta(t_N)) \to \alpha := \inf\{J(x_\theta(t_1),...,x_\theta(t_N)) \ ; \ x_\theta \text{ is a solution to (2)}\}.$$

as $n \to +\infty$. Since $\mathcal{A}_\theta$ is compact, we may assume (up to taking a sub-sequence) that (x^n_θ) strongly-weakly converges to an admissible solution x_θ to (4). Now, ϕ is continuous, hence, x_θ is an optimal solution to (2) as well.

Moreover, by using the continuity of J, we deduce that

$$J(x_\theta(t_1), ..., x_\theta(t_N)) = \alpha,$$

by letting $n \to +\infty$. The result follows. $\square$

Since (H1)–(H2) is satisfied, we set $L := L_K$ where $K := B[0, R]$. We now turn to the existence of a solution to the bilevel optimization problem. For this purpose, we introduce the following hypotheses:

(H) For every $\theta \in \Theta$, the lower-level OCP (2) has a unique solution $x_\theta \in \mathcal{A}_\theta$.

(H') For every $(\theta, x_\theta) \in \Theta \times \mathcal{A}_\theta$ such that x_θ is optimal for (2), for every $(\theta_n) \in \Theta^{\mathbb{N}}$ converging to θ as $n \to \infty$, there is an optimal solution x_{θ_n} to (2) with $\theta = \theta_n$ such that the sequence (x_{θ_n}) strongly-weakly converges to x_θ as $n \to +\infty$.

Theorem 2.4. *Under* (H) *or* (H'), *the bilevel optimization problem* (1)–(2) *has a solution* (θ^*, x_{θ^*}).

Proof. Let us set $\alpha := \inf_{\theta \in \Theta}\{J(x_\theta(t_1), ..., x_\theta(t_N))\ ;\ x_\theta \in \Psi^o(\theta)\}$. There exists $(\theta_n) \in \Theta^{\mathbb{N}}$ such that for every $n \in \mathbb{N}$, there is $x_{\theta_n} \in \Psi^o(\theta_n)$ such that

$$J(x_{\theta_n}(t_1), ..., x_{\theta_n}(t_N)) \to \alpha,$$

as $n \to +\infty$. Recall that for all $n \in \mathbb{N}$, one has $x_{\theta_n}(0) = \xi$. Since Θ is compact, we may assume that (θ_n) converges (up to a subsequence) to some $\theta^* \in \Theta$. For every $n \in \mathbb{N}$, one can write for a.e. $t \in [0, T]$:

$$\dot{x}_{\theta_n}(t) = f_{\theta_n}(x_{\theta_n}(t)) + u_{\theta_n}(t) g^1_{\theta_n}(x_{\theta_n}(t)) = f_{\theta^*}(x_{\theta_n}(t)) + u_{\theta_n}(t) g^1_{\theta^*}(x_{\theta_n}(t)) + r_n(t),$$

where

$$r_n(t) := f_{\theta_n}(x_{\theta_n}(t)) - f_{\theta^*}(x_{\theta_n}(t)) + u_{\theta_n}(t)(g^1_{\theta_n}(x_{\theta_n}(t)) - g^1_{\theta^*}(x_{\theta_n}(t))).$$

Using Lemma 2.3, we deduce from (H1) that $\|r_n\|_{L^2([0,T],\mathbb{R}^n)} \to 0$ as $n \to +\infty$. Now, the velocity set associated with the control system $\dot{x} = f_{\theta^*}(x) + u(t) g_{\theta^*}(x)$ is a non-empty compact convex subset of $\mathbb{R}^n$ for every $x \in \mathbb{R}^n$. Hence, we are in a position to apply Theorem 1.11 from [9] about compactness of solutions to a control system. We deduce that there exist $u^* \in \mathcal{U}$ and an absolutely continuous function $x^* : [0, T] \to \mathbb{R}^n$ such that (x^*, u^*) satisfies $x^*(0) = \xi$ and

$$\dot{x}^*(t) = f_{\theta^*}(x^*(t)) + u^*(t) g_{\theta^*}(x^*(t)) \quad \text{a.e. } t \in [0, T],$$

moreover, up to a subsequence, (x_{θ_n}) strongly-weakly converges to x^* over $[0, T]$. Let us set $x^*_{\theta^*} := x^*$, $u^*_{\theta^*} := u^*$. We wish now to prove that the pair $(x^*_{\theta^*}, u^*_{\theta^*})$ is a solution to the lower problem (2) with $\theta := \theta^*$. For this purpose, note that for every $n \in \mathbb{N}$, one has:

$$\phi(x_{\theta_n}(T)) \le \phi(y_n(T)), \tag{5}$$

where y_n is the unique solution to the Cauchy problem

$$\begin{cases} \dot{y}_n(t) & = f_{\theta_n}(y_n(t)) + u(t) g_{\theta_n}(y_n(t)) \quad \text{a.e. } t \in [0, T], \\ y_n(0) & = \xi \end{cases} \tag{6}$$

for a given arbitrary $u \in \mathcal{U}$.

As previously, the preceding ordinary differential equation can be rewritten

$$\dot{y}_n(t)=f_{\theta^*}(y_n(t))+u(t)g_{\theta^*}(y_n(t))+f_{\theta_n}(y_n(t))-f_{\theta^*}(y_n(t))+u(t)(g_{\theta_n}(y_n(t))-g_{\theta^*}(y_n(t))),$$

for a.e. $t \in [0,T]$, so that by a similar argumentation as for proving the convergence of (x_{θ_n}) to x^*, we may assume that, up to a sub-sequence, (y_n) strongly-weakly converges over $[0,T]$ to an absolutely continuous function $\tilde{y}$ satisfying

$$\dot{\tilde{y}}(t) = f_{\theta^*}(\tilde{y}(t)) + \tilde{u}(t)g_{\theta^*}(\tilde{y}(t)) \quad \text{a.e. } t \in [0,T],$$

for some $\tilde{u} \in \mathcal{U}$. By letting $n \to +\infty$ in (6), we obtain that $\tilde{u} = u$ almost everywhere over $[0,T]$ (note that the control u in (6) is independent on n). By passing to the limit in (5), we deduce that $\phi(x_{\theta^*}(T)) \leq \phi(\tilde{y}(T))$ where $\tilde{y}$ is any admissible solution to the control system $\dot{x} = f_{\theta^*}(x) + u(t)g_{\theta^*}(x)$ (since $u \in \mathcal{U}$ is arbitrary).

Suppose now that (H) is satisfied. It follows that $\alpha := \inf_{\theta\in\Theta}\{J(x_\theta(t_1), ..., x_\theta(t_N))\}$ where x_θ denotes the unique solution to (2) for any fixed value of $\theta \in \Theta$. From the uniform convergence of (x_{θ_n}) to x_{θ^*}, we get that $\alpha = J(x_{\theta^*}(t_1), ..., x_{\theta^*}(t_N))$ which ends the proof in this case.

Suppose now that (H') is satisfied. Let us given an optimal solution $\tilde{x}_{\theta^*} \in \mathcal{A}_{\theta^*}$ to (2) with $\theta = \theta^*$. It follows that there is a sequence $(\tilde{x}_{\theta_n}) \in \mathrm{AC}([0,T], \mathbb{R}^n)^{\mathbb{N}}$ such that for every $n \in \mathbb{N}$, $\tilde{x}_{\theta_n}$ is an optimal solution to (2) and such that $(\tilde{x}_{\theta_n})$ strongly-weakly converges to $\tilde{x}_{\theta^*}$ as $n \to +\infty$. Since

$$\forall n \in \mathbb{N}, \quad J(x_{\theta_n}(t_1), ..., x_{\theta_n}(t_N)) \leq J(\tilde{x}_{\theta_n}(t_1), ..., x_{\theta_n}(t_N)),$$

we obtain that $x_{\theta^*} \in \Psi^o(\theta^*)$ and moreover that $\alpha = J(x_{\theta^*}(t_1), ..., x_{\theta^*}(t_N))$ by letting $n \to +\infty$. This ends the proof. □

Remark 2.5. In the bilevel context, supposing that the lower-level OCP has a unique solution means that the response of the follower to the leader is unique and it simplifies the approach since Ψ^o is a singleton. In our setting, this can be verified for instance in the linear quadratic case, but, for a non-linear OCP like (2), we cannot prove in general that the optimal solution is unique (typically if the problem has symmetries). This question is related to the differentiability of the value function. However, uniqueness can still be established for specific problems whose solutions are concatenations of bang and singular arcs. This property is rather problem-dependent. For instance, the problem analyzed in [23] admits a unique optimal control of this form (this is proved using a clock form argumentation). Hypothesis (H') pertains to the closedness properties of the multifunction Ψ^o, which, to the best of our knowledge, is a delicate question in our infinite dimensional setting. We refer to [28] (and references herein) for sufficient conditions on the data defining a bilevel problem that ensure the existence of a solution (in the optimistic or pessimistic case). □

At the lower-level, the PMP provides the following properties on a solution x_{θ^*} to (2) with $\theta = \theta^*$ associated with some control $u_{\theta^*} \in \mathcal{U}$.

Proposition 2.6. *There is an absolutely continuous function $p_{\theta^*} : [0,T] \to \mathbb{R}^n$ (written as a row vector) such that*

$$\left\{\begin{array}{ll} \dot{p}_{\theta^*}(t) & = -p_{\theta^*}(t)Df_{\theta^*}(x_{\theta^*}(t)) - u_{\theta^*}(t)p_{\theta^*}(t)Dg^1_{\theta^*}(x_{\theta^*}(t)) \quad \text{a.e. } t \in [0,T], \\ p_{\theta^*}(T) & = -\nabla\phi(x_{\theta^*}(T))^\top. \end{array}\right. \tag{7}$$

Moreover, if $\varphi_{\theta^*}(t) := p_{\theta^*}(t) \cdot g^1_{\theta^*}(x_{\theta^*}(t))$ denotes the switching function, the Hamiltonian maximization condition in the PMP *reads as follows:*

$$u_{\theta^*}(t) = \text{sign}(\varphi_{\theta^*}(t)),$$

for a.e. $t \in [0,T]$ such that $\varphi_{\theta^*}(t) \neq 0$.

Recall that when φ_{θ^*} is of constant sign over a time interval, the control has a bang arc $u_{\theta^*} = +1$ or $u_{\theta^*} = -1$ (the number of switching times possibly being infinite) whereas if φ_{θ^*} vanishes over a time interval, then the control has a singular arc. For more details about singular arcs, see *e.g.*, [4, 5, 3]. The difficulty in analyzing the lower-level OCP comes from the fact that we do not know the structure of an optimal control so that the use of the PMP for studying the bilevel problem (1)–(2) may not be suitable, that is why, a direct method was preferred in [21]. However, we shall next see that this issue can be circumvented when considering a regularization of the lower problem.

3. Regularization of the lower-level OCP

In order to regularize the lower-level OCP (2), we proceed as follows. Let us given $n-1$ arbitrary smooth vector fields $g^2, ..., g^n : \mathbb{R}^n \to \mathbb{R}^n$ (of class C^1) and let us consider the Cauchy problem

$$\left\{\begin{array}{ll} \dot{x}(t) & = f_\theta(x(t)) + u_1(t)g^1_\theta(x(t)) + \varepsilon \sum_{j=2}^{n} u_j(t)g^j(x(t)) \quad \text{a.e. } t \in [0,T], \\ x(0) & = \xi, \end{array}\right. \tag{8}$$

where $(\theta, \varepsilon) \in \Theta \times \mathbb{R}^*_+$ and admissible controls $\mathbf{u} := (u_1, ..., u_n)$ are now with values in the admissible control set $\mathcal{U}_n$ defined as

$$\mathcal{U}_n := \{\mathbf{u} : [0,T] \to \mathbb{R}^n \; ; \; \mathbf{u}(t) \in B[0_{\mathbb{R}^n}, 1] \; \text{ for a.e. } t \in [0,T]\}.$$

We suppose that the vector fields g^j for $2 \leq j \leq n$ do not depend on $\theta \in \Theta$, as they are not involved in the original problem, but only in the approximation method. Nevertheless, they could depend on θ (which would not change our approach). Moreover, we suppose that they satisfy the following condition:

(H3) For every $\theta \in \Theta$ and for every $x \in \mathbb{R}^n$, $\text{span}\{g^1_\theta(x), g^2(x), ..., g^n(x)\} = \mathbb{R}^n$.

Note that (H3) implies that g^1_θ never vanishes. The existence of smooth vector fields g_i, $2 \leq i \leq n$, can be established using arguments from differential geometry and considering the orthogonal of $g^1_\theta(x)$ at each point $x \in \mathbb{R}^n$. This hypothesis is crucial in our approach (see also [29, 30]). Indeed, thanks to the Hamiltonian maximization condition in the PMP, we will be able to explicitly compute an optimal solution to the regularized optimal control problem:

$$\underset{\mathbf{u}\in\mathcal{U}_n}{\text{minimize}}\ \phi(x(T)) \tag{9}$$

where x denotes the unique solution to (8) associated with $\mathbf{u} \in \mathcal{U}_n$. Without any loss of generality, we assume that the vector fields g^i, $2 \le i \le n$ are bounded (this avoids the blow-up phenomenon in (8)). Also, we are interested in letting $\varepsilon \downarrow 0$, hence, in the sequel, ε is taken in $(0, 1]$. Proceeding in the spirit of Lemma 2.3, and after possibly increasing $R \ge 0$, we may assume that for every $\theta \in \Theta$, for every $\varepsilon \in (0, 1]$, every admissible solution $x_{\theta,\varepsilon}$ to (17) satisfies

$$\forall t \in [0, T], \ \ \|x_{\theta,\varepsilon}(t)\|_{\mathbb{R}^n} \le R.$$

Since the control system is affine w.r.t. the input $\mathbf{u}$ and the admissible control set is compact, for every $(\theta, \varepsilon) \in \Theta \times (0, 1]$, the OCP (9) has an optimal solution $\mathbf{u}_{\theta,\varepsilon} \in \mathcal{U}_n$ and $x_{\theta,\varepsilon}$ stands for the associated solution to the control system. Next, we denote by $H_{\theta,\varepsilon} : \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^n \to \mathbb{R}$ the Hamiltonian associated with this OCP and defined as:

$$H_{\theta,\varepsilon}(x, p, v) := p \cdot f_\theta(x) + v_1 p \cdot g^1_\theta(x) + \varepsilon \sum_{j=2}^{n} v_j p \cdot g^j(x),$$

for every $(x, p, v) \in \mathbb{R}^n \times \mathbb{R}^n \times \mathbb{R}^m$ and every $(\theta, \varepsilon) \in \Theta \times (0, 1]$.

Proposition 3.1. *Fix $(\theta, \varepsilon) \in \Theta \times (0, 1]$ and let $(x_{\theta,\varepsilon}, \mathbf{u}_{\theta,\varepsilon})$ be an optimal solution to (9). There is an absolutely continuous function $p_{\theta,\varepsilon} : [0, T] \to \mathbb{R}^n$ (written as a row vector) satisfying the (backward) adjoint system:*

$$\begin{cases} \dot p_{\theta,\varepsilon}(t) & = -D_x H_{\theta,\varepsilon}(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t), \mathbf{u}_{\theta,\varepsilon}(t)) \quad \text{a.e. } t \in [0, T], \\ p_{\theta,\varepsilon}(T) & = -\nabla\phi(x_{\theta,\varepsilon}(T))^\top. \end{cases} \tag{10}$$

Moreover, the optimal control $u_{\theta,\varepsilon}$ is given by:

$$\begin{cases} u^1_{\theta,\varepsilon}(t) = \dfrac{p_{\theta,\varepsilon}(t) \cdot g^1_\theta(x_{\theta,\varepsilon}(t))}{\sqrt{(p_{\theta,\varepsilon}(t) \cdot g^1_\theta(x_{\theta,\varepsilon}(t)))^2 + \varepsilon^2 \sum_{k=2}^{n} (p_{\theta,\varepsilon}(t) \cdot g^k(x_{\theta,\varepsilon}(t)))^2}}, \\ u^j_{\theta,\varepsilon}(t) = \dfrac{\varepsilon p_{\theta,\varepsilon}(t) \cdot g^j_\theta(x_{\theta,\varepsilon}(t))}{\sqrt{(p_{\theta,\varepsilon}(t) \cdot g^1_\theta(x_{\theta,\varepsilon}(t)))^2 + \varepsilon^2 \sum_{k=2}^{n} (p_{\theta,\varepsilon}(t) \cdot g^k(x_{\theta,\varepsilon}(t)))^2}}, \end{cases} \tag{11}$$

for $2 \le j \le n$ and for almost every $t \in [0, T]$.

Proof. The existence of $p_{\theta,\varepsilon}$ follows from the application of the PMP to (9). Additionally, the triplet $(x_{\theta,\varepsilon}, p_{\theta,\varepsilon}, \mathbf{u}_{\theta,\varepsilon})$ satisfies the Hamiltonian maximization condition:

$$\mathbf{u}_{\theta,\varepsilon}(t) \in \underset{\omega\in B[0,R]}{\operatorname{argmax}}\ H_{\theta,\varepsilon}(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t), \omega) \quad \text{a.e. } t \in [0, T],$$

that can be rewritten as follows: for a.e. $t \in [0, T]$, $\mathbf{u}_{\theta,\varepsilon}(t)$ is a solution to the linear optimization problem under a convex constraint:

$$\min_{\omega\in B[0_{\mathbb{R}^n},R]} F_{\theta,\varepsilon,t}(\omega) := \omega_1 p_{\theta,\varepsilon}(t) \cdot g^1_\theta(x_{\theta,\varepsilon}(t)) + \varepsilon \sum_{j=2}^{n} \omega_j p_{\theta,\varepsilon}(t) \cdot g^j(x_{\theta,\varepsilon}(t)).$$

If the minimum of $F_{\theta,\varepsilon,t}$ is achieved at some point ω in $B(0_{\mathbb{R}^n}, R)$, then, one must have $\nabla F_{\theta,\varepsilon,t}(\omega) = 0$. This means that $p_{\theta,\varepsilon}(t)$ should be orthogonal to $g^1_\theta(x_{\theta,\varepsilon}(t))$ and to each vector $g^j(x_{\theta,\varepsilon}(t))$ for $1 \leq j \leq n$. Hence, using (H3), the vector $p_{\theta,\varepsilon}(t)$ should be zero, but, since the adjoint equation is linear w.r.t. the covector, $p_{\theta,\varepsilon}$ is necessarily identically equal to zero which is a contradiction. It follows that for a.e. $t \in [0, T]$, $\mathbf{u}_{\theta,\varepsilon}(t)$ belongs to the boundary of $B[0_{\mathbb{R}^n}, R]$. To obtain (11), we use the Hamiltonian maximization condition. For a fixed value of t, the optimal control is a solution to an NLP[1] where the cost is linear and the constraint set is $B[0_{\mathbb{R}^n}, 1]$. Since the optimal control takes its values on the unit sphere of $\mathbb{R}^n$, (11) follows from a direct application of the KKT conditions (see also [29, 30]). □

Note that once $(x_{\theta,\varepsilon}, \mathbf{u}_{\theta,\varepsilon})$ is fixed, the covector $p_{\theta,\varepsilon}$ as defined by (10) is uniquely defined (since no initial-terminal constraints are considered). This property is no longer valid when such additional conditions on the the initial and terminal state are considered. For technical purpose, we set for $2 \leq j \leq n$:

$$\begin{cases} \psi_1(x,p) := \dfrac{p \cdot g^1_\theta(x)}{\sqrt{(p \cdot g^1_\theta(x))^2 + \varepsilon^2 \sum_{k=2}^n (p \cdot g^k(x))^2}}, \\ \psi_j(x,p) := \dfrac{\varepsilon p \cdot g^j_\theta(x)}{\sqrt{(p \cdot g^1_\theta(x))^2 + \varepsilon^2 \sum_{k=2}^n (p \cdot g^k(x))^2}}, \end{cases} \tag{12}$$

so that from (12), $u^1_{\theta,\varepsilon}$ and $u^j_{\theta,\varepsilon}$ can be expressed in feedback form of state and costate functions as follows:

$$u^1_{\theta,\varepsilon}(t) = \psi_1(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t)) \ \text{ and } \ u^j_{\theta,\varepsilon}(t) = \psi_j(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t)), \tag{13}$$

for a.e. $t \in [0, T]$ and for $2 \leq j \leq n$. The next proposition follows from the preceding computations.

Proposition 3.2. *Let $(\theta, \varepsilon) \in \Theta \times (0, 1]$ and consider an extremal $(x_{\theta,\varepsilon}, p_{\theta,\varepsilon}, u_{\theta,\varepsilon})$ of (9) such that $u_{\theta,\varepsilon}$ is an optimal solution to (9). Then, the state-adjoint system can be written*

$$\begin{cases} \dot{x}_{\theta,\varepsilon}(t) = f_\theta(x_{\theta,\varepsilon}(t)) + \psi_1(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t)) g^1_\theta(x_{\theta,\varepsilon}(t)) \\ \qquad\qquad + \varepsilon \displaystyle\sum_{j=2}^n \psi_j(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t)) g^j(x_{\theta,\varepsilon}(t)), \\ \dot{p}_{\theta,\varepsilon}(t) = -p_{\theta,\varepsilon}(t) D f_\theta(x_{\theta,\varepsilon}(t)) - \psi_1(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t))\, p_{\theta,\varepsilon}(t) D g^1_\theta(x_{\theta,\varepsilon}(t)) \\ \qquad\qquad - \varepsilon \displaystyle\sum_{j=2}^n \psi_j(x_{\theta,\varepsilon}(t), p_{\theta,\varepsilon}(t))\, p_{\theta,\varepsilon}(t) D g^j(x_{\theta,\varepsilon}, (t)), \end{cases} \tag{14}$$

for a.e. $t \in [0, T]$, together with the mixed initial-terminal condition

$$x_{\theta,\varepsilon}(0) = \xi \ \text{ and } \ p_{\theta,\varepsilon}(T) = -\nabla\phi(x_{\theta,\varepsilon}(T))^\top. \tag{15}$$

[1] Nonlinear Programming Problem.

Remark 3.3. Given $(\theta,\varepsilon)\in\Theta\times(0,1]$, solving (14)–(15) amounts to the resolution of a two-points boundary problem which can be achieved numerically by a shooting method (see, *e.g.*, [11] and references herein). For this purpose, let us denote by $(\tilde{x}_{\theta,\varepsilon},\tilde{p}_{\theta,\varepsilon})$ the unique solution to (14) such that $\tilde{x}_{\theta,\varepsilon}(0)=\xi$ and $\tilde{p}_{\theta,\varepsilon}(0)=p_0$ where $p_0\in\mathbb{R}^n$ is a parameter. The indirect method consists in finding a zero of the shooting function $S_{\theta,\varepsilon}:\mathbb{R}^n\to\mathbb{R}^n$ where $S_{\theta,\varepsilon}$ is defined as:

$$S_{\theta,\varepsilon}(p_0):=\|\tilde{p}_{\theta,\varepsilon}(T)+\nabla\phi(\tilde{x}_{\theta,\varepsilon}(T))\|^2_{\mathbb{R}^n}.$$

This can be done by a Newton-like's method after finding an accurate guess of the initial condition (via a direct method for instance [7]). In Section 4.1, the numerical resolution of such a problem in the context of a resource allocation model is based on [26].

4. Convergence analysis and numerical illustration

For any fixed $\varepsilon\in(0,1]$, the regularized bilevel optimization problem under consideration reads as follows:

$$\underset{\theta\in\Theta}{\text{minimize}}\ J(x_{\theta,\varepsilon}(t_1),...,x_{\theta,\varepsilon}(t_N)) \tag{16}$$

$$\text{subject to}\ \begin{cases} x_{\theta,\varepsilon}\in \underset{x\in\mathrm{AC}([0,T],\mathbb{R}^n)}{\operatorname{argmin}}\ \phi(x(T))\quad\text{s.t.}\\ \dot{x}(t)=f_\theta(x(t))+u_1(t)g^1_\theta(x(t))+\varepsilon\sum_{j=2}^n u_j g^j(x(t))\ \ \text{a.e. } t\in[0,T],\\ x(0)=\xi\ \text{ and } \mathbf{u}\in\mathcal{U}_n. \end{cases} \tag{17}$$

Next, we write $(x_{\theta,\varepsilon},\mathbf{u}_{\theta,\varepsilon})$ an optimal pair to (17) where $\mathbf{u}_{\theta,\varepsilon}\in\mathcal{U}_n$. As in Problem (1)–(2), we consider optimistic solutions. Hence, for $\varepsilon\in(0,1]$, let us define a multifunction $\Psi^o_\varepsilon:\Theta\rightrightarrows\mathrm{AC}([0,T],\mathbb{R}^n)$ as:

$$\Psi^o_\varepsilon(\theta):=\underset{x_{\theta,\varepsilon}\in\mathrm{AC}([0,T],\mathbb{R}^n)}{\operatorname{argmin}}\{J(x_{\theta,\varepsilon}(t_1),...,x_{\theta,\varepsilon}(t_N))\ ;\ x_{\theta,\varepsilon}\text{ is a solution to (17)}\},$$

where $\theta\in\Theta$. Moreover, let $\mathcal{A}_{\theta,\varepsilon}$ denotes the set of admissible solutions to the Cauchy problem in (17) where $\mathbf{u}\in\mathcal{U}_n$. A solution to the above bilevel problem has to be understood as follows.

Definition 4.1. A pair $(\theta_\varepsilon,x_{\theta_\varepsilon,\varepsilon})\in\Theta\times\mathrm{AC}([0,T],\mathbb{R}^n)$ is said to be a *solution* to (16)–(17) if:

$$\forall\theta\in\Theta,\ J(x_{\theta_\varepsilon,\varepsilon}(t_1),...,x_{\theta_\varepsilon,\varepsilon}(t_N))\le J(x_{\theta,\varepsilon}(t_1),...,x_{\theta,\varepsilon}(t_N)), \tag{18}$$

where for all $\theta\in\Theta$, $x_{\theta,\varepsilon}\in\Psi^o_\varepsilon(\theta)$. □

Observe that if $(\theta_\varepsilon,x_{\theta_\varepsilon,\varepsilon})$ is an optimal solution to (16)–(17), then, $(x_{\theta_\varepsilon,\varepsilon},\mathbf{u}_{\theta_\varepsilon,\varepsilon})$ is a solution to the OCP (17) with $\theta:=\theta_\varepsilon$. By employing a similar argumentation as for proving Proposition 2.4, one can show that for each $\varepsilon\in(0,1]$, Problem (16)–(17) has an optimal solution $(\theta_\varepsilon,x_{\theta_\varepsilon,\varepsilon})$ under one of the two following hypotheses:

(H_ε) For every $(\varepsilon, \theta) \in (0,1] \times \Theta$, the lower-level OCP (17) has a unique solution $x_{\theta,\varepsilon} \in \mathcal{A}_{\theta,\varepsilon}$.

(H'_ε) For every $(\theta, x_\theta) \in \Theta \times \mathcal{A}_\theta$ such that x_θ is optimal for (2), for all $\varepsilon \in (0,1]$, there is an optimal solution $x_{\theta,\varepsilon}$ to (17) such that $(x_{\theta,\varepsilon})$ strongly-weakly converges to x_θ as $\varepsilon \downarrow 0$.

Note that for every $\varepsilon > 0$, $(x_{\theta_\varepsilon,\varepsilon}, \mathbf{u}_{\theta_\varepsilon,\varepsilon})$ is an optimal pair to the OCP (17) with $\theta := \theta_\varepsilon$. Our aim now is to relate solutions to (16)–(17) to a solution to (1)–(2) as ε tends to zero. Since Θ is compact, we may assume (up to taking a subsequence) that (θ_ε) converges to θ^* as ε goes to zero where $\theta^* \in \Theta$. We start by examining for a fixed value of the parameter $\theta \in \Theta$, the behavior of optimal solutions to the lower-level problem (17) as $\varepsilon \downarrow 0$ (see also [29, 30]).

Proposition 4.2. *For every $\theta \in \Theta$, there is a pair $(x_\theta^*, u_\theta^*) \in \mathrm{AC}([0,T], \mathbb{R}^n) \times \mathcal{U}$ where x_θ^* is the unique solution to the Cauchy problem*

$$\left\{ \begin{array}{ll} \dot{x}(t) & = f_\theta(x(t)) + u_\theta^*(t) g_\theta(x(t)) \quad \text{a.e. } t \in [0,T], \\ x(0) & = \xi, \end{array} \right. \tag{19}$$

such that, up to a sub-sequence, $(x_{\theta,\varepsilon})$ strongly-weakly converges to x_θ^* as ε tends to zero. Moreover, (x_θ^*, u_θ^*) is an optimal solution to (2).

Proof. Using the hyptheses (H2)–(H3), the sequence $(x_{\theta,\varepsilon})$ is uniformly bounded w.r.t. ε in $L^\infty([0,T], \mathbb{R}^n)$ and so is $(\dot{x}_{\theta,\varepsilon})$. Since the velocity set associated with the system (8) is convex, the existence of (x_θ^*, u_θ^*) as well as the strong-weak convergence follows from the application of [9, Theorem 1.11]. Now, given an arbitrary $u \in \mathcal{U}$, the augmented control $\mathbf{u} := (u, 0, ..., 0)$ belongs to $\mathcal{U}_n$ so that by optimality of $u_{\theta,\varepsilon}$, one has $\phi(x_{\theta,\varepsilon}(T)) \leq \phi(x(T))$ where x is the unique solution to the Cauchy problem $\dot{x} = f_\theta(x) + u(t) g_\theta^1(x)$, $x(0) = \xi$. By letting $\varepsilon \downarrow 0$, we deduce that (x_θ^*, u_θ^*) is an optimal solution to (2) □

We now examine the convergence of an optimal sequence $(\theta_\varepsilon, x_{\theta_\varepsilon,\varepsilon})$ to (16)–(17) to a solution to the bilevel problem (1)–(2) as $\varepsilon \downarrow 0$. For this purpose, let

$$\alpha := \inf_{\theta \in \Theta} \{ J(x_\theta(t_1), ..., x_\theta(t_N)) \; ; \; x_\theta \in \Psi^o(\theta) \},$$

and let us introduce the following hypothesis:

(H”) Every solution x_θ to (2) is the limit as $\varepsilon \downarrow 0$ (for the strong-weak topology) of a sequence $(x_{\theta,\varepsilon}) \in \mathcal{A}_{\theta,\varepsilon}^{\mathbb{N}}$ such that for every $\varepsilon \in (0,1]$, $x_{\theta,\varepsilon}$ is a solution to (17).

Our main result is as follows.

Theorem 4.3. *Suppose that for every $\varepsilon \in (0,1]$, Problem* (16)–(17) *has a solution $(\theta_\varepsilon, x_{\theta_\varepsilon,\varepsilon})$. Then, there is $(\theta^*, x_{\theta^*}) \in \Theta \times \mathcal{A}_{\theta^*}$ such that, up to a subsequence, (θ_ε) converges to θ^*, $(x_{\theta_\varepsilon,\varepsilon})$ strongly-weakly converges to x_{θ^*} as $\varepsilon \downarrow 0$, and x_{θ^*} is a solution to* (17). *Furthermore, if either* (H) *or* (H”) *is verified, then, (θ^*, x_{θ^*}) is an optimal solution to* (1)–(2) *and $J(x_{\theta_\varepsilon,\varepsilon}(t_1), ..., x_{\theta_\varepsilon,\varepsilon}(t_N)) \to \alpha$ as $\varepsilon \downarrow 0$.*

Proof. For every $\varepsilon \in (0,1]$, the control system satisfied by the pair $(x_{\theta_\varepsilon,\varepsilon}, \mathbf{u}_{\theta_\varepsilon,\varepsilon})$ can be written:

$$
\begin{aligned}
\dot{x}_{\theta_\varepsilon,\varepsilon}(t) &= f_{\theta_\varepsilon}(x_{\theta_\varepsilon,\varepsilon}(t)) + u^1_{\theta_\varepsilon,\varepsilon}(t) g_{\theta_\varepsilon}(x_{\theta_\varepsilon,\varepsilon}(t)) + \varepsilon \sum_{j=2}^{n} u^j_{\theta_\varepsilon,\varepsilon}(t) g^j(x_{\theta_\varepsilon,\varepsilon}(t)) \\
&= f_{\theta^*}(x_{\theta_\varepsilon,\varepsilon}(t)) + u^1_{\theta_\varepsilon,\varepsilon}(t) g_{\theta^*}(x_{\theta_\varepsilon,\varepsilon}(t)) + \tilde{r}_\varepsilon(t),
\end{aligned}
$$

where for a.e. $t \in [0,T]$, $\tilde{r}_\varepsilon(t)$ is defined by:

$$
\begin{aligned}
\tilde{r}_\varepsilon(t) := {}& f_{\theta_\varepsilon}(x_{\theta_\varepsilon,\varepsilon}(t)) - f_{\theta^*}(x_{\theta_\varepsilon,\varepsilon}(t)) + u^1_{\theta_\varepsilon,\varepsilon}(t)(g_{\theta_\varepsilon}(x_{\theta_\varepsilon,\varepsilon}(t)) - g_{\theta^*}(x_{\theta_\varepsilon,\varepsilon}(t))) \\
&+ \varepsilon \sum_{j=2}^{n} u^j_{\theta_\varepsilon,\varepsilon}(t) g^j(x_{\theta_\varepsilon,\varepsilon}(t)).
\end{aligned}
$$

Due to (H1), one has $\|\tilde{r}_\varepsilon\|_{L^2([0,T],\mathbb{R}^n)} \to 0$ as $\varepsilon \downarrow 0$. Note that for every $\varepsilon \in (0,1]$, the control $u^1_{\theta_\varepsilon,\varepsilon}$ satisfies $|u^1_{\theta_\varepsilon,\varepsilon}(t)| \leq 1$ for every $\varepsilon \in (0,1]$ and a.e. $t \in [0,T]$. From [9, Theorem 1.11], there is $u_{\theta^*} \in \mathcal{U}$ such that, up to a sub-sequence, $(x_{\theta_\varepsilon,\varepsilon})$ strongly-weakly converges to x_{θ^*} over $[0,T]$ where x_{θ^*} denotes the unique solution to $\dot{x} = f_{\theta^*}(x) + u_{\theta^*}(t) g^1_{\theta^*}(x)$, $x(0) = \xi$. Now, $x_{\theta_\varepsilon,\varepsilon}$ is by definition a solution to the lower-level OCP (17). Let us then consider the control $(u, 0, ..., 0) \in \mathcal{U}_n$ where $u \in \mathcal{U}$ is arbitrary. Obviously, one has $\phi(x_{\theta_\varepsilon,\varepsilon}(T)) \leq \phi(x(T))$ where x is the unique solution to the Cauchy problem $\dot{x} = f_\theta(x) + u(t) g^1_\theta(x)$, $x(0) = \xi$ over $[0,T]$. Letting $\varepsilon \downarrow 0$ proves that x_{θ^*} is a solution to (2).

Suppose now that (H) is fulfilled. Since for every $\varepsilon \in (0,1]$ and for every $\theta \in \Psi^o_\varepsilon(\theta)$, one has

$$
J(x_{\theta_\varepsilon,\varepsilon}(t_1), ..., x_{\theta_\varepsilon,\varepsilon}(t_N)) \leq J(x_{\theta,\varepsilon}(t_1), ..., x_{\theta,\varepsilon}(t_N)), \tag{20}
$$

letting $\varepsilon \downarrow 0$ gives us:

$$
J(x_{\theta^*}(t_1), ..., x_{\theta^*}(t_N)) \leq J(x_\theta(t_1), ..., x_\theta(t_N)), \tag{21}
$$

where (x_θ) is the uniform limit of $(x_{\theta,\varepsilon})$ as $\varepsilon \downarrow 0$. Now, $x_{\theta,\varepsilon}$ is a solution to the lower problem (17), so, for every $u \in \mathcal{U}$, $\phi(x_{\theta,\varepsilon}(T)) \leq \phi(\tilde{x}_{\theta,\varepsilon}(T))$ where $\tilde{x}_{\theta,\varepsilon}$ is the unique solution to the regularized control system associated with the control $(u, 0, ..., 0) \in \mathcal{U}_n$. So, letting again $\varepsilon \downarrow 0$, we deduce that for every $u \in \mathcal{U}$, one has $\phi(x_\theta(T)) \leq \phi(x(T)))$ where x denotes the unique solution to the Cauchy problem $\dot{x} = f_\theta(x) + u(t) g^1_\theta(x)$, $x(0) = \xi$. Hence, x_θ is the unique solution to (2). If we combine the preceding property with (21), we deduce that (θ^*, x_{θ^*}) is an optimal solution to the bilevel optimization problem (1)–(2).

Suppose now that (H") is fulfilled and let $x_\theta \in \Psi^o(\theta)$ and $(x_{\theta,\varepsilon})$ as in (H"). Since $(\theta_\varepsilon, x_{\theta_\varepsilon,\varepsilon})$ is optimal for (16)–(17), we have (20) so that letting $\varepsilon \downarrow 0$ gives us (21). As $x_\theta \in \Psi^o(\theta)$ is arbitrary, this proves that (θ^*, x_{θ^*}) is optimal for (16)–(17).

Finally, under (H) or (H"), (θ^*, x_{θ^*}) is optimal, and in conclusion we deduce that $\alpha = J(x_{\theta^*}(t_1), ..., x_{\theta^*}(t_N))$. By the uniform convergence of $(x_{\theta_\varepsilon,\varepsilon})$ to x_{θ^*} as $\varepsilon \downarrow 0$, we obtain that $J(x_{\theta_\varepsilon,\varepsilon}(t_1), ..., x_{\theta_\varepsilon,\varepsilon}(t_N)) \to \alpha$ as wanted. This ends the proof. □

Remark 4.4. As for proving the existence of a solution to the bilevel optimization problem (1)–(2) or to (16)–(17), it is not evident to bypass the hypothesis that the lower-level problem has always a unique solution. We observe that (H") is in the same spirit as (H') and that it concerns only the original problem. It would be interesting to explore more into details if for any fixed value of $\theta \in \Theta$, any optimal trajectory of the lower-level OCP can be approached by an optimal solution to the regularized lower problems as $\varepsilon \downarrow 0$. □

Recall that u_{θ^*} (the optimal control associated with x_{θ^*}) is a concatenation of bang arcs (for which $u_{\theta^*}(t) = +1$ or $u_{\theta^*}(t) = -1$ depending on the sign of the switching function) and singular arcs (for which u_{θ^*} is with intermediate values in $[-1, 1]$) and that the number of switchings of the control can be infinite. We have the following property related to convergence of optimal controls at the lower-level.

Corollary 4.5. *Under the hypotheses of Theorem* 4.3*, suppose that* $u_{\theta^*}(t) = +1$ *for almost every* t *in some time interval* $[t_1, t_2] \subset [0, T]$*. Then, one has* $u^1_{\theta_\varepsilon,\varepsilon}(t) \to u_{\theta^*}(t)$ *for a.e.* $t \in [t_1, t_2]$.

Proof. Recall that for a.e. $t \in [0, T]$ such that $\varphi_{\theta^*}(t) \neq 0$, one has:

$$u_{\theta^*}(t) = \frac{p_{\theta^*}(t) \cdot g^1_{\theta^*}(x_{\theta^*}(t))}{|p_{\theta^*}(t) \cdot g^1_{\theta^*}(x_{\theta^*}(t))|},$$

and that $(x_{\theta_\varepsilon,\varepsilon})$ strongly-weakly converges to x_{θ^*} as $\varepsilon \downarrow 0$. Moreover, the sequence of covectors $(p_{\theta_\varepsilon,\varepsilon})$ satisfies the following properties:

- $(p_{\theta_\varepsilon,\varepsilon})$ is uniformly bounded in $L^\infty([0, T], \mathbb{R}^n)$ (this follows from the adjoint equation, the boundedness of admissible controls, and the fact that the sequence $(\nabla\psi(x_{\theta_\varepsilon,\varepsilon}(T)))$ is uniformly bounded w.r.t. ε) ;
- The velocity set associated with the adjoint equation is linear since the state-adjoint system is affine w.r.t. the control.

Hence, up to taking a sub-sequence, we may assume that there is $p \in \mathrm{AC}([0, T], \mathbb{R}^n)$ such that, $(p_{\theta_\varepsilon,\varepsilon})$ strongly-weakly converges to p over $[0, T]$. Moreover, by taking the limit as $\varepsilon \downarrow 0$ in the state-adjoint system, one can see that p satisfies (7), hence, $p = p_{\theta^*}$ over $[0, T]$. Recall that $(g^1_{\theta_\varepsilon})$ uniformly converges to g_{θ^*} over $B[0, R]$ as $\varepsilon \downarrow 0$. From the uniform convergence over $[0, T]$ of $(x_{\theta_\varepsilon,\varepsilon}, p_{\theta_\varepsilon,\varepsilon})$ to $(x_{\theta^*}, p_{\theta^*})$ as $\varepsilon \downarrow 0$, we deduce that:

$$u^1_{\theta_\varepsilon,\varepsilon}(t) = \psi_1(x_{\theta_\varepsilon,\varepsilon}(t), p_{\theta,\varepsilon}(t)) \to \frac{p_{\theta^*}(t) \cdot g^1_{\theta^*}(x_{\theta^*}(t))}{|p_{\theta^*}(t) \cdot g^1_{\theta^*}(x_{\theta^*}(t))|} = \mathrm{sign}(\varphi_{\theta^*}(t)) \quad \text{a.e. } t \in [t_1, t_2],$$

as $\varepsilon \downarrow 0$, which ends the proof. □

Remark 4.6. If the solution u_{θ^*} to the lower-level problem in (1)–(2) has a singular arc over a time interval $[t_1, t_2]$, convergence of optimal controls almost everywhere is no longer valid since $t \mapsto p_{\theta^*}(t) \cdot g^1_{\theta^*}(x_{\theta^*}(t))$ vanishes over $[t_1, t_2]$. Nevertheless, weak convergence of $(u_{\theta_\varepsilon,\varepsilon})$ to u_{θ^*} in $L^2([0, T], \mathbb{R})$ remains [29, 30]. □

4.1. Illustration of the regularization technique on a bilevel problem

We wish now to illustrate our methodology on a concrete example arising in the field of microbial resource allocation models. Our objective is to approach a solution to a bilevel optimization problem by a solution to its regularized version for a (fixed) small value of the parameter $\varepsilon > 0$. This means that we will solve (16)–(17) for some value of $\varepsilon > 0$ in this biological context.

1. Presentation of the model. Let us introduce the microalgae model borrowed from [21, 20]. For more details on the modeling part, we refer to these articles and references herein. The resource allocation problem we consider is as follows

$$\min_{|u(\cdot)|\leq 1} z(T) \ \text{ s.t. } \begin{cases} \dot{c} = k_p \dfrac{pI}{K+pI}(1-c) - k_R c(1-c-p), & c(0) \in (0,1), \\ \dot{p} = \dfrac{(1+u(t))}{2} k_R c(1-c-p) - k_P \dfrac{p^2 I}{K+pI}, & p(0) \in (0,1), \\ \dot{z} = -\dfrac{k_p pI}{K+pI}, & z(0) = 0. \end{cases} \tag{22}$$

Here, (c,p) are state variables and z is an auxiliary state to recast the original Lagrange problem in [21] into a Mayer problem. Note that the original control is $v \in [0,1]$ that we have written $v = \frac{1+u}{2}$ with $u \in [-1,1]$. The parameter $I > 0$ stands for light intensity. The parameter $\theta := (K, k_R, k_p)$ to be determined so that the optimal solution of (22) fits experimental data is with values in $\Theta := I_1 \times I_2 \times I_3$ where for $i = 1,2,3$, $I_i \subset \mathbb{R}_+^*$ is a non-empty compact interval. Given measurements of the state variables at the instants t_i, $1 \leq i \leq N$, this leads to the bilevel optimization problem

$$\underset{\theta \in \Theta}{\text{minimize}} \sum_{i=1}^{N} ((c_\theta(t_i) - \tilde{c}_i)^2 + (p_\theta(t_i) - \tilde{p}_i)^2),$$

where $(c_\theta, p_\theta, z_\theta)$ is an optimal solution to the above OCP and $\tilde{c}_i, \tilde{p}_i$ are given data at the instants $t_i \in [0,T]$. If we set $x := (c,p,z)$, the control system can be rewritten

$$\dot{x} = f_\theta(x) + u(t) g_\theta^1(x),$$

where $f_\theta, g_\theta^1 : \mathbb{R}^3 \to \mathbb{R}^3$ are given by the resource allocation model (note that f_θ and g_θ^1 are of class C^∞ over $\Omega \times \mathbb{R}_-$ with $\Omega := \{(c,p) \in [0,1] \times [0,1] \ ; \ c + p \leq 1\}$ representing the invariant domain for the dynamics). Since the control variable u appears only in the second component of the controlled vector field g_θ^1, we may set $g^2(x) := (1,0,0)^\top$ and $g^3(x) := (0,0,1)^\top$ without any loss of generality, so that the regularized control system writes:

$$\dot{x} = f_\theta(x) + u_1(t) g_\theta^1(x) + \varepsilon u_2(t) g^2(x) + \varepsilon u_3(t) g^3(x),$$

together with the control constraint $u_1^2 + u_2^2 + u_3^2 \leq 1$. Also, g^2 and g^3 are constant, so, the adjoint equation becomes

$$\dot{p}_{\theta,\varepsilon}(t) = -p_{\theta,\varepsilon}(t) D f_\theta(x(t)) - u_1(t) p_{\theta,\varepsilon}(t) D g_\theta^1(x(t)) \quad \text{a.e. } t \in [0,T],$$

together with the terminal condition $p_{\theta,\varepsilon}(T) = (0,0,-1)^\top$.

2. Resolution of the lower OCP. In the context of (22), Figure 1 depicts for a fixed value of the parameter $\theta \in \Theta$ an optimal solution to the lower-level OCP (2) as well as solutions to the lower-level OCP (17) for small values of the parameter $\varepsilon > 0$. The solution of (2) is obtained via a direct method [6] whereas the resolution of (17) is done via an indirect method solving the two-points boundary value problem given by (14) in this setting. Simulations have been carried out with `Julia (v1.11.4)` using the `OptimalControl.jl` [6] and `DifferentialEquations.jl` [26] packages to solve respectively the OCP (2) and the OCP (17). Following Proposition 4.2, regularized trajectories converge to the solution of the original lower problem which is of type bang-singular-bang with two switching times [21].

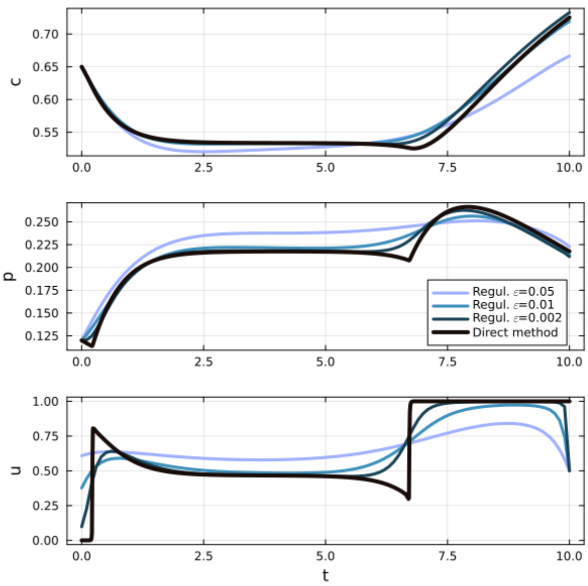

Figure 1: Optimal solutions to the lower OCP (17) obtained for small values of $\varepsilon > 0$ via an indirect method which consists in solving the boundary value problem (14) in the case of the resource allocation problem. The bang-singular-bang solution to the lower OCP (2) is obtained via a direct method.

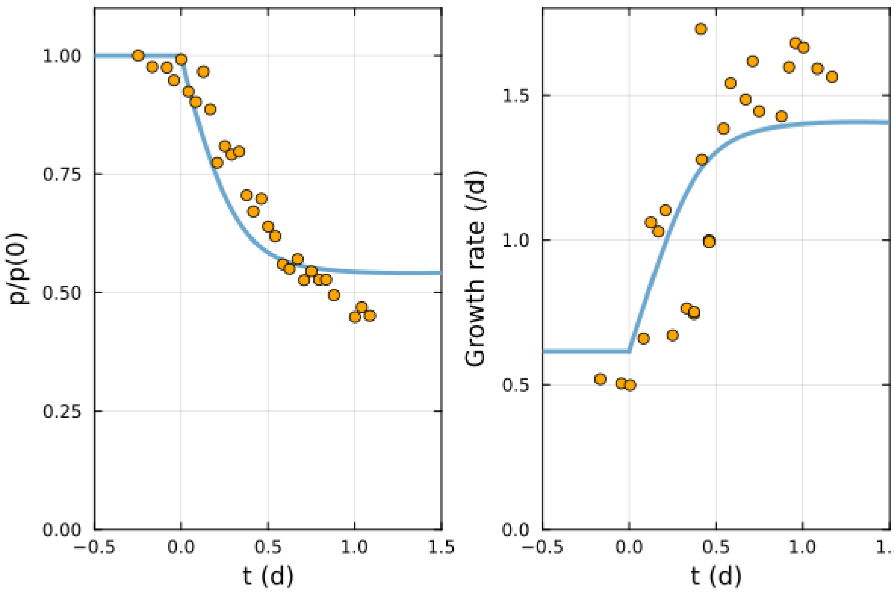

Figure 2: Bilevel optimization problem to represent microalgae dynamics in response to a light upshift. The trajectory obtained as a solution of the bilevel problem fits relatively well the experimental data from [25]. The x-axis represents time (in days d), while the y-axis shows the variables p and the cellular growth rate μ.

3. Approximation of the bilevel problem. We now search the parameter values such that an optimal solution to the lower problem fits a given data set. For this purpose, we consider the experiment of [25] with the microalgae *Thalassiosira weisflogii*, in which the chlorophyll content (an indicator of the variable p) and the cellular growth rate $\mu := k_R c(1-c-p)/(1-c)$ were measured after a light upshift [21].

The bilevel problem (16)–(17), modified to consider the available measurements, is solved for $\varepsilon = 0.01$ using the Nelder-Mead method with the `NLopt.jl` package for the upper problem [18], while solving the lower-level OCP (17) via an indirect method using `DifferentialEquations.jl` package. In conclusion, the resulting optimal trajectory is in relatively good agreement with the experimental data, see Figure 2. We also obtained as a solution an optimal value of the parameter θ, namely $\theta^* = (86.3, 5.30, 2.31)$ (in $\mu\text{mol} \cdot \text{m}^{-2}\text{s}^{-1} \times \text{d}^{-1} \times \text{d}^{-1}$).

5. Concluding remarks

In this paper, we have introduced a classical regularization procedure to approach solutions to a bilevel optimization problem involving an OCP at the lower-level. The primary novelty lies in applying this technique (originally developed for optimal control) within a bilevel optimization framework, with a focus on computational aspects. Moreover, the technique we employ, which is based on the indirect method in optimal control, differs from other approaches that solve the lower-level problem via a direct method. A key advantage of the indirect method after regularization is that it provides an exact expression for the optimal control, whereas without regularization techniques, one needs to know the optimal control's structure in advance. Note that we can also expect the regularized lower-level OCP to have a unique solution (although we do not have a certificate for this).

The convergence of the resulting approximated problem has been obtained in a simple framework with a single input and no mixed initial-terminal constraint. Moreover, we have presented a concrete case study to validate our approach. It could be possible to extend our methodology to the more general case where the lower OCP is with multiple inputs, a free terminal time, and with mixed initial-terminal constraints. Note that this regularization technique applies only to affine control systems w.r.t. the inputs which already cover a large class of application models in several fields such as in mechanics or in biology where the control enters linearly into the system.

One of the primary concerns in this bilevel problem is that the optimal control problems (at the lower-level) may have multiple solutions since they are non-convex and non-linear w.r.t. the state. It follows that, at the upper level, the evaluation of the cost function may not necessarily lead to the minimal value among all possible optimal solutions. We supposed that either the solution at the lower-level is unique (which is a standard assumption in the bilevel setting) or that optimal solutions to the original lower-level OCP can be approached by optimal solutions to the regularized ones. A deeper insight into the question of uniqueness of optimal controls for Mayer OCP governed by affine control systems w.r.t. the input could be addressed as well an examination whether optimal solutions to the non-perturbed lower-level OCP can be approached by optimal solutions to the regularized OCP. Such issues could deserve further developments in the context of bilevel optimization involving an OCP at the lower-level.

Acknowledgments. We are grateful to A. Rapaport, and to J. Salomon for fruitful discussions on the subject.

Hedy was a remarkable person with a visionary approach to Mathematics. He leaves behind the memory of a colleague with whom, over the years, many discussions took

place at the University of Montpellier concerning optimal control and optimization. This article is naturally dedicated to him.

References

[1] A. Abdel Wahab, P. Bettiol: *Necessary optimality conditions for a class of bilevel problems*, J. Optim. Theory Appl. 208/1 (2026), art.no. 30, 30p.

[2] A. Abdel Wahab, P. Bettiol: *Necessary optimality conditions for minimax multiprocesses*, Appl. Math. Optim. 91/1 (2024), art.no. 15, 32p.

[3] T. Bayen, F. Mairet: *Approximation of chattering arcs in optimal control problems governed by mono-input affine control systems*, Set-Valued Var. Analysis 32/3 (2024), art.no. 21, 21p.

[4] B. Bonnard, M. Chyba: *Singular Trajectories and their Role in Control Theory*, Mathématiques et Appl. Vol. 40, Springer, Paris (2002).

[5] U. Boscain, B. Piccoli: *Optimal Syntheses for Control Systems on 2-D Manifolds*, Mathématiques et Appl. Vol. 43, Springer, Berlin (2004).

[6] J. B. Caillau, O. Cots, J. Gergaud, P. Martinon, S. Sed: *OptimalControl.jl: a Julia package to model and solve optimal control problems with ODE's*, url: 10.5281/zenodo.13336563 (2024).

[7] J. B. Caillau, R. Ferretti, E. Trélat, H. Zidani: *An algorithmic guide for finite-dimensional optimal control problems*, in: *Numerical Control. Part B*, Handbook of Numerical Analysis Vol. 24, E. Zuazua et al. (eds.), Elsevier/North Holland, Amsterdam (2023) 559–626.

[8] F. H. Clarke: *Functional Analysis, Calculus of Variations and Optimal Control*, Springer, London (2013).

[9] F. H. Clarke, Y. S. Ledyaev, R. J. Stern, P. R. Wolenski: *Nonsmooth Analysis and Control Theory*, Grad. Texts in Mathematics Vol. 178, Springer, New York (1998).

[10] F. H. Clarke, R. Vinter: *Optimal multiprocesses*, SIAM J. Control Optim. 27/5 (1989) 1072–1091.

[11] O. Cots, R. Dutto, S. Jan, S. Laporte: *A bilevel optimal control method and application to the hybrid electric vehicle*, Optimal Control Appl. Methods 46/3 (2025) 1104–1119.

[12] S. Dempe: *Foundations of Bilevel Programming*, Nonconvex Optimization and Its Applications Vol. 61, Kluwer, Dordrecht (2002).

[13] S. Dempe, V. Kalashnikov, Gerardo A. Prez-Valds, N. Kalashnykova: *Bilevel Programming Problems: Theory, Algorithms and Applications to Energy Networks*, Energy Systems, Springer, Berlin (2015).

[14] S. Dempe, B. S. Mordukhovich, A. B. Zemkoho: *Sensitivity analysis for two-level value functions with applications to bilevel programming*, SIAM J. Optim. 22/4 (2012) 1309–1343.

[15] J. P. Dussault, M. Haddou, A. Kadrani, T. Migot: *On approximate stationary points of the regularized mathematical program with complementarity constraints*, J. Optim. Theory Appl. 186 (2020) 504–522.

[16] J.-P. Dussault, M. Haddou, T. Migot: *Mathematical programs with vanishing constraints: constraint qualifications, their applications, and a new regularization method*, Optimization 68/2-3 (2019) 509–538.

[17] F. Fisch, J. Lenz, F. Holzapfel, G. Sachs: *On the solution of bilevel optimal control problems to increase the fairness in air races*, J. Guidance Control Dynamics 35/4 (2012) 1292–1298.

[18] S. G. Johnson: *The NLopt nonlinear-optimization package*, https://github.com/stevengj/nlopt (2014).

[19] C. Lobry: *Deux remarques sur la commande bang bang des systèmes semi linéaires*, Banach Center Publ. 1 (1976) 139–145.

[20] F. Mairet, T. Bayen: *Parameter estimation for dynamic resource allocation in microorganisms: A bi-level optimization problem*, IFAC-PapersOnLine 53/2 (2020) 16814–16819.

[21] F. Mairet, T. Bayen: *The promise of dawn: microalgae photoacclimation as an optimal control problem of resource allocation*, J. Theor. Biology 515 (2021), art.no. 110597.

[22] K. Mombaur, J. P. Laumond: *From human to humanoid locomotion-an inverse optimal control approach*, Autonomous Robots 28 (2010) 369–383.

[23] J. Moreno: *Optimal time control of bioreactors for the wastewater treatment*, Optimal Control Appl. Methods 20/3 (1999) 145–164.

[24] L. S. Pontryagin, V. G. Boltyanskiy, R. V. Gamkrelidze, E. F. Mishchenko: *Mathematical Theory of Optimal Processes*, International Series of Monographs on Pure and Applied Mathematics Vol. 55, Pergamon Press, Oxford (1964).

[25] A. F. Post, Z. Dubinsky, K. Wyman, P. G. Falkowski: *Kinetics of light-intensity adaptation in a marine planktonic diatom*, Marine Biology 83 (1984) 231–238.

[26] C. Rackauckas, Q. Nie: *Differentialequations.jl – a performant and feature-rich ecosystem for solving differential equations in julia*, J. Open Research Software Vol. 5/1 (2017).

[27] A. Rapaport, D. Dochain: *Minimal time control of fed-batch processes for growth functions with several maxima*, IEEE Trans. Autom. Control 56/11 (2011) 2671–2676.

[28] A. Sinha, P. Malo, K. Deb: *A review on bilevel optimization: From classical to evolutionary approaches and applications*, IEEE Trans. Evol. Comput. 22/2 (2017) 276–295.

[29] C. Silva, E. Trélat: *Asymptotic approach on conjugate points for minimal time bang-bang controls*, Systems Control Letters 59/11 (2010) 720–733.

[30] C. Silva, E. Trélat: *Smooth regularization of bang-bang optimal control problems*, IEEE Trans. Autom. Control 55/11 (2010) 2488–2499.

[31] P. Spinelli, G. S. Rakotonirainy: *Minimum time problem synthesis*, Systems Control Letters 10/4 (1988) 281–290.

[32] J. J. Ye: *Necessary conditions for bilevel dynamic optimization problems*, SIAM J. Control Optim. 33/4 (1995) 1208–1223.

[33] J. J. Ye: *Optimal strategies for bilevel dynamic problems*, SIAM J. Control Optim. 35/2 (1997) 512–531.

[34] A. J. Zaslavski: *Necessary optimality conditions for bilevel minimization problems*, Nonlinear Analysis 75/3 (2012) 1655–1678.