ON THE CONVERGENCE OF MIN/SUP POINTS IN OPTIMAL CONTROL PROBLEMS

ADIB BAGH

Received 26 March 2001

We modify the definition of lopsided convergence of bivariate functionals to obtain stability results for the min/sup points of some control problems. In particular, we develop a scheme of finite dimensional approximations to a large class of non-convex control problems.

1. Introduction

Let \( X \) be a topological space and consider the following problem:

\[
\text{Find } \bar{x} \text{ such that } f_0(\bar{x}) = \inf_{x \in \bar{X}} f_0(x), \text{ subject to } x \in C \subset X.
\]

In case \( f_0 \) and \( C \) are convex, a variety of primal/dual numerical methods can be used to find the saddle points of a Lagrangian \( L \) associated with this problem \([4, 7]\). These methods take advantage of the fact that the search for the saddle points of \( L \) is unconstrained, or conducted over sets much simpler than \( C \). Furthermore, we can introduce penalties on \( f_0 \) in a way that regularizes \( L \) and makes it smoother. This in turn leads to a more convenient optimality condition of the form \((0, 0) \in \partial L\). In the case of a non-convex problem, similar methods can be used to find the saddle points of an augmented Lagrangian, and these points can be used to obtain a solution to the original problem (see \([7, 8]\), and \([11, \text{Chapter 10, Sections } I \text{ and } K^*]\)).

The primal/dual methods are often combined with approximating \( L \). The notion of epi/hypo convergence introduced by Attouch and Wets in \([2]\) provides a setting for constructing a sequence \( L_n \) in such way that the saddle points \((\bar{x}_n, \bar{y}_n)\) of \( L_n \) converge to the saddle points of \( L \). In this paper, however, we are interested in problems where \( L \) has a min/sup point rather than a saddle point,
and we are also interested in approximating $L$ with a sequence of Lagrangians of simpler forms.

The notion of lopsided convergence of bivariate functionals defined on $X \times Y$, where $X$ and $Y$ are topological vector spaces, was introduced by Attouch and Wets in [1] in order to study the stability of min/sup points. A very similar notion was also introduced and studied in details by Lignola, Loridan, and Morgan in [5, 6]. In this paper, we will use the concept of lopsided convergence to approximate control problems. However, the original definition of lopsided convergence in [1] cannot be used directly due to the lack of compactness of the perturbation space for most control problems. Therefore, we modify the definition of lopsided convergence to better suit our applications.

In Section 3, we review different notions of convergence for bivariate functions. In Section 4, we show how the modified lopsided convergence can provide a simple way to recover a number of stability results that already exist in the literature. Section 5 contains the main application; we develop a scheme to approximate non-convex optimal control problems with a sequence of finite dimensional problems. The modified definition of lopsided convergence we introduce in Section 3 has further applications in two level programming and problems of existence of Stackelberg equilibria in a non compact setting. These applications, however, will be left for a subsequent paper.

2. Preliminaries

In this section, we review some basic definitions of set convergence that we will use in the following sections. Most of these definitions can be found in [4, 11].

Let $(X, \tau)$ be a topological vector space. Let $C_n$ be subsets of $X$. We define the limit inferior and the limit superior of the collection $C_n$

\[
\tau - \text{Li} C_n = \{ x \in X | \exists \text{ sequence } x_n \tau \rightarrow x, x_n \in C_n \},
\]

\[
\tau - \text{Ls} C_n = \{ x \in X | \exists \text{ subsequence } x_{n_k} \tau \rightarrow x, x_{n_k} \in C_{n_k} \}.
\]  

(2.1)

We define the following notions of set convergence

\[
C_n \xrightarrow{\tau^+} C_0, \quad \text{if} \ C_0 \subseteq \tau - \text{Li} C_n,
\]

\[
C_n \xrightarrow{\tau^-} C_0, \quad \text{if} \ \tau - \text{Ls} C_n \subseteq C_0.
\]  

(2.2)

We say $C_n$ converge to $C_0$ in the Painlevé-Kuratowski sense, and we write $C_n \xrightarrow{p^+} C$, if

\[
C_n \xrightarrow{\tau^+} C, \quad C_n \xrightarrow{\tau^-} C_0.
\]  

(2.3)
Let $f : X \to \bar{\mathbb{R}}$, where $\bar{\mathbb{R}}$ is the set of extended real numbers. The function $f$ is \textit{sequentially lower semi-continuous} (lsc), if for all $x \in X$ and $x_n \xrightarrow{\tau} x$, 
\begin{equation}
\liminf_n f(x_n) \geq f(x). \tag{2.4}
\end{equation}
Similarly, $f$ is \textit{sequentially upper semi-continuous} (usc), if for all $x \in X$ and $x_n \xrightarrow{\tau} x$, 
\begin{equation}
\limsup_n f(x_n) \leq f(x). \tag{2.5}
\end{equation}
The domain of $f$ is 
\begin{equation}
\text{dom } f = \{ x \in X \mid f(x) < \infty \}. \tag{2.6}
\end{equation}
and $f$ is proper, if dom $f \neq \emptyset$ and $f$ never assumes the value $-\infty$. The \textit{epigraph} of $f$ is 
\begin{equation}
\text{epi } f = \{(x, \alpha) \in X \times \mathbb{R} \mid \alpha \geq f(x)\}. \tag{2.7}
\end{equation}
The function $f$ is \textit{coercive}, if for any $\alpha$,
\begin{equation}
\{ x \mid |f(x)| \leq \alpha \} \text{ is bounded.} \tag{2.8}
\end{equation}

We say $f_n \tau$-epi-converge to $f$ on the topological space $(X, \tau)$, and we write $f_n \xrightarrow{\tau} f$, if
\begin{enumerate}[\quad (i)]
\item $\forall x \in X, \forall x_n \xrightarrow{\tau} x$ such that 
\begin{equation}
\liminf_n f(x_n) \geq f(x). \tag{2.8}
\end{equation}
\item $\forall x \in X, \exists x_n \xrightarrow{\tau} x$ such that 
\begin{equation}
\limsup_n f(x_n) \leq f(x). \tag{2.9}
\end{equation}
\end{enumerate}
The set of minimizers of $f$ is denoted by $\text{argmin } f = \{ x \in X \mid f(x) = \inf_X f(x) \}$.

\textbf{Theorem 2.1 (see [11, Theorem 7.31]).} Let $(X, \tau)$ be a topological space and let $f_n : X \to \bar{\mathbb{R}}$ be a collection of functions that $\tau$-epi-converge to a proper function $f_0$. Then, 
\begin{equation}
\tau - \text{Lsargmin } f_n \subseteq \text{argmin } f_0. \tag{2.10}
\end{equation}
Furthermore, if $\tau - \text{Lsargmin } f_n$ is not empty, then $\liminf f_n = \text{inf } f_0$.

Let $X$ and $Y$ be topological spaces and suppose that $(\cdot, \cdot) : X \times Y \to \mathbb{R}$ is a bilinear continuous map. Let $f : X \to \bar{\mathbb{R}}$ be a convex function. We define $f^* : Y \to \bar{\mathbb{R}}$, the conjugate of $f$, by 
\begin{equation}
f^*(y) = \sup_{x \in X} \{ (x, y) - f(x) \}. \tag{2.11}
\end{equation}
Consider the functions $F_0: X \to \bar{R}$ and $G: X \to \bar{R}$, where $G$ is lsc and convex. Let $F(x) = F_0(x) + G(\Lambda(x))$, where $\Lambda: X \to X$ is a continuous map. Then, a Lagrangian of $F$, with respect to a dual space $Y$, can be defined by

$$L(x, y) = F_0(x) + \langle \Lambda(x), y \rangle - G^*(y), \quad (2.12)$$

where $G^*(y) = \sup_{x \in X} \{ \langle x, y \rangle - G(x) \}$. (2.13)

The bivariate function $L$ is a Lagrangian in the sense that

$$F(x) = \sup_{y \in Y} L(x, y). \quad (2.14)$$

We now consider a certain class of integral functionals on $L^p$ spaces. Consider first the function $f: [0, T] \times \mathbb{R}^n \to \bar{R}$. We say $f$ is a normal integrand, if $f$ is measurable in the first variable and lsc in the second. We define the integral functional $I_f$ over the space $L^r([0, T], \mathbb{R}^n)$ for $r \in [1, +\infty]$,

$$I_f(u) = \int_0^T f(t, u(t))dt. \quad (2.15)$$

Suppose $f$ is a normal integrand that is also convex in the second argument, then $I_f$ is a proper, convex, lsc (and weakly lsc) function on $L^r$. Furthermore, if $r \in [1, +\infty]$, then the conjugate of $I_f$ is defined over $(L^r([0, T], \mathbb{R}^n)^*)$ by the following equation

$$I_f^*(v) = \int_0^T f^*(t, v(t))dt, \quad (2.16)$$

where for every $t$, $f^*(t, \cdot)$ is conjugate of $f(t, \cdot)$ over $\mathbb{R}^n$ ([4, Proposition 2.1, Chapter IX]). Note that for $r = +\infty$, and a conjugate space $(L^\infty)^*$, the above formula will not be valid.

We now list the definitions that we need regarding bivariate functions. Let $K: X \times Y \to \bar{R}$ where $X$ and $Y$ are topological spaces. We say $K$ is proper, if the function $V: X \to \bar{R}$ defined by $V(x) = \sup_{y \in Y} K(x, y)$ is proper. A saddle point for $K$ is a pair $(\bar{x}, \bar{y})$ such that

$$K(\bar{x}, y) \leq K(\bar{x}, \bar{y}) \leq K(x, \bar{y}), \quad \forall x \in X, \forall y \in Y. \quad (2.17)$$

A min/sup point for $K$ is a point $\bar{y} \in X$ such that

$$\sup_{x \in X} K(x, y) \leq \sup_{y \in Y} K(x, y), \quad \forall y \in Y. \quad (2.18)$$

We write arg sp$K$ to denote the set of saddle points of $K$, and we write arg min sup$K$ to denote its set of min/sup points.

Finally, in all what follows, we will use the standard convention of $\infty - \infty = \infty$. This will allow us to deal with constraints on the variables $x$ and $y$ in a consistent manner.
3. Lopsided convergence

We start by reviewing some notions of convergence for bivariate functions. Let \((X, \tau)\) and \((Y, \sigma)\) be topological spaces and consider a sequence of bivariate functionals \(K_n : X \times Y \to \bar{R}\). The sequence \([K_n]\) epi/hypo converges to \(K_0\) [2], if

(i) For all \((x, y) \in X \times Y\), for all \(x_n \xrightarrow{\tau} x\), there exists \(y_n \xrightarrow{\sigma} y\) such that
\[
\liminf K_n(x_n, y_n) \geq K_0(x, y).
\]

(ii) For all \((x, y) \in X \times Y\), for all \(y_n \xrightarrow{\sigma} y\), there exists \(x_n \xrightarrow{\tau} x\) such that
\[
\limsup K_n(x_n, y_n) \leq K_0(x, y).
\]

Epi/hypo convergence implies the convergence of saddle points.

Theorem 3.1 (see [2]). If \(K_n\) epi/hypo converge to \(K_0\), then
\[
\tau - Ls\arg sp K_n \subset \arg sp K_0.
\]

Definition 3.2 is the original definition of the lopsided convergence [3].

Definition 3.2. The sequence \([K_n]\) lopsided converges to \(K_0\), denoted by \(K_n \xrightarrow{lo} K_0\), if

(i) For all \((x, y) \in X \times Y\), for all \(x_n \xrightarrow{\tau} x\), there exists \(y_n \xrightarrow{\sigma} y\) such that
\[
\liminf K_n(x_n, y_n) \geq K_0(x, y).
\]

(ii) For all \((x, y) \in X \times Y\), for all \(y_n \xrightarrow{\sigma} y\), there exists \(x_n \xrightarrow{\tau} x\) such that for all \(y\in Y\) and for all \(y_n \xrightarrow{\sigma} y\),
\[
\limsup K_n(x_n, y_n) \leq K_0(x, y).
\]

Note that in the definition of epi/hypo convergence there is a symmetry between parts (i) and (ii). The term "lopsided" is meant to emphasize the lack of such symmetry in Definition 3.2. Theorem 3.3 is the main theorem regarding lopsided convergence that concerns us.

Theorem 3.3 (see [1]). If \(K_n \xrightarrow{lo} K_0\), \(Y\) is a compact space, then
\[
\tau - Ls\argminsup K_n \subset \argminsup K_0.
\]

It is clear from the definitions that lopsided convergence implies epi/hypo convergence but the converse is not true. The compactness of \(Y\) is essential in the proof of Theorem 3.3 even though the compactness of \(Y\) is not required for
On the convergence of min/sup points in optimal control problems
the existence of a min/sup point [1]. In this paper, however, we will investigate
control problems where the space \( Y \) is a space of perturbations for the state
variables which is not compact in most cases (or not compact in a topology that
is compatible with the continuity properties the problem has). Therefore, we
introduce the following modification of Definition 3.2.

**Definition 3.4.** Consider two topologies \( \tau_1 \) and \( \tau_2 \) on the space \( X \) and assume
that \( \tau_1 \) is stronger than \( \tau_2 \). Similarly, consider two topologies \( \sigma_1 \) and \( \sigma_2 \) on \( Y \)
and assume that \( \sigma_1 \) is stronger than \( \sigma_2 \). The sequence \( \{K_n\} \) converges to \( K_0 \),
\( K_n \xrightarrow{\text{lo}2} K_0 \), if

(i) For all \((x, y) \in X \times Y\), for all \( x_n \xrightarrow{\tau_2} x \), there exists \( y_n \xrightarrow{\sigma_1} y \) such that
\[ \liminf_n K_n(x_n, y_n) \geq K_0(x, y). \] (3.7)

(ii) For all \( x \in X \), there exists \( x_n \xrightarrow{\tau_1} x \) such that for any sequence \( \{y_n\} \) such
that \( K_n(x_n, y_n) \) is bounded below, there exists a subsequence \( n_k \), there exists
\( \epsilon_k \downarrow 0 \), there exists \( w_{nk} \in Y \) such that eventually
\[ K_0(x, w_{nk}) = K_0(x_{nk}, y_{nk}) - \epsilon_k. \] (3.8)

Note that when \( \tau_1 = \tau_2, \sigma_1 = \sigma_2, \) and \( Y \) is compact, the two definitions of
lopsided convergence are equivalent. Now we can prove the following theorem.

**Theorem 3.5.** If \( K_n \xrightarrow{\text{lo}2} K_0 \) and \( K_0 \) is proper then
\[ t_2 = \text{argminsup} K_n \subset \text{argminsup} K_0. \] (3.9)

The proof of Theorem 3.5 follows immediately from the following lemma
and Theorem 2.1.

**Lemma 3.6.** Let \( X, Y \) be topological spaces as in Definition 3.4. Let \( K_n : X \times Y \to \mathbb{R} \), where \( K_0 \) is proper. Let
\[ V_n(x) = \sup_{y \in Y} K_n(x, y), \quad n = 0, 1, 2, \ldots. \] (3.10)

Assume

(i) For every \((x, y) \in X \times Y\), for all \( x_n \xrightarrow{\text{lo}2} x \), there exists \( y_n \xrightarrow{\text{lo}2} y \) such that
\[ \liminf_n K_n(x_n, y_n) \geq K_0(x, y). \] (3.11)

(ii) For every \( x \in X \), there exists \( x_n \xrightarrow{\text{lo}2} x \) such that for any sequence \( \{y_n\} \),
where \( K_n(x_n, y_n) \) is bounded below, there exists a subsequence \( n_k \) and there
exist \( \epsilon_k \downarrow 0 \) and \( w_{nk} \in Y \) such that eventually
\[ K_0(x, w_{nk}) = K_0(x_{nk}, y_{nk}) - \epsilon_k. \] (3.12)
Then,  
\[ \tau_2 = \text{LSargmin } V_0 \subset \text{argmin } V_0. \]  
(3.13)

Furthermore, if \( \tau_2 = \text{LSargmin } V_0 \) is not empty, then  
\[ \liminf V_0 = \liminf V_0. \]  
(3.14)

**Proof.** Condition (i) implies that for all \((x, y) \in X \times Y\), for all \( x_n \xrightarrow{\tau_2} x \), there exists \( y_n \xrightarrow{\sigma_1} y \) such that  
\[ \liminf_{n} \sup_{y \in Y} K_0(x_n, y) \geq \liminf_{n} \sup_{y \in Y} K_0(x_n, y_n) \geq K_0(x, y). \]  
(3.15)

Hence,  
\[ \liminf_{n} V_0(x_n) \geq V_0(x). \]  
(3.16)

We also claim that for all \( x \in X \), there exists \( x_n \xrightarrow{\tau_1} x \) such that  
\[ \limsup_{n} V_0(x_n) \leq V_0(x). \]  
(3.17)

Assume not, and assume that \( V_0(x) < +\infty \). Then, for every sequence \( x_n \xrightarrow{\tau_1} x \), we have \( \limsup_{n} V_0(x_n) = \beta > V_0(x) \). Thus, there exists a subsequence of \( V_0(x_n) \) that converges to \( \beta \). To simplify the notation, we will also use \( n \) to index this subsequence. Hence, there exist \( x > 0 \), \( y \in Y \), a sequence \( \{y_n\} \) in \( Y \), and there exists \( n_0 \) such that for all \( n \geq n_0 \),  
\[ K_0(x, y_n) > K_0(x, y) + \epsilon. \]  
(3.18)

Assumption (ii) implies that there exist \( w_{n} \) and \( \epsilon_{n} \) such that eventually  
\[ K_0(x, w_{n}) > K_0(x, y_n) - \epsilon_{n}. \]  
(3.19)

Hence, for some \( n_2 > n_0 \) and for some \( \epsilon' > 0 \), we have  
\[ K_0(x, w_{n}) > K_0(x, y) + \epsilon'. \]  
(3.20)

Thus,  
\[ V_0(x) \geq K_0(x, y) + \epsilon', \]  
(3.21)

which contradicts the definition of \( V_0(x) \).

Finally (3.16) and (3.17) yield the required conclusion via Theorem 2.1. \( \square \)

**Remark 3.7.** In **Definition 3.4** of lopsided convergence, the requirements that \( y_n \xrightarrow{\sigma_2} y \) in part (i) and that \( w_{n} \xrightarrow{\sigma_2} 0 \) in part (ii) were not used in the proof of **Lemma 3.6**. In fact, these requirements are not needed for the conclusion of **Theorem 3.5**. However, they are needed to make the notion of lopsided convergence stable with respect to perturbation of some classes of functions (e.g., the class of \( \tau_2 \times \sigma_2 \) continuous biaffine functions).
Remark 3.8. There is a number of variations of the definition of epi/hypo convergence. These variations consider more than one topology on the spaces $X$ and $Y$ (see [3, 12]) in a way that is very similar to what we have in Definition 3.4.

Remark 3.9. At this point, it may seem that the original definition of lopsided convergence can be directly used in problems defined over infinite dimensional spaces when the Lagrangian is coercive in the $y$ variable. It may seem that all what is needed in this case is to restrict the Lagrangian to a bounded set of $Y$ and to choose a weak enough topology that will make this bounded set compact. This, however, is not true, as the last remark of Section 5 will illustrate. In fact, in some of our applications, a topology that is weak enough to make the modified space $Y$ compact will make it very difficult for us to verify condition (ii) of Definition 3.2.

4. Applications I

In our first application, we will consider problems with Lagrangians of the following form (for examples and details see [4, Chapter VII]):

$$L(u, v) = J(u) + \langle v, \Phi(u) \rangle + \delta_A(u) - \delta_B(v),$$  \hspace{1cm} (4.1)

where $L : U \times V \to \mathbb{R}$ and $U$ and $V$ are Hilbert spaces, $J : U \to \mathbb{R}$ is a convex, lsc, Gâteau differentiable function, $A$ and $B$ are nonempty closed convex subsets of $U$ and $V$, respectively. $\delta_A$ and $\delta_B$ are the indicator functions of these sets ($0$ on the set and $+\infty$ outside it).

The map $\Phi : A \to V$ is Lipschitzian in the sense that

$$\|\Phi(u_2) - \Phi(u_1)\| \leq c\|u_1 - u_2\|, \hspace{1cm} \forall u_1, u_2 \in A. \hspace{1cm} (4.2)$$

And for every $v \in V$, the function $u \mapsto \langle v, \phi(u) \rangle$ is convex and lsc on $A$, where $\langle \cdot, \cdot \rangle$ is the inner product of $V$. Note that our assumptions imply that for all $v$, for all $u_n \xrightarrow{n} u$, we have

$$\liminf_{n \to \infty} \langle v, \phi(u_n) \rangle \geq \phi(v, u). \hspace{1cm} (4.3)$$

A number of primal/dual methods that take advantage of the special form of the Lagrangian can be used to attack this problem (cf. [4, Chapter VII]). Most of these methods, however, require the following additional conditions on $J$, $A$, and $B$:

(a) If is bounded
(b) $J$ is coercive in the sense that for all $u_1, u_2 \in A$,

$$J(u_2) - J(u_1) + \alpha \|u_2 - u_1\| \geq \alpha \|u_2 - u_1\|, \hspace{1cm} \alpha > 0. \hspace{1cm} (4.4)$$

In this section, we are interested in problems that do not satisfy the above conditions. Therefore, we perturb the original Lagrangian in such a way that the
resulting Lagrangians will be
\[ L_n(u,v) = J(u) + \langle v, \Phi(u) \rangle + \frac{1}{n} \| u \|^2 + \delta_A(u) - \delta_B(v), \]  \hspace{1cm} (4.5)
where \( B_n = B \cap r_n \mathcal{B} \), where \( \mathcal{B} \) is the unit ball in \( V \), and \( r_n \) increases to infinity.

Clearly, \( L_n \) satisfies conditions (a) and (b). We now show that \( L_n \) lopsided converges to \( L \).

**Theorem 4.1.** Let \( \tau_1 \) and \( \tau_2 \) be, respectively, the norm and the weak topology on \( U \). Similarly, let \( \sigma_1 \) and \( \sigma_2 \) be, respectively, the norm and weak topology on \( V \). Let \( L_0 \) and \( L_n \) be defined by (4.1) and (4.5), then
\[ L_n \rightharpoonup L_0. \]  \hspace{1cm} (4.6)

**Proof.** Let \( (u,v) \in U \times V \) such that \( v \in B \). Let \( u_n \rightharpoonup u, v_n \to v \), and \( v_n \in B_n \). Due to (4.3), and since every convex lsc function on a Banach space is also weakly lsc, we immediately obtain
\[ \liminf_n \left[ J(u_n) + \delta_A(u_n) + \langle v_n, \Phi(u_n) \rangle + \frac{1}{n} \| u_n \|^2 - \delta_B(v_n) \right] \geq J(u) + \delta_A(u) + \langle v, \Phi(u) \rangle - \delta_B(v), \]  \hspace{1cm} (4.7)
which is condition (i) of Definition 3.4.

To verify condition (ii), we let \( u \in U \). For any sequence \( v_n \), there exists \( \epsilon_n \) such that
\[ J(u) + \delta_A(u) + \langle v_n, \Phi(u) \rangle - \delta_B(v_n) \geq J(u) + \delta_A(u) + \langle v_n, \Phi(u) \rangle + \frac{1}{n} \| u_n \|^2 - \delta_B(v_n) \] \[ - \epsilon_n. \]  \hspace{1cm} (4.8)
\[ \square \]

**Remark 4.2.** The above approximation method can be used when \( B \) is unbounded. Moreover, in some control problems, the set \( B \) has the following form:
\[ B = \{ u \in L^p(\Omega) \text{ such that } \| \text{grad} u \| \leq M \}, \]  \hspace{1cm} (4.9)
where \( \Omega \) is a bounded open set in \( \mathbb{R}^n \) and \( M \) is some constant. The set \( B \) is bounded in \( L^p \) due to the Poincaré inequality. However, the bound of \( B \) (the Poincaré constant for the region \( \Omega \)) may not be available. In this case, it is convenient numerically to use the above approximation and replace \( B \) with an increasing sequence of balls in \( L^p \) even though \( B \) is bounded.

Finally, we show how the modified lopsided convergence can be used to recover some stability results regarding some classical control problems.
Consider the following problem: Minimize
$$J(u) = \int_Q |y(t,x,u) - y_0|^2 dx dt$$
subject to
$$\|\text{grad } y(t,x,u)\| \leq 1 \quad \text{a.e.} \quad (t,x) \in Q,$$
where $y_0(t,x) \in L^2([0,T];H^1_0(\Omega))$ is a given function, and $y(t,x,u) \in L^2([0,T];H^1_0(\Omega))$ is the unique weak solution of the following dynamics:
$$\begin{align*}
\frac{\partial y}{\partial t} - M\Delta y &= f + u \quad \text{in } Q, \\
y &= 0 \quad \text{on } \partial \Omega \times [0,T], \\
y(0,x,u) &= y_0(x), \quad \forall x \in \Omega,
\end{align*}$$
where $\Omega$ is a subset in $\mathbb{R}^n$, $T > 0$, $Q = [0,T] \times \Omega$, $y_0 \in L^2(\Omega)$, and $f \in L^2(Q)$. The Lagrangian of the above problem can be expressed as (see [4, Section 3.1] for details)
$$L(u,v) = \int_Q \left[ \frac{1}{2} \|\text{div}(u)\|^2 - \text{div}(u) \cdot z + u \cdot \text{grad} \phi + u \cdot v \right] dx dt,$$
$$\text{where } \phi(t,x) = y(t,x,0), \quad z(t,x,u) = y(t,x,u) - y(t,x,0), \quad z_\epsilon(t,x) = y_\epsilon(t,x) - y(t,x,0), \quad \text{and } \phi = y(0) \text{ (see [4, Section 1, Chapter VII]).}$$
Solving the original control problem corresponds to finding $u$, where the following infimum is attained:
$$\inf_{u \in L^2(Q)} \sup_{v \in B} L(u,v),$$
where $A = \{u \mid u \in L^2(Q), \text{div } u \in L^2(Q)\}$ and $B = \{v \mid v \in L^2(Q) \text{ and } |v(t,x)| \leq 1 \ \text{a.e.}\}$. Note that, despite the fact that new formulation of the problem involves a supremum and an infimum, the format of $L$ and the simplicity of the set $B$ make the new formulation of the problem easier to solve numerically. The Lagrangian in (4.13), however, does not satisfy requirement (b) for the direct application of numerical methods. Furthermore, (4.5) may not have a saddle point under our current assumptions, and the standard methods of approximating saddle points (via epi/hypo convergence) cannot be used. Therefore, for $\epsilon > 0$, we introduce
$$L_\epsilon(u,v) = \int_Q \left[ \frac{1}{2} \|\text{div}(u)\|^2 - \text{div}(u) \cdot z + (1 + \epsilon) |u|^2 + u \cdot \text{grad} \phi + u \cdot v \right] dx dt.$$
Now for every $\epsilon > 0$, the Lagrangian $L_\epsilon(u,v)$ satisfies conditions $(a)$ and $(b)$, and a standard numerical method, such as the Uzawa method [4], can be used.
to find a saddle point \((u, v)\) for \(L\). If we show that, as \(\epsilon \rightarrow 0\), \(L\) lopsidedly converges to \(L\), then we will know that every cluster point of \(\{u_{\epsilon}\}\) is a min/sup point of \(L\). This is precisely the claim of the following theorem.

**Theorem 4.3** (also [4, Proposition 3.4, Chapter VII]). Let \(L_{\epsilon}(u, v)\) and \(L(u, v)\) be defined by (4.13) and (4.15). As \(\epsilon \rightarrow 0\), we have

\[
L_{\epsilon}(u, v) \overset{\text{lo2}}{\longrightarrow} L(u, v),
\]

\[w = \text{LSargminsup}L_{\epsilon}(u, v) \subset \text{argminsup}L(u, v).\]  

Furthermore, if \(w = \text{LSargminsup}L_{\epsilon}(u, v) \neq \emptyset\), then

\[
\inf V \sup U L_{\epsilon}(u, v) \longrightarrow \inf V \sup U L(u, v).\]  

**Proof.** The \(L^2\) norm in the space \(A = \{u | u \in L^2(\Omega)^n, \text{div} v \in L^2(\Omega)\}\) is weakly lsc, and the conditions of **Definition 3.4** can be easily verified with the appropriate choices for the topologies \(\tau_1, \tau_2, \sigma_1,\) and \(\sigma_2\).

\(\square\)

5. **Application II**

In this section, we consider the following control problem, minimize, over \(L_p = L_p([0, T]; \mathbb{R}^n)\), with \(p \in [2, +\infty)\), the functional

\[
I_1(u) = \int_0^T \phi(t, u(t), x(t))dt,
\]

where \(x(\cdot) : [0, T] \rightarrow \mathbb{R}^k\) is a solution to

\[
Lx = u,
\]

and \(L\) is an operator from \(C([0, T], \mathbb{R}^k)\) to \(L^1\). The states and controls are subject to the constraint

\[
\{t, f(t, u(t))\} \in \text{gph} E \quad \text{a.e.},
\]

where \(E\) is a set-valued map from \([0, T]\) to \(\mathbb{R}^j\) and \(\text{gph} E\) is the graph of the map \(E\).

**Assumptions on the operator \(L\).** The operator \(L\) has an inverse \(\Lambda\) such that \(u_{\epsilon}\) converges weakly to \(u\) in \(L^p\) which implies that \(\Lambda u_{\epsilon}\) converges pointwise to \(\Lambda u\).

The following are examples of operators that satisfy our assumptions.

**Example 5.1.** \(L\) is the linear differential operator given by

\[
x'(t) = A(t)x(t) + B(t)u(t), \quad x(0) = u,
\]

with the standard assumptions on \(A\) and \(B\).
On the convergence of min-sup points in optimal control problems

Example 5.2. Let \( L \) be a differential operator representing an evolution equation,
\[
    x'(t) = g(t, u(t), x(t)), \quad x(0) = a,
\]
under the usual assumptions of growth and Lipschitz continuity on \( g \).

Example 5.3. Any differential operator \( L \) whose inverse can be expressed by
an integral equation of the form
\[
    x(t) = \int_0^T g(t, s)u(s)ds,
\]
where \( g(t, \cdot) \) is in \( L^2 \).

Assumptions on the constraint (5.3). The map \( f: [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^j \) is measurable in the first argument, continuous in the rest. The map \( E \) is nonempty and convex-valued with a closed graph. We also posit a growth condition, there exists a function \( s \in L^2 \) such that
\[
    \sup_{z \in E(t)} \|z\| \leq s(t) \quad \text{a.e.}
\]

The constraint (5.3) can model a combination of constraints such as
\[
    u(t) \in U(t), \quad x(t) \in V(t), \quad M(t, x(t), u(t)) \in C(t),
\]
where \( U, V, \) and \( C \) are set-valued maps. In this case, we only need to define
\[
    E(t) = U(t) \times V(t) \times C(t),
    f(t, u(t), x(t)) = (u(t), x(t), M(t, x(t), u(t))).
\]

Assumptions on the cost function \( I \). We will assume that \( I \) is finite-valued, weakly lower semi-continuous, and strongly upper semi-continuous over \( L^2 \).

These assumptions can be formulated in terms of standard conditions on \( \phi \)
(i) The map \( \phi: [0, T] \times \mathbb{R}^* \times \mathbb{R} \to \mathbb{R} \) is measurable in the first argument, convex in the second, and continuous in the third.
(ii) There exists a function \( \Phi: \mathbb{R}^+ \to \mathbb{R}^+ \) such that for all \( t, y, \) and \( z \) such that
\[
    \Phi(|z|) \leq \phi(t, z, y).
\]
(iii) For all \( u \in L^2 \), there exists a neighborhood \( V \) of \( u \) and a function \( h \in L^1 \) such that for every \( u' \in V \)
\[
    \phi(t, u'(t), x'(t)) \leq h(t),
\]
where \( x' = \Lambda u' \).
The above conditions imply that $I_1$ is finite-valued. Note that assuming that $I_1$ is finite-valued does not limit the scope of this model, since we will deal with the constraints on the controls when we deal with constraint (5.3). Furthermore, conditions (i) and (ii) imply the weak lower semi-continuity of $I_1$ (see [4, Theorem 2.1, Chapter VIII]). The (strong) upper semi-continuity of $I_1$ follows from (iii), Fatou’s lemma, and Lebesgue dominated convergence theorem.

Our goal in this section is to construct finite dimensional approximations for the above non-convex (partially convex) control problem. In [9, 10], Rockafellar developed the full duality theory for a similar type of control problems. However, he was mainly interested in cases where the existence of saddle points for the Lagrangians is guaranteed, and therefore he required the cost function to be convex (in the control and the state) and the dynamics to be linear. We will be able to relax these conditions, since we are only interested in the partial duality of the problem and since we only assume the existence of min/sup points.

To simplify the notation, we will assume that $f : [0, T] \times \mathbb{R}^n \times \mathbb{R}^k \to \mathbb{R}^n$. Hence, for all $u$ and for all $x$ such that $\mu \lambda H x = u$, we have $f(t, u(t), x(t)) \in L^2$ because of the growth condition on $E$. We introduce an exact but finite penalty function.

Let $\theta : [0, T] \times \mathbb{R}^n$ be a normal, convex (in the second argument) integrand

$$
\theta(t, w) = \begin{cases} 0, & \text{if } (t, w) \in \text{gph } E, \\ > 0, & \text{otherwise}. \end{cases}
$$

More specifically, we define

$$
\theta(t, z) = d(z, E(t)),
$$

where $d(z, C) = \inf_{c \in C} \|c - z\|$ for any closed subset $C$. Note that $\theta(t, \cdot)$ is finite over $\mathbb{R}^n$.

Now the problem can be expressed as minimizing, over $L^p$, the functional

$$
F(u) = I_1(u) + I_2(w),
$$

where $I_1 : L^p \to \mathbb{R}$ is defined by

$$
I_1(u) = \int_0^T \phi(t, u(t), x(t)) dt,
$$

and

$$
I_2(w) = \int_0^T \theta(t, w(t)) dt.
$$

A Lagrangian associated with the above problem is given by $L : L^p \times L^2 \to \mathbb{R}$ with

$$
L(u, v) = I_1(u) - I_2^*(v) + \Gamma(v, w),
$$

where $\Gamma(v, w)$ is the convex conjugate of $\theta(t, z) = d(z, E(t))$, and $\Lambda u = x$. 

Adib Bagh 47
On the convergence of min/sup points in optimal control problems

\[ I^*_2 = \int_0^T \theta^*(t, u(t)) \, dt, \quad (5.18) \]
\[ \theta^*(t, z) = \sup_{y \in \mathbb{R}^n} \left\{ \langle y, z \rangle - \theta(t, y) \right\}, \quad (5.19) \]

where \( \theta \) is still given by (5.13), and

\[ MGammaH(u, v) = \int_0^T v(t) w(t) \, dt, \quad (5.20) \]

where again \( w(t) = f(t, u(t), x(t)) \), and \( MLambdaHu = x \). A direct calculation of \( \theta^*(t, z) \), when \( \theta \) is given by (5.13), gives us

\[ \theta^*(t, z) = \sigma_{E(t)}(z) + \delta_B(z), \quad (5.21) \]

where \( \sigma_{E(t)}(z) = \sup_{y \in E(t)} \langle y, z \rangle \) is the support function of the set \( E(t) \), and \( B \) is the unit ball in \( \mathbb{R}^n \) (see [11, Example 11.26, Chapter 11] for details). Note that \( \theta^*(t, \cdot) \) is convex, proper, and lsc. Moreover, it is coercive over \( \mathbb{R}^n \) since \( \theta(t, \cdot) \) is finite everywhere.

The fact that \( F(u) = \sup_{v \in \mathbb{R}^n} L(u, v) \) follows from the definition of the conjugate function and from the fact that the conjugate of \( I_2 \) is actually given by (5.18) (see also (2.12) in the Preliminaries).

Using exact penalties for the joint state and control constraints causes serious computational complications (see [10] for details). Therefore, we introduce a sequence of finite non exact penalties \( \theta_n : [0, T] \times \mathbb{R}^n \rightarrow \mathbb{R} \) such that \( \theta_n(t, \cdot) \) increases continuously to \( \theta(t, \cdot) \). More specifically, we will take Moreau envelopes of \( \theta_n \),

\[ \theta_n(t, z) = \inf_{y \in \mathbb{R}^n} \left\{ \theta(t, y) + \frac{n}{2} \| y - z \|^2 \right\}, \quad (5.22) \]

These approximating functions are strictly convex and differentiable. Moreover, the conjugates of \( \theta_n^*(t, w) \) are given by (see [11, Chapter 11])

\[ \theta_n^*(t, z) = \sigma_{E(t)}(z) + \delta_B(z) + \frac{1}{2n} \| z \|^2. \quad (5.23) \]

Since \( \theta_n(t, \cdot) \) is strictly convex, \( \theta_n^*(t, \cdot) \) is differentiable on the interior of its domain (see [11, Theorem 11.13]).

We now use a method developed by Wright [12] to discretize the control problem in the primal and dual variables at the same time. Let \( \Pi_n \) be an increasing sequence of partitions of \([0, T]\)

\[ \Pi_n = (0 = a_0 < a_1 < \cdots < a_{n-1} < a_n = T), \quad (5.24) \]

such that \( |a_n - a_{n-1}| \to 0 \) uniformly in \( x \), as \( n \to + \infty \).
We define the following sets:

\[ \begin{align*}
U_n &= \{ u \in L^p \mid u \text{ is constant on } [a_n s - 1, a_n s], s = 0, 1, \ldots, T \}, \\
V_n &= \{ v \in L^2 \mid v \text{ is constant on } [a_n s - 1, a_n s], s = 0, 1, \ldots, T \}.
\end{align*} \]

The sequence of approximating Lagrangians that we will consider is

\[ \begin{align*}
L_n(u, v) &= I_1(u) + \delta U_n(u) - I_2^*(v) + \Gamma(u, v),
\end{align*} \]  

where \( \delta U_n \) and \( \delta V_n \) are the indicator functions of \( U_n \) and \( V_n \), respectively, and

\[ \begin{align*}
I_1(u) &= \int_0^T \phi(t, u(t)) dt, \\
I_2^*(v) &= \int_0^T \theta^*(t, v(t)) dt + \delta V_n(v),
\end{align*} \]

where \( I_2^* : L^2 \to \mathbb{R} \) and \( \theta^* \) is given by (5.19).

The standard convention of \( -\infty = \infty \) implies that \( L(u, v) = +\infty \), if \( u \notin U_n \) or \( v \notin V_n \). Moreover, \( L(u, v) = -\infty \), if \( u \notin U_n \) and \( v \notin V_n \). Now, the problem of finding the min/sup points of \( L_n \) is a finite dimensional problem. Also note that for a step function \( v \), calculating \( \theta^*(t, v(t)) \) reduces to solving a number of simple problems of convex minimization (concave maximization) in \( \mathbb{R}^n \).

Furthermore, the discretized problem will be governed by a difference equation (see [12] for details).

Since \( L_n \) is not convex in \( u \) (\( \phi \) is not convex in \( x \) and \( \Gamma \) is not convex in \( u \)), \( L \) and \( L_n \) may not possess saddle points. However, we can use the method of augmented Lagrangians developed by Rockafellar in [8] to find the min/sup points of \( L_n \). The idea of this method is to construct an augmented Lagrangian \( L_n(u, v, r) \), for some \( r \in (0, +\infty) \), and then use some standard primal/dual numerical method to find \( (\bar{u}_n, \bar{v}_n) \), the saddle points of \( L_n(u, v, r) \). For every \( n \), the point \( \bar{u}_n \) will then be a min/sup point for \( L_n \) (see [8] and [11, Section 4*], Chapter 11), also see [7, Chapter 17] for an actual algorithm using augmented Lagrangians). This approach will allow us to approximate, discretize, and numerically solve the original problem despite its lack of convexity.

Finally, we state the main theorem of the section which will show that our approximation scheme will actually work. 

**Theorem 5.4.** Let \( L_n \) and \( L \) be defined by (5.17) and (5.26). Then

\[ w = \arg \min \sup L_n \subset \arg \min \sup L. \]

Before we prove Theorem 5.4, we list a number of basic lemmas that we will need in our proof.
Lemma 5.5. Consider the functions $I_1$ and $I_{1,u}$ defined by (5.16) and (5.27). Consider also the sets $U_n$ (5.25). Let $u \in L^p$, then there exists a sequence $\hat{u}_n \in U_n$ such that
\begin{equation}
\lim I_{1,u}(\hat{u}_n) = I_1(u).
\end{equation}
Proof. The proof is an immediate result of the fact that step functions are dense in $L^p$ for $p \in [2, +\infty)$ and of our assumption that $I_1$ is norm continuous. □

Lemma 5.6. Let $v \in L^2$ such that $v(t) \in B \ a.e., \ where \ B \ is \ the \ unit \ ball \ in \ \mathbb{R}^n$. Then there exists a sequence $\hat{v}_n \to v$ such that $v_n \in B \cap V_n$ and
\begin{equation}
\limsup_n \int_0^T \sigma_{E_0}(\hat{v}_n(t))dt \leq \int_0^T \sigma_{E_0}(v(t))dt.
\end{equation}
Proof. Using some elementary facts from measure theory (see [12, Lemmas 1, 2] for details) and using the definition of the support function, we can find a sequence $\hat{v}_n \to v$ such that $v_n(t) \in B, \ \hat{v}_n(t) \to v(t), \ and \ v_n(t) \leq v(t) \ a.e. \ in \ [0,T]$. Hence, for all $n$, we have
\begin{equation}
\int_0^T \sigma_{E_0}(v_n(t))dt \leq \int_0^T \sigma_{E_0}(v(t))dt.
\end{equation}
and the assertion of the lemma is immediate. □

Lemma 5.7. Let $v \in L^2$ such that $v(t) \in B \ a.e. \ Then, \ there \ exists \ a \ sequence \ \hat{v}_n \to v \ such \ that \ v_n(t) \in B \cap V_n$ and
\begin{equation}
\limsup_n I^*_2(\hat{v}_n) \leq I^*_2(v).
\end{equation}
Proof. Lemma 5.6 and the definition of $I^*_2$ given by (5.19) and (5.28). □

Proof of Theorem 5.4. In light of Theorem 3.5, we only need to show that $L_n$ lopsided converges to $L$, when $\tau_1$ and $\tau_2$ are, respectively, the norm and weak topologies on $L^p$, and $\sigma_1$ and $\sigma_2$ are also, respectively, the norm and weak topologies on $L^2$.

Verifying part (i) of Definition 3.4.
For all $(u,v) \in L^p \times L^2$ such that $v \in \text{dom } I^*_2$, take $\hat{v}_n$ as in Lemma 5.7. Note that, for all $u_n \rightarrow u$, we have
\begin{equation}
\liminf_n I_1(u_n) \geq I_1(u)
\end{equation}
due to our assumption that $I_1$ is weak lsc. Moreover,
\begin{equation}
\lim_n \left[ \Gamma(\hat{u}_n, v_n) \rightarrow \Gamma(u,v) \right] = 0.
\end{equation}
It is clear now from our choice of $\hat{v}_n$ and from (5.34) and (5.35) that part (i) of Definition 3.4 is satisfied.

Verifying part (ii) of Definition 3.4.

For all $(u,v) \in L^p \times L^2$ such that $\sup (u(v) < +\infty$, take $\hat{u}_n \to u$ as in Lemma 5.5. Now, for any $v_n \in L^2$ such that $-\infty < I_{1,n}(\hat{u}_n, v_n) = I_{1,n}(\hat{u}_n) - I_{2,n}(\hat{v}_n) + \Gamma(\hat{u}_n, v_n)$, we must have $\{v_n(t)| \in B$ a.e. Hence, $\{v_n\}$ must be bounded in $L^2$. Hence,

\[
\lim_{n} \left[ \Gamma(\hat{u}_n, v_n) \to \Gamma(u, v) \right] = 0.
\]

(5.36)

Therefore, there exists $\epsilon_n \to 0$ such that eventually

\[
I_1(u) - I_2(v_n) + \Gamma(u, v_n) \geq I_{1,n}(\hat{u}_n) - I_{2,n}(\hat{v}_n) + \Gamma(\hat{u}_n, v_n) - \epsilon_n
\]

(5.37)

and part (ii) of the definition is verified.

Remark 5.8. Theorem 5.4 is only a stability result, in the sense that $\text{argminsup} L$ might be empty. However, standard conditions on $\phi$ can be added to assure that $\text{argminsup} L$ and $\text{argminsup} L$ will not be empty. In this case, the numerical method we suggested will produce a sequence of pairs $(u_n, v_n) \in L^p \times L^2$ which are the saddle points of the augmented approximating Lagrangians. The sequence $v_n$ will be bounded in $L^2$, which is important for numerical reasons, but its weak cluster point may fail to be a solution to the dual problem. More specifically, a weak \times weak cluster point of $(u_n, v_n)$, may fail to be a saddle point for the Lagrangian of the original control problem. However, as Theorem 5.4 shows, any weak cluster point of $u_n$ will be a solution of the original control problem. Finally, if we know a priori that the original control problem has saddle points, then Theorem 5.4 can be easily modified to show stability of the saddle points with respect to the discretization and approximation scheme of this section.

Remark 5.9. The convexity of $\phi$ in the second variable was needed only to obtain the weak lower semi-continuity of $I_1$. In case the cost function is known to be weakly lsc on $L^p$, no such assumption is needed. A cost function that depends on the state only would be an example of such a case.

Remark 5.10. We finally elaborate further on Remark 3.9 of Section 3; under the appropriate coercivity conditions, we may attempt to use the original definition of lopsided convergence, with $Y$ taken as a large enough ball in $L^2$ and $\sigma$ as the (relative) weak topology. In this case, $Y$ is compact in $\sigma$. However, verifying condition (ii) of Definition 3.2 is, in essence, equivalent to verifying
On the convergence of min/sup points in optimal control problems

that for all \( u_n \rightharpoonup u \) in \( Y \),

\[
\liminf_n \int \sigma_{E(t)} v_n(t) dt \geq \int \sigma_{E(t)} v(t) dt.
\]  

(5.38)

Such statement cannot be verified without further conditions on the set-valued map \( E \).

References


Adib Bagh: Department of Mathematics, University of California-Davis, CA 95616, USA
E-mail address: bagh@math.ucdavis.edu