# Inverse optimal control with polynomial optimization

Edouard Pauwels[1,2], Didier Henrion[1,2,3], Jean-Bernard Lasserre[1,2]

Draft of March 20, 2014

**Abstract**

In the context of optimal control, we consider the inverse problem of Lagrangian identification given system dynamics and optimal trajectories. Many of its theoretical and practical aspects are still open. Potential applications are very broad as a reliable solution to the problem would provide a powerful modeling tool in many areas of experimental science. We propose to use the Hamilton-Jacobi-Bellman sufficient optimality conditions for the direct problem as a tool for analyzing the inverse problem and propose a general method that attempts at solving it numerically with techniques of polynomial optimization and linear matrix inequalities. The relevance of the method is illustrated based on simulations on academic examples under various settings.

# 1 INTRODUCTION

In brief the *Inverse Optimal Control Problem* (IOCP) can be stated as follows. Given a system dynamics

$$\dot{x}(t) = f(x(t), u(t))$$

with possibly state and/or control constraints

$$x(t) \in X, \ \ u(t) \in U, \ \ t \in [0, T]$$

and a set of trajectories

$$(x(t; x_0), u(t; x_0))_{t \in [0,T], \, x_0 \in X}$$

parametrized by time and initial states, and stored in a database, the goal is to find a *Lagrangian* function

$$l : X \times U \to \mathbb{R}$$

---

[1]CNRS; LAAS; 7 avenue du colonel Roche, F-31400 Toulouse; France.

[2]Université de Toulouse; LAAS, F-31400 Toulouse, France.

[3]Faculty of Electrical Engineering, Czech Technical University in Prague, Technická 2, CZ-16626 Prague, Czech Republic

such that all state and control trajectories in the database are *optimal trajectories* for the direct Optimal Control Problem (OCP) with integral cost

$$\int_0^T l(x(t), u(t))dt,$$

with fixed or free terminal time $T$.

Inverse problems in the context of calculus of variations and control are old topics that arouse a renewal of interest in the context of optimal control, especially in humanoid robotics. Actually, even the well-posedness of the IOCP is an issue and a matter of debate between the robotics and computer science communities.

## 1.1 Context

The problem of variational formulation of differential equations (or the inverse problem of the calculus of variations) dates back to the *19-th* century. The one dimensional case is found in [9], historical remarks are found in [10, 30]. Necessary and sufficient conditions for the existence and the enumeration of solutions to this problem have been investigated since, see also [28] for a survey about recent developments. Notice that calculus of variations problems correspond to the particular choice of dynamics $f = u$ in the OCP.

Kalman first formulated the inverse problem in the context of linear quadratic regulator (LQR) [16] which triggered research efforts in this realm [3, 15, 13]. Departing from the linear case, Hamilton-Jacobi theory is used in [29] to recover quadratic value functions, in [23] to prove existence theorems for a class of inverse control problems, and in [7] to generalize results obtained for LQR. In a slightly different context, [12] linked Lyapunov and optimal value functions for optimal stabilization problems. More recently the well-posedness issue was addressed in [24] in the context of LQR. Robustness and continuity aspects with respect to Lagrangian variations were investigated in [8], results about well-posedness and experimental requirements were exposed in [2], both in the context of Dubbins dynamics and strictly convex positive Lagrangians.

On a more practical side, motivated by [4], the authors in [22] have proposed an algorithm based on the ability to solve the direct problem for parametrized Lagrangians and on a formulation of the IOCP as a (finite-dimensional) nonlinear program. Similar approaches have been proposed in the context of Markov Decision Processes [1, 27]. In a sense these methods are "blind" to the problem structure as testing optimality at each current candidate solution is performed by solving numerically the associated direct OCP. To further exploit the problem structure, we can use explicit analytical optimality conditions for the direct OCP. For instance the use of necessary optimality conditions for solving numerically the IOCP have already been proposed in recent works. In [26] the direct problem is first discretized and the IOCP is expressed as a (finite-dimensional) nonlinear program. Then the *residual* technique of [17] for inverse parametric optimization is applied to the Karush-Kuhn-Tucker optimality conditions of the nonlinear program to account for optimality of the database trajectories. On the other hand, [14] proposes to use the maximum principle for kinetic parameter estimation.

Surprisingly, in all the above references, only the seminal theoretical works of [29, 23, 7] are based on (or use) the Hamilton-Jacobi-Bellman equation (HJB in short), whereas the HJB provide a well-known sufficient condition for optimality and a perfect practical tool for verification. One reason might be that the HJB is rarely used for solving the direct OCP and is rather used only as a verification tool.

## 1.2   Contribution

We claim and hope to convince the reader that the HJB optimality equation (in fact even a certain relaxation of HJB) is a very appropriate tool not only for analyzing but also for solving the IOCP. Indeed:

(a) The HJB optimality equation provides an almost perfect criterion to express optimality for database trajectories.

(b) The HJB optimality equation (or its relaxation) *sheds light* on the many potential pitfalls related to the IOCP when treated in full generality. Among them, the most concerning issue is the ill-posedness of the inverse problem and the existence of solutions carying very little physical meaning. The HJB condition can be used as a guide to restrict the search space for candidate Lagrangians. Previous approaches to deal with this problem, implicitly or explicitly, involve strong constraints on the class of functions among which candidate Lagrangians are searched [22, 26, 8, 2]. This allows, in some cases, to alleviate the ill-posedness issue and to provide theoretical guarantees regarding the possibility to recover a true Lagrangian from the observation of optimal trajectories [8, 2]. Departing from these approaches, the search space restrictions that we propose stems from either simple relations between the candidate Lagrangian and optimal value function associated with the OCP, obvious from the HJB equation, or from purely sparsity biased estimation arguments. We do *not* require an *a priori* (and always questionable) selection of the "type" of candidate Lagrangian (e.g. coming from some "physical" observations and/or remarks).

(c) Last but not least, a relaxation of the HJB optimality equation can be readily translated into a positivity condition for a certain function on some set. If the vector field $f$ is polynomial, and the state and/or control constraint sets $X$ and $U$ are basic semi-algebraic then a natural strategy is to consider *polynomials* as candidate Lagrangians and (approximated) optimal value functions associated with the direct OCP. Within this framework, using powerful positivity certificates *à la Putinar* to look for an optimal solution of the IOCP reduces to solving a hierarchy of semidefinite programs (SDP), or linear matrix inequalities (LMI) of increasing size. Importantly, a distinguishing feature of this approach is *not* to rely on iteratively solving a direct OCP. Notice that since the 1990s, the availability of reasonably efficient SDP solvers [31] has strengthened the interest in optimization problems with polynomial data. Examples of applications of such positivity certificates are given in [18] and [20] for direct optimization and optimal control, and in [19] for inverse (static) optimization.

The paper is organized as follows. In Section 2 we properly define the IOCP. Next, in section 3, we present the conceptual ideas based on HJB theory, polynomial optimization and LMI. We emphasize that these ideas allow to highlight potential pitfalls of the

IOCP. A practical implementation of the discussed conceptual method is proposed in section 4. Finally, in section 5 we first illustrate the outputs of the method on a simple one-dimensional example and then provide promising results on more complex academic examples.

# 2   PROBLEM FORMULATION

## 2.1   Notations

If $A$ is a topological vector space, $\mathcal{C}(A)$ represents the set of continuous functions from $A$ to $\mathbb{R}$ and $\mathcal{C}^1(A)$ represents the set of continuously differentiable functions from $A$ to $\mathbb{R}$. Let $X \subseteq \mathbb{R}^{d_X}$ denote the state space and $U \subseteq \mathbb{R}^{d_U}$ denote the control space which are supposed to be compact subsets of Euclidean spaces. System dynamics are represented by a continuously differentiable vector field $f \in \mathcal{C}^1(X \times U)^{d_X}$ which is supposed to be known. Terminal state constraints are represented by a set $X_T \subseteq X$ which is also given. Let $B_n$ denote the unit ball of the Euclidean norm in $\mathbb{R}^n$, and let $\partial X$ denote the boundary of set $X$.

## 2.2   Direct OCP

Given a Lagrangian

$$l_0 \in \mathcal{C}(X \times U),$$

we consider a direct OCP of the form:

$$
\begin{aligned}
v_0(t, z) := \ &\inf_u \int_t^T l_0(x(s), u(s))ds \\
&\text{s.t. } \dot{x}(s) = f(x(s), u(s)), \\
&\quad\ x(s) \in X,\ u(s) \in U,\ s \in [t, T], \\
&\quad\ x(t) = z,\ x(T) \in X_T
\end{aligned}
\tag{OCP}
$$

with final time $T \geq t$. In this formulation, the final time $T$ might be given or free, in which case it is a variable of the problem and the *value function* $v_0$ does not depend on $t$. For the rest of this paper, direct problems of the form of (OCP) are denoted as

$$\text{OCP}(l_0, z).$$

*Remark.* More general problem classes could be considered. Indeed, there is no terminal cost and the problem is stationary. Further comments are found in next sections.

*Remark.* The infimum in (OCP) might not be attained.

## 2.3   IOCP

The inverse problem consists of recovering the Lagrangian of the direct OCP based on:

- knowledge of the dynamics $f$ as well as state and/or control constraint sets $X, U, X_T$,

- observation of optimal trajectories stored in a database indexed by some index set $I$

$$\mathcal{D} = \{(t_i, x_i, u_i)\}_{i \in I} \in ([0,T] \times X \times U)^I.$$

*Remark.* For complete trajectories, the index set $I$ is a continuum. For real world data, it is a finite set.

More specifically, the inverse problem consists of finding $l \in \mathcal{C}(X \times U)$ such that $\mathcal{D}$ is a subset of optimal trajectories of problem $\mathrm{OCP}(l, x_i(t))$ for all $t \in [0, T]$. A solution of this problem would be an operator mapping back $\mathcal{D}$ to $l$ such that the input data satisfy optimality conditions for problem $\mathrm{OCP}(l, x_i(t))$.

# 3  HAMILTON-JACOBI-BELLMAN FOR THE INVERSE PROBLEM

## 3.1  Hamilton-Jacobi-Bellman theory

For the rest of this paper, we define the following linear operator acting on Lagrangians and value functions:

$$\mathcal{L} \colon (l, v) \to l + \frac{\partial v}{\partial t} + \frac{\partial v}{\partial x}^T f.$$

A well-known sufficient condition for optimality in problem (OCP) follows from Hamilton-Jacobi-Bellman (HJB) theory, see e.g. [11, 6].

**Proposition 3.1.** *Suppose that there exist state and control trajectories $(x_0(s), u_0(s))_{s \in [t,T]}$ such that*

$$\dot{x}_0(s) = f(x_0(s), u_0(s)), \tag{A1}$$
$$x_0(s) \in X, \ u_0(s) \in U, \ s \in [t, T],$$
$$x_0(t) = z, \ x_0(T) \in X_T.$$

*Suppose in addition that there exists a function $v_0 \in \mathcal{C}^1([t, T] \times X)$ such that*

$$\mathcal{L}(l_0, v_0)(s, x, u) \geq 0, \ \forall (s, x, u) \in [t, T] \times X \times U, \tag{1}$$
$$v_0(T, x) = 0, \ \forall x \in X_T, \tag{2}$$
$$\mathcal{L}(l_0, v_0)(s, x_0(s), u_0(s)) = 0, \ \forall s \in [t, T]. \tag{3}$$

*Then the state and control trajectories $(x_0(s), u_0(s))_{s \in [t,T]}$ are optimal solutions of the direct problem (OCP).*

*Proof.* Recall that $X, U$ are compact. So integrating (1) along *any* feasible state and control trajectories $(x(s), u(s))$ with initial state $z$ at time $t$, and using (2) yields $\int_t^T l_0(x(s), u(s)) ds \geq v_0(t, z)$ whereas using (3) one has $\int_t^T l_0(x_0(s), u_0(s)) ds = v_0(t, z)$. □

Observe that (1)-(2) are just a *relaxation* of the HJB equation and from Proposition 3.1 (1)-(3) provide a certificate of optimality of the proposed trajectory. Additional assumptions on problem structure are required to make this condition necessary, see *e.g.* [11]. Moreover, for most direct problems these conditions can not be met in the usual sense and *viscosity solutions* are needed, see *e.g.* [6]. Our approach is based on a classical interpretation as well as a relaxation of these sufficient conditions alone.

*Remark.* In the free terminal time setting, the conditions (1)-(2) can be simplified because $v_0$ does not depend on the initial time $t$ any more.

## 3.2 Inverse problem: main idea

We use HJB relations to characterize some approximate solutions of the inverse problem. The idea is based on the following weakening of Proposition 3.1.

**Proposition 3.2.** *Let $(x_0(s), u_0(s))_{s\in[t,T]}$, with $x_0(t) = z$, be such that (A1) is satisfied. Suppose that there exist a real $\epsilon$ and functions $l \in \mathcal{C}(X \times U)$, $v \in \mathcal{C}^1([0,T] \times X)$ such that*

$$\mathcal{L}(l,v)(s,x,u) \geq 0, \ \forall (s,x,u) \in [t,T] \times X \times U, \tag{4}$$

$$v(T,x) = 0, \ \forall x \in X_T, \tag{5}$$

$$\int_t^T \mathcal{L}(l,v)(s,x_0(s),u_0(s))ds \leq \epsilon. \tag{6}$$

*Then, the input trajectory $(x_0(s), u_0(s))_{s\in[t,T]}$ is $\epsilon$-optimal for problem $OCP(l,z)$.*

*Proof.* Proceeding as in the proof of Proposition 3.1, $v(t,z)$ is a lower bound on the value function of problem $OCP(l,z)$ and from the last linear constraint (6) we deduce that

$$\int_t^T l(x_0(s), u_0(s))ds \leq v(t,z) + \epsilon.$$

$\square$

Proposition 3.2 can be easily extended to the case of multiple trajectories. The main advantage is that we have a certificate of $\epsilon$-optimality. Observe that 0-optimality is in principle impossible to obtain because the optimal value function $v$ of a direct OCP is in general non-differentiable; however as soon as $v$ is continuous then by the Stone-Weierstrass Theorem $v$ can be approximated on the compact $[t,T] \times X$ as closely as desired by a polynomial and so $\epsilon$-optimality is indeed achievable for arbitrary $\epsilon > 0$.

*Remark.* The conditions (4)-(6) do not ensure that the infimum of the direct problem with cost $l(x,u)$ is attained. However the fact that a trajectory is given ensures that it is feasible.

## 3.3 Multiple and trivial solutions

The construction of multiple solutions to the inverse problem is trivial given the tools of Proposition 3.2. Consider the free terminal time setting and suppose that the pair $(l,v)$

satisfies conditions (4)-(6) for some $\epsilon > 0$. Take a differentiable $\tilde{v}$ such that $\tilde{v}(T, x) = 0$ for $x \in X_T$. Then the pair $(l - f^T v, v + \tilde{v})$ satisfies constraints (4)-(6) for the same $\epsilon$. The possibility to construct such solutions stems from the existence of trivial solutions to the problem.

As we already mentioned, the constraint (4) is positively homogeneous in $(l, v)$ and therefore, the pair $(0, 0)$ is always feasible. In other words the trivial cost $l = 0$ is always an optimal solution of the inverse problem, independently of input trajectories. But the well-posedness issue is even worse than this. Consider a pair of functions $(l, v)$ such that $\mathcal{L}(l, v) = 0$ on the domain of interest, then any feasible trajectory of the direct problem will be optimal for $l$.

**Example 1.** Consider the following one dimensional free terminal time direct setting

$$X = U = B_1, \ X_T = \{0\}, \ f(x, u) = u.$$

The pair of functions $(l, v) = (-2xu, x^2)$ satisfies constraint (4) and $\mathcal{L}(l, v) = 0$. Any feasible trajectory, $(x_0(s), u_0(s))_{s \in [t, T]}$ with $t < T$, $x_0(t) = z$, and such that (A1) is satisfied, is optimal for $\text{OCP}(l, z)$. Indeed

$$\int_t^T l(x_0(s), u_0(s)) ds = \int_t^T \frac{d}{dt}(-x_0(s)^2) ds = x_0(t)^2.$$

Such pairs are solutions of the IOCP in the sense that was proposed in the previous section. However, these solutions do not have any physical interpretation, because they do not depend on input trajectories. In addition, because the solutions of the IOCP form a *convex cone*, the existence of such solutions allows to construct multiple solutions to the IOPC. To avoid this, one possibility is to include an additional normalizing constraint of the form:

$$\mathcal{A}(\mathcal{L}(l, v)) = 1 \tag{7}$$

for some linear functional $\mathcal{A}$ on the space of continuous functions. This can be viewed as a search space reduction as we intersect the cone defined by (4) with an affine space.

*Remark.* As we already mentioned, we do not consider terminal cost and limit ourselves to stationary problems. This setting was chosen to avoid more ill-posedness of the same type and keep the presentation clear.

*Remark.* Previous practical methods [22, 26] and theoretical work [24, 8, 2] include, implicitly or explicitly, constraints of the form of (7) but only enforce them on the candidate Lagrangian $l$.


## 3.4 Considering multiple trajectories

Considering a single trajectory as input for the IOCP leads to Lagrangians that enforce closedness to this trajectory, which may have little physical meaning.

**Example 2.** Consider the direct problem with free terminal time

$$X = U = B_2, \ X_T = \{0\}, \ f(x, u) = u, \ l(x, u) = 1.$$

The optimal control is $u(x) = \frac{-x}{||x||_2}$ and the optimal value function is $v(x) = ||x||_2$. Consider the trajectory

$$(x_0(s), u_0(s))_{s \in [0,1]} = ((s - 1, 0), (1, 0))_{s \in [0,1]}.$$

This trajectory is optimal for $\text{OCP}(l, (-1, 0))$ but it is also optimal for $\text{OCP}(||u - (1, 0)||_2^2, (-1, 0))$. However, the second Lagrangian only captures a constant of the particular trajectory considered.

## 3.5 Effect of discretization

In practical settings, one rarely has access to complete trajectories. Typical experiments produce discrete samples from trajectories and possibly with additional experimental noise. The database consists of a set of $n \in \mathbb{N}$ points $\{(t_i, x_i, u_i)\}_{i=1...n}$. In this case, we replace the integral in (6) by a discrete sum:

$$\frac{1}{n} \sum_{i=1}^{n} \mathcal{L}(l, v)(t_i, x_i, u_i) \le \epsilon. \tag{8}$$

In this setting, it is generically possible to find Lagrangians that satisfy constraints (4) and (6) with $\epsilon = 0$.

**Example 3.** Consider, in a free terminal time setting a discrete dataset $\{(x_i, u_i)\}_{i=1...n}$. The pair of functions $((x, u) \to \prod_i ||x_i - x||^2 ||u_i - u||^2, 0)$ satisfies constraint (4) and (8) with $\epsilon = 0$.

Because of experimental noise and random perturbations of the input trajectories, one can find Lagrangians which fit tightly a particular sample of trajectories. This does not give much insight about the true nature of the original trajectories. Furthermore, the proposed Lagrangian may be far from optimal if the database was fed up with a different sample of trajectories. In statistics this well-known phenomenon bears the name of *overfitting* (see *e.g.* [5] for a nontechnical introduction and [32] for an overview of the mechanisms it involves). A common approach to avoid overfitting is to introduce biases in the estimation procedure with search space restrictions or regularization terms. We adopt such a strategy in the practical method described in the next section.

*Remark.* It is clear that this discretization, although intuitive, arises many questions regarding the effect of noise and of sample size in practical IOCP, rarely discussed and even mentioned in previous works. These theoretical considerations are not specific to the method we propose. They are beyond the scope of this paper and constitute a strong motivation for future research work.

# 4 A PRACTICAL IMPLEMENTATION

We have seen that the HJB theory is very useful to address well-posedness issues regarding the inverse problem. In full generality the HJB equations (or their relaxations) are computationally intractable. On the other hand, in a polynomial and semi-algebraic context, some powerful positivity certificates from real algebraic geometry permit to translate the relaxation (4) of the HJB equations into an appropriate LMI hierarchy, hence amenable to practical computation. Therefore, in the sequel, we assume that $f$ is a polynomial and $X$, $U$ and $X_T$ are basic semi-algebraic sets. Recall that $G \subset \mathbb{R}^d$ is basic semi-algebraic whenever there exists polynomials $\{g_i\}_{i=1\ldots m}$ such that $G = \{x : g_i(x) \geq 0, i = 1 \ldots m\}$. In addition, we assume that the input database is indexed by a finite set: $\mathcal{D} = \{(t_i, x_i, u_i)\}_{i=1,\ldots,n}$.

## 4.1 Problem formulation

We consider the following program

$$\begin{aligned}
&\inf_{l,v,\epsilon} \ \epsilon + \lambda ||l||_1 \\
&\text{s.t. } \mathcal{L}(l,v)(t,x,u) \geq 0, \ \ \forall (t,x,u) \in [0,T] \times X \times U, \\
&\qquad v(T,x) = 0, \ \ \forall x \in X_T, \\
&\qquad \tfrac{1}{n} \sum_{i=1}^{n} \mathcal{L}(l,v)(t_i, x_i, u_i) \leq \epsilon, \\
&\qquad \mathcal{A}(\mathcal{L}(l,v)) = 1
\end{aligned} \tag{IOCP}$$

where $l$ and $v$ are polynomials, $\epsilon$ is a real, $\lambda > 0$ is a given regularization parameter, and $||.||_1$ denotes the $\ell_1$ norm of a polynomial, *i.e.* the sum of absolute values of its coefficients when expanded in the monomial basis. The first two constraints come from the relaxation (4) of the HJB equations while the third constraint comes from the *fit* constraint (8). Finally, the last affine constraint is meant to avoid the trivial solutions that satisfy $\mathcal{L}(l,v) = 0$. The $\ell_1$ norm is not differentiable around sparse vectors (with entries equal to zero) and has therefore a sparsity promoting role which allows to bias solutions of the problem toward Lagrangians with few nonzero coefficients. This regularization affects the problem well-posedness and will prove to be essential in numerical experiments.

## 4.2 Implementation details

Linear constraints are easily expressed in term of polynomial coefficients. A classical lifting allows to express the $\ell_1$ norm minimization as a linear program. The normalization functional $\theta$ is chosen to be an integral over a box contained in $S$. To express nonnegativity of polynomials over a compact basic semi-algebraic set of the form $G = \{x : g_i(x) \geq 0, i = 1, \ldots, m\}$, we invoke powerful positivity certificates from real algebraic geometry. Indeed, if a polynomial $p$ is positive on $G$ then by Putinar's Positivstellensatz [25] it can be written

as

$$p = p_0 + \sum_{i=1}^{m} g_i p_i, \quad p_i \in \Sigma^2, \quad i = 1, \ldots, m \tag{9}$$

where $\Sigma^2$ denotes the set of sum of squares (SOS) polynomials. Hence (9) provides a useful certificate that $p$ is nonnegative on $G$. Moreover, as membership to $\Sigma^2$ reduces to semidefinite programming, the constraint (9) is easily expressed as an LMI whose size depends on the degree bound allowed for the SOS polynomials $p_i$ in (9). Therefore, replacing the positivity constraint in (IOCP) with the constraint (9) allows to express problem (IOCP) as a hierarchy of LMI problems [31] indexed by the degree bounds on the SOS in (9). Thus each LMI of the hierarchy can be solved efficiently (of course up to some size limitations). We use the SOS module of the `YALMIP` toolbox [21] to manipulate and express polynomial constraints at a high level in `MATLAB`.

# 5 NUMERICAL EXPERIMENTS

## 5.1 General setting

In our numerical experiments we considered several direct problems of the same form as (OCP). That is, we give ourselves compact sets $X$, $U$, $X_T$, the dynamics $f$, and a Lagrangian $l_0$. We take known examples for which the optimal control law can be computed. Given these, we generate randomly $n$ data points $\mathcal{D} = \{(t_i, x_i, u_i)\}_{i=1\ldots n}$ in the domain, such that $u_i$ is the optimal control value at point $x_i$ and time $t_i$. For a given value of $\lambda$, we compute a solution $l$ of problem (IOCP). We can then measure how $l$ is close to $l_0$. Note that in some simulations, the control is corrupted with noise.

## 5.2 Benchmark direct problems

### 5.2.1 Minimum exit time in dimension 1

For this problem we take

$$X = U = B_1, \ X_T = \partial X, \ l_0 = 1, \ f = u.$$

The optimal law for this problem is $u = \text{sign}(x)$ and the value function is $v_0(x) = 1 - |x|$.

### 5.2.2 Minimum exit time in dimension 2

For this problem we take

$$X = U = B_2, \ X_T = \partial X, \ l_0 = 1, \ f = u.$$

The optimal law for this problem is $u = \frac{x}{||x||_2}$ and the value function is $v_0(x) = 1 - ||x||_2$.

### 5.2.3 Minimum exit norm in dimension 2

For this problem we take

$$X = U = B_2, \ X_T = \partial X,$$
$$l_0 = ||x||_2^2 + ||u||_2^2, \ f = u.$$

The optimal law for this problem is $u = x$ and the value function is $v_0(x) = 1 - ||x||_2^2$.

### 5.2.4 Fixed time double integrator with quadratic cost

For this problem we take

$$X = B_2, \ U = [-r, r], \ T = 1,$$
$$l_0 = x^T \begin{pmatrix} 2 & \frac{1}{4} \\ \frac{1}{4} & 1 \end{pmatrix} x + u^2,$$
$$f = \begin{pmatrix} 0 & 1 \\ 0 & 0 \end{pmatrix} x + \begin{pmatrix} 0 \\ 1 \end{pmatrix} u$$

for a big enough value of $r$. This is an LQR problem, the optimal control is of the form $u(t) = -K(t)x(t)$ where $K(t)$ is obtained by solving the corresponding Riccati differential equation.

## 5.3 Numerical results

### 5.3.1 Illustration on a one dimensonal problem

We consider the one dimensional minimum exit time problem in 5.2.1. The results are presented in Figure 1. We compare the output of (IOCP) with ($\lambda = 1$) and without ($\lambda = 0$) regularization. The main comments are as follows.

- Given any symmetric differentiable concave function $v$ vanishing on $\{-1, 1\}$, the pair $(l = |v'|, v)$ solves problem (IOCP) with $\epsilon = 0$.

- Given any polynomial $v$ vanishing on $\{-1, 1\}$, and any positive polynomial $p$ on $[0, T] \times X \times U$, the pair $(l = (1 - u^2)p - uv', v)$ solves problem (IOCP) with $\epsilon = 0$. Note that this solution only captures the fact that $|u| = 1$.

- Any convex combination of solutions of the types mentioned above also solve problem (IOCP). It is therefore very hard, in the absence of regularization ($\lambda = 0$), to recover the true Lagrangian.

- The sparsity inducing effect of $\ell_1$-norm regularization allows to recover the true Lagrangian ($\lambda = 1$).

- The value function associated to this Lagrangian is not smooth around the origin and therefore hard to approximate, hence the value of the error is high.
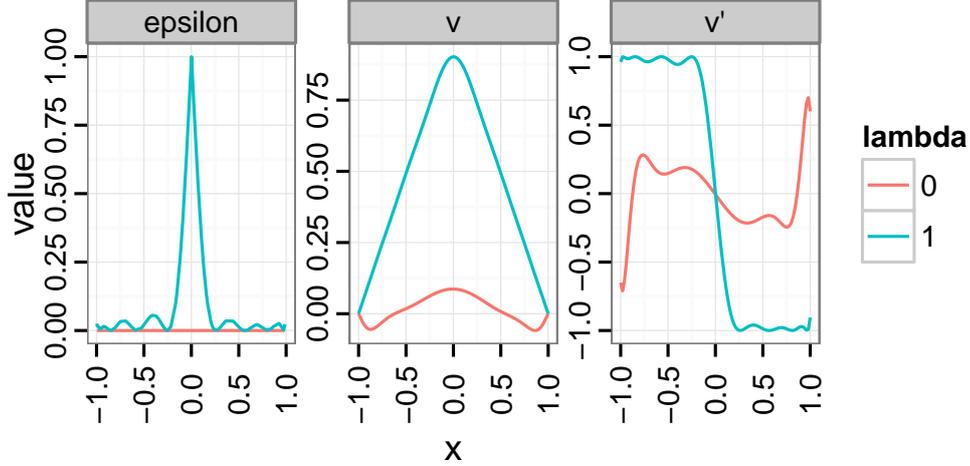
Figure 1: Solution for the one dimensional minimum exit time problem, effect of the regularization parameter $\lambda$. The first column is the distribution of the error $\epsilon$, the second is a representation of the value function $v$ and the third column is a representation of its derivative for solutions of problem (IOCP) with and without regularization. We take 100 points on the segment. Lagrangian $l$ and value function $v$ are both polynomials of degree 16.

### 5.3.2 Lagrangian estimation in dimension 2

We consider the following settings

    A  Problem 5.2.2 with $\{x_i\}$ sampled on $B_2$.

    A'  Problem 5.2.2 with $\{x_i\}$ sampled on $B_2 \setminus \frac{1}{2}B_2$.

    B  Problem 5.2.3 with $\{x_i\}$ sampled on $B_2$.

    C  Problem 5.2.4 with $\{x_i\}$ sampled uniformly on the unit $\ell_\infty$ ball.

In all cases, $l$ has degree 4 and $v$ has degree 10.

**Estimation error**   Since the inverse problem is positively homogeneous, our objective is to recover a Lagrangian $l_0$, up to a positive multiplicative factor. Since we use polynomials, the Lagrangians we estimate can be represented as vectors in the monomial basis. We use the following metric to account for the estimation error:

$$\inf_\alpha \frac{||l_0 - \alpha l||_2}{||l_0||_2} = \left(1 - \frac{\langle l_0, l\rangle^2}{||l_0||_2^2 ||l||_2^2}\right)^{\frac{1}{2}} \tag{10}$$

where the scalar product of polynomials is defined as the usual scalar product of their vectors of coefficients in the monomial basis.
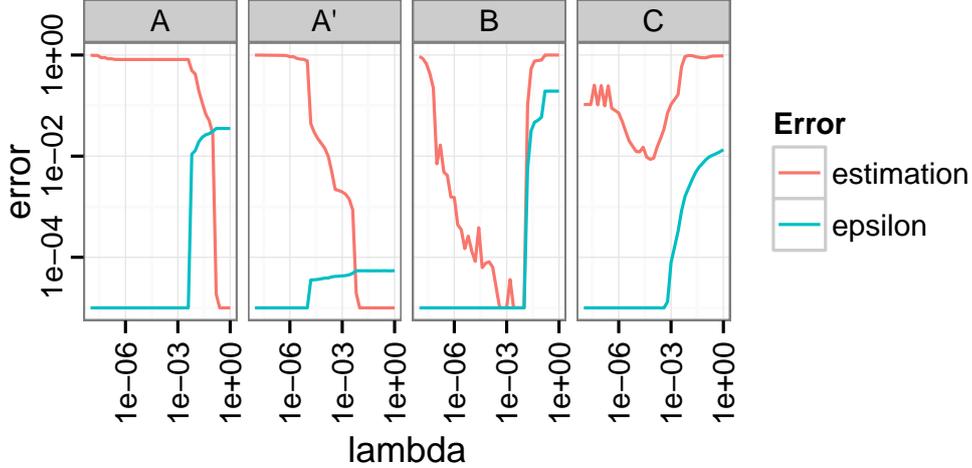
Figure 2: Deterministic setting in dimension 2, error versus variation of regularization parameter $\lambda$. A: minimum exit time. A': minimum exit time sampling far from the origin. B: minimum exit norm. C: double integrator. Estimation error is the metric defined in (10). Epsilon error is the value of the term $\epsilon$ in program (IOCP). For A, A' and B, we take 20 random points and corresponding control value. For C, we take 50 random points and time with corresponding control. For all problems, Lagrangian $l$ is a polynomial of degree 4 and value function $v$ is a polynomial of degree 10.

**Deterministic setting**  The results for the four problems are presented in Figure 2. For all problems, $l$ is of degree 4. Therefore, for problems A, A' and B, $l$ is represented by a 70-dimensional vector and for problem $C$, it is a 35-dimensional vector. When the estimation error is close to 1, we estimate a Lagrangian $l$ that is orthogonal to $l_0$ in the monomial basis. For all problems we are able to recover the true Lagrangian with good accuracy for some value of the regularization parameter $\lambda$. In the absence of regularization, we do not recover the true Lagrangian at all. This highlights the important role of $\ell_1$ regularization which allows to bias the estimation toward *sparse polynomials*. When the estimation error is minimal, the value of $\epsilon$ is reasonably low, depending on how the value function can be approximated by a polynomial. For example, A' shows lower $\epsilon$ value because we avoid sampling database points close to the nondifferentiable point of the true value function. In example C, the value function is not a polynomial and therefore harder to approximate. The estimation accuracy is still very reasonable.

**Stochastic setting**  The results for the four problems are presented in Figure 3. The setting is similar to the deterministic case of the previous paragraph, except that we add uniform noise (of maximum magnitude $10^{-1}$) to the control input. Therefore, the problem is harder than the one presented in the previous paragraph. Several samples and noise realizations are considered to highlight the global trends. These simulations show that despite the stochastic corruption of the input control, we are still able to recover the true Lagrangian with reasonable accuracy. As one could expect, increasing the number of datapoints allows to recover the true Lagrangian with a better accuracy.
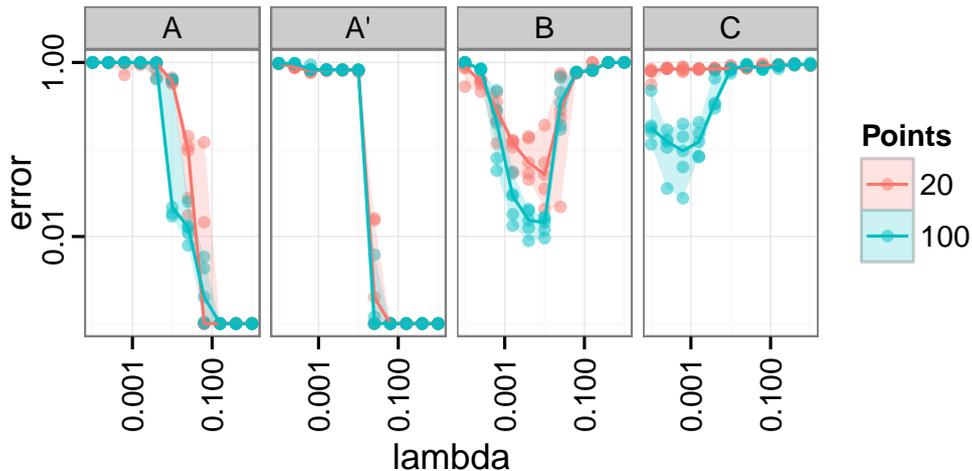
Figure 3: Stochastic setting in dimension 2, estimation error versus variation of regularization parameter $\lambda$. A: minimum exit time. A': minimum exit time sampling far from the origin. B: minimum exit norm. C: double integrator. Estimation error is the metric defined in (10). We vary the number of random points. The control is corrupted by uniform noise and we repeat the experiment five times for each value of $\lambda$. For all problems, Lagrangian $l$ is a polynomial of degree 4 and value function $v$ is a polynomial of degree 10.

# 6    CONCLUSIONS AND FUTURE WORKS

We have presented how Hamilton-Jacobi-Bellman sufficient condition can be used to analyse the inverse problem of optimal control and proposed a practical method based on polynomial optimization and linear matrix inequality hierarchies to provide a candidate solution to this problem. Numerical results suggest that the method is able to estimate accurately Lagrangians comming from various optimal control problems.

For the specific examples proposed, the optimality conditions allow to highlight many sources of ill-posedness for the inverse problem. In addition to a relaxation of the optimality conditions, we added a constraint and a penalization to circumvent ill-posedness. Numerical simulations support the idea that these are essential to estimate a Lagrangian accuartely. We do not rely on strong bias toward specific candidate Lagrangians and are able to perform accurate estimation in many different settings using the same regularization technique. A natural question that arises here is to find necessary conditions under which this is possible.

The motivation for the use of Hamilton-Jacobi-Bellman optimality condition is the guarantees it provides when one has access to complete noiseless trajectories. However, practical experimental settings require to consider the effects of discretization and additive noise. Along these lines, consistency and asymptotic properties of the proposed estimation procedure with respect to random discretization and noise are natural questions.

Finally, it is necessary to carry out further experiments on real world datasets in order to determine if the proposed method works on practical inverse problems, our primary target being those coming from humanoid robotics [4, 22].

14

## ACKNOWLEDGMENTS

## References

[1] P. Abbeel, A. Y. Ng. Apprenticeship learning via inverse reinforcement learning. In *Proceedings of the International Conference on Machine Learning*, ACM, 2004.

[2] A. Ajami, J.-P. Gauthier, T. Maillot, U. Serres. How humans fly. *ESAIM: Control, Optimisation and Calculus of Variations*, 19(04):1030–1054, 2013.

[3] B. D. O. Anderson, J. B. Moore. *Linear optimal control*. Prentice-Hall, 1971.

[4] G. Arechavaleta, J.-P. Laumond, H. Hicheur, A. Berthoz. An optimality principle governing human walking. *IEEE Transactions on Robotics*, 24(1):5–14, 2008.

[5] M. A. Babyak. What you see may not be what you get: a brief, nontechnical introduction to overfitting in regression-type models. *Psychosomatic medicine*, 66(3):411–421, 2004.

[6] M. Bardi, I. Capuzzo-Dolcetta. *Optimal control and viscosity solutions of Hamilton-Jacobi-Bellman equations*. Springer, 2008.

[7] J. Casti. On the general inverse problem of optimal control theory. *Journal of Optimization Theory and Applications*, 32(4):491–497, 1980.

[8] F.C. Chittaro, F. Jean, P. Mason. On inverse optimal control problems of human locomotion: Stability and robustness of the minimizers. *Journal of Mathematical Sciences*, 195(3):269–287, 2013.

[9] G. Darboux. *Leçons sur la théorie générale des surfaces et les applications géométriques du calcul infinitésimal*, volume 3, §§604-605, Gauthier-Villars, 1894.

[10] J. Douglas. Solution of the inverse problem of the calculus of variations. *Transactions of the American Mathematical Society*, 50(1):71–128, 1941.

[11] P. Falb, M. Athans. *Optimal Control. An Introduction to the Theory and Its Applications*. McGraw-Hill, 1966.

[12] R. A. Freeman, P. V. Kokotovic. Inverse optimality in robust stabilization. *SIAM Journal on Control and Optimization*, 34(4):1365–1391, 1996.

[13] T. Fujii, M. Narazaki. A complete optimality condition in the inverse problem of optimal control. *SIAM Journal on Control and Optimization*, 22(2):327–341, 1984.

[14] K. Hatz, J. P. Schlöder, H. G. Bock. Estimating parameters in optimal control problems. *SIAM Journal on Scientific Computing*, 34(3):A1707–A1728, 2012.

[15] A. Jameson, E. Kreindler. Inverse problem of linear optimal control. *SIAM Journal on Control*, 11(1):1–19, 1973.

[16] R. E. Kalman. When is a linear control system optimal? *Journal of Basic Engineering*, 86(1):51–60, 1964.

[17] A. Keshavarz, Y. Wang, S. Boyd. Imputing a convex objective function. *Proceedings of the IEEE International Symposium on Intelligent Control*, pp. 613–619, 2011.

[18] J. B. Lasserre. Global optimization with polynomials and the problem of moments. *SIAM Journal on Optimization*, 11(3):796–817, 2001.

[19] J. B. Lasserre. Inverse polynomial optimization. *Mathematics of Operations Research*, 38(3):418–436, 2013.

[20] J. B. Lasserre, D. Henrion, C. Prieur, E. Trélat. Nonlinear optimal control via occupation measures and LMI relaxations. *SIAM Journal on Control and Optimization*, 47(4):1643–1666, 2008.

[21] J. Löfberg. Pre-and post-processing sum-of-squares programs in practice. *IEEE Transactions on Automatic Control*, 54(5):1007–1011, 2009.

[22] K. Mombaur, A. Truong, J. P. Laumond. From human to humanoid locomotion–an inverse optimal control approach. *Autonomous robots*, 28(3):369–383, 2010.

[23] P. Moylan, B. D. O. Anderson. Nonlinear regulator theory and an inverse optimal control problem. *IEEE Transactions on Automatic Control*, 18(5):460–465, 1973.

[24] F. Nori, R. Frezza. Linear optimal control problems and quadratic cost functions estimation. *Proceedings of the Mediterranean Conference on Control and Automation*, volume 4, 2004.

[25] M. Putinar. Positive polynomials on compact semi-algebraic sets. *Indiana University Mathematics Journal*, 42(3):969–984, 1993.

[26] A. S. Puydupin-Jamin, M. Johnson, T. Bretl. A convex approach to inverse optimal control and its application to modeling human locomotion. *Proceedings of the IEEE International Conference on Robotics and Automation*, pp. 531–536, 2012.

[27] N. D. Ratliff, J. A. Bagnell, M. A. Zinkevich. Maximum margin planning. *Proceedings of the ACM International Conference on Machine Learning*, pp. 729–736, 2006.

[28] D. J. Saunders. Thirty years of the inverse problem in the calculus of variations. *Reports on Mathematical Physics*, 66(1):43 – 53, 2010.

[29] F. Thau. On the inverse optimum control problem for a class of nonlinear autonomous systems. *IEEE Transactions on Automatic Control*, 12(6):674–681, 1967.

[30] E. Tonti. Variational formulation for every nonlinear problem. *International Journal of Engineering Science*, 22(11):1343–1371, 1984.

[31] L. Vandenberghe, S. Boyd. Semidefinite programming. *SIAM Review*, 38(1):49–95, 1996.

[32] V. N. Vapnik. An overview of statistical learning theory. *IEEE Transactions on Neural Networks*, 10(5):988–999, 1999.