Explicit Solutions to the Bellman Equation for Semilinear Systems

David Ohlin[1], Richard Pates[1] and Murat Arcak[2]

***Abstract*— This paper presents sufficient conditions for optimal control of systems with dynamics given by a linear operator, in order to obtain an explicit solution to the Bellman equation that can be calculated in a distributed fashion. Further, the class of Linearly Solvable MDP is reformulated as a continuous-state optimal control problem. It is shown that this class naturally satisfies the conditions for explicit solution of the Bellman equation, motivating the extension of previous results to semilinear dynamics to account for input nonlinearities. The applicability of the given conditions is illustrated in scenarios with linear and quadratic cost, corresponding to the Stochastic Shortest Path and Linear-Quadratic Regulator problems.**

I. Introduction

For what classes of optimal control problems can we expect the existence of efficient algorithms? Leading lights of the field are the A^* algorithm [1] for Stochastic Shortest Path problems (SSP), the Linear-Quadratic Regulator problem (LQR) [2] and Linearly solvable Markov Decision Processes (LDP) [4]. The objective of this paper is to formalize the connection between these seemingly disparate instances, giving a set of sufficient conditions to guarantee that the Bellman equation can be decoupled and solved explicitly.

In [4], the class of LDP is identified as a subset of MDP with cost based on the Kullback-Liebler distance between an underlying autonomous transition function and the controlled dynamics. The key feature of an explicit linear equation for the solution to the Bellman equation is leveraged to find the optimal control. In order to illustrate the general applicability of our conditions, we reformulate the class of LDP as a continuous-state problem and show that the resulting modified cost and input constraints together with the dynamics share properties of SSP and LQR.

II. Problem setup

We consider the infinite-horizon optimal control problem

$$\begin{aligned} \text{Minimize} \quad & \sum_{t=0}^{\infty} \|x\|_{h(P)} \\ \text{subject to} \quad & x(t+1) = \mathcal{A}_P x(t), \quad x(0) = x_0 \\ & P \in \mathcal{P}, \quad x(t) \in \mathcal{X} \end{aligned} \tag{1}$$

This work is partially funded by the Wallenberg AI, Autonomous Systems and Software Program (WASP) funded by the Knut and Alice Wallenberg Foundation.

[1] Department of Automatic Control, Lund University, Box 118, SE-221 00 LUND, Sweden. The authors are with the ELLIIT Strategic Research Area at Lund University. (e-mail: {david.ohlin, richard.pates}@control.lth.se).

[2] Department of Electrical Engineering and Computer Sciences, UC Berkeley, 569 Cory Hall, Berkeley, CA 94720, USA (e-mail: arcak@berkeley.edu).

where $\mathcal{X}$ is a proper cone and $\mathcal{A}_P$ is a bounded linear operator parameterized by P. Let $\mathcal{P}$ be a closed subset of some Hilbert space. Define the norm as

$$\|x\|_w := \langle w, x \rangle \quad \text{for } x \in \mathcal{X},\ w \in \text{int}(\mathcal{X}^*) \tag{2}$$

where $\mathcal{X}^*$ is the dual cone of $\mathcal{X}$. Further, let the cone $\mathcal{X}$ be invariant under the dynamics, i.e. $\mathcal{A}_P x \in \mathcal{X}$ for all $x \in \mathcal{X}$. The codomain of the weighting function $h(P)$ is restricted, $h : \mathcal{P} \to \text{int}(\mathcal{X}^*)$, guaranteeing a positive immediate cost. Define $\{\mathcal{P}_i\}_{i=1}^n$ as a partition of the constraint set,

$$\mathcal{P} = \mathcal{P}_1 \times \cdots \times \mathcal{P}_n. \tag{3}$$

The existence of such a partition with specific structure in relation to the cost and dynamics of (1), such that both are additively separable, is key to enable decomposition of the Bellman equation. This condition is formalized in the following assumption:

Assumption 1: Let $\{\mathcal{P}_i\}_{i=1}^n$ be a partition as in (3) with $P_i \in \mathcal{P}_i$ such that

$$\mathcal{A}_P = \sum_{i=1}^{n} \mathcal{A}_{P_i} \ \text{ and } \ h(P) = \sum_{i=1}^{n} h_i(P_i), \tag{4}$$

The functions $h_i(P_i)$ are $\mathcal{X}^*$-convex and $\mathcal{A}_{P_i} x$ are $\mathcal{X}$-convex in the parameter P_i for $x \in \mathcal{X}$. Further, let $\langle h_i(P_i), x \rangle$ be coercive with regard to P_i for any $x \in \mathcal{X}$.

A. The Linear-Quadratic Regulator

Consider the linear dynamics $y(t+1) = (A + BK)y(t)$ with $y \in \mathbb{R}^n$, for some matrices A, B, i.e., assuming static feedback $u = Ky$ with gain $K \in \mathbb{R}^{n \times m}$ chosen freely. We illustrate here only the case of $m = n$ to simplify the exposition. In order to obtain a problem on the form (1), let $x(t) = y(t)y(t)^\top$. This gives the expression $\mathcal{A}_K x = (A + BK)x(A + BK))^\top$ for the dynamics with the domain $\mathcal{X}$ given by the semidefinite cone. Restricting the dynamics matrix $A + BK$ to be invertible yields invariance of $\mathcal{X}$ under $\mathcal{A}_K$. The immediate cost for symmetric weight matrices Q and R is

$$\begin{aligned} \|x\|_{Q+K^\top RK} &= \langle Q + K^\top RK, x \rangle \\ &= \text{tr}((Q + K^\top RK)^\top x) \\ &= y^\top (Q + K^\top RK) y \end{aligned}$$

using the Frobenius inner product on $\mathcal{X}$. This is equivalent to the typical quadratic cost for $Q + K^\top RK \succ 0$. Finding the K that solves (1) is then equivalent to solving the LQR problem.

B. Stochastic Shortest Path

As shown in [3], SSP can be modeled by the optimal control problem

$$\begin{aligned}
\text{Minimize} \quad & \sum_{t=0}^{\infty} \left[s^\top x(t) + r^\top u(t) \right] \text{ over } \{u(t)\}_{t=0}^{\infty} \\
\text{subject to} \quad & x(t+1) = Ax(t) + Bu(t) \\
& u(t) \geq 0, \quad x(0) = x_0 \in \mathbb{R}_+^n \\
& \mathbf{1}^\top u_1(t) \quad \leq \quad E_1^\top x(t) \\
& \quad \vdots \qquad\qquad \vdots \\
& \mathbf{1}^\top u_n(t) \quad \leq \quad E_n^\top x(t)
\end{aligned} \tag{5}$$

Here, the input signal $u \in \mathbb{R}^m$ is partitioned into n subvectors u_i, each containing m_i elements, so that $m = \sum_{i=1}^n m_i$. This is a special case of (1). Given static feedback $u = Kx$, we let $\mathcal{A}_K x = (A+BK)x$ and invariance of $\mathcal{X} = \mathbb{R}_+^n$ under the dynamics corresponds to the condition $A + BK \geq 0$. This holds given an appropriate choice of the matrix E, governing the input constraints in (5). The immediate cost is expressed as the weighted 1-norm

$$\begin{aligned}
\|x\|_{s+K^\top r} &= (s + K^\top r)^\top |x| \\
&= (s + K^\top r)^\top x \text{ for } x \in \mathcal{X}.
\end{aligned}$$

The inclusion $s + K^\top r \in \text{int}(\mathcal{X}^*)$ can be fulfilled by requiring $s > 0$, $r \geq 0$, guaranteeing observability of the state.

III. Main Result

The following theorem gives a sufficient condition for the decomposition and explicit solution of the Bellman equation for (1). Additionally, a program, which is demonstrated to be convex in the cases of interest below, gives the optimal cost function.

Theorem 1: Let Assumption 1 hold. Then, the following statements are equivalent:

(i) The problem (1) has a finite value for all x_0.

(ii) There exists $\lambda \in \mathcal{X}^*$ satisfying the equation

$$\lambda = \sum_{i=1}^{n} \min_{P_i \in \mathcal{P}_i} h_i(P_i) + \mathcal{A}_{P_i}^* \lambda \tag{6}$$

(iii) The value of the program

$$\begin{aligned}
\text{Maximize} \quad & \|x_0\|_\lambda \text{ over } \lambda \in \mathcal{X}^* \\
\text{subject to} \quad & \sum_{i=1}^{n} \left(\min_{P_i \in \mathcal{P}_i} h_i(P_i) + \mathcal{A}_{P_i}^* \lambda \right) - \lambda \in \mathcal{X}^*
\end{aligned}$$

is bounded for $x_0 \in \mathcal{X}$.

Further, the maximum value of the program in (iii) and the optimal value of (1) is given by $\|x_0\|_\lambda$, with λ solving (6). The optimally controlled dynamics are given by

$$P_i = \operatorname*{argmin}_{P_i \in \mathcal{P}_i} h_i(P_i) + \mathcal{A}_{P_i}^* \lambda. \tag{7}$$

IV. Linearly Solvable MDP

The LDP framework of [4] can be reformulated as a continuous-state problem on the form (1), with dynamics

$$x(t+1) = Px(t).$$

Consider the objective function

$$\|x\|_{h(P)} = s^\top x + \text{diag}(P^\top \log(P \oslash \overline{P}))^\top x + \pi^\top x \tag{8}$$

where

$$\pi = (I - P^\top)\mathbf{1} \odot \log((I - P^\top)\mathbf{1} \oslash \overline{p}_g). \tag{9}$$

Here, $\overline{p}_g$ is the constant vector of unmodified transition rates from non-goal states to the goal state in the original system, and $\overline{P}$ are a constant given matrix defining the autonomous dynamics. This construction makes the immediate cost equal to that of [4], with the final term $\pi^\top x$ representing the cost incurred by transitions to the goal states in the original model. We can simplify the constraint set, as the cost associated with control in states with $x_i = 0$ naturally vanishes in the formulation (8).

$$\mathcal{P} = \{P : P^\top \mathbf{1} \leq \mathbf{1},\ P \geq 0\} \tag{10}$$

with equality $p_i^\top \mathbf{1} = 1$ for row i if the ith element of $\overline{p}_g$ is zero. In order to show applicability of Theorem 1 we first introduce a proposition concerning the immediate cost:

Proposition 4.1: The immediate cost (8) is a valid norm for $s > 0$.

Proof: The term $(h(P) - s)^\top x$ is equal to the KL-divergence between two distributions (see [4]), the controlled and autonomous dynamics, weighted by the state vector, and is thus nonnegative. It follows then from positivity of the dynamics that $h(P)^\top x \geq 0$ with equality only in the case $x = 0$ as a consequence of $s > 0$. ∎

The above reformulation together with the application of Theorem 1 shows the close connection between traditional results for optimal control of linear systems [2] and the linear cost achieved for LDP in [4]. Any sparsity in the autonomous dynamics $\overline{P}$ is preserved in the optimal solution, similar to the case of linear cost and dynamics [3].

References

[1] Peter E. Hart, Nils J. Nilsson, and Bertram Raphael. "A Formal Basis for the Heuristic Determination of Minimum Cost Paths". In: *IEEE Transactions on Systems Science and Cybernetics* 4.2 (1968), pp. 100–107.

[2] Rudolf E. Kálmán. "Contributions to the Theory of Optimal Control". In: *Bol. Soc. Mat. Mexicana* 5 (1960), pp. 102–119.

[3] David Ohlin, Anders Rantzer, and Emma Tegling. *Heuristic Search for Linear Positive Systems*. 2024. arXiv: `2410.17220 [math.OC]`.

[4] Emanuel Todorov. "Linearly-solvable Markov decision problems". In: *Advances in Neural Information Processing Systems*. Ed. by B. Schölkopf, J. Platt, and T. Hoffman. Vol. 19. MIT Press, 2006.