Distributed dynamic programming for discrete-time stochastic control, and idempotent algorithms☆

William M. McEneaney *

Department of Mechanical and Aerospace Engineering, MC 0411, University of California, San Diego, 9500 Gilman Drive, La Jolla, CA 92093-0411, USA

ARTICLE INFO

ABSTRACT

Idempotent methods have been found to be extremely fast for the solution of dynamic programming equations associated with deterministic control problems. The original methods exploited the idempotent (e.g., max-plus) linearity of the associated semigroup operator. It is now known that curse-of-dimensionality-free idempotent methods do not require this linearity. Instead, it is sufficient that certain solution forms are retained through application of the associated semigroup operator. Here, we see that idempotent methods may be used to solve some classes of stochastic control problems. The key is the use of the idempotent distributive property. We demonstrate this approach for a class of nonlinear, discrete-time stochastic control problems.

© 2010 Elsevier Ltd. All rights reserved.

1. Introduction

It is now well known that many classes of deterministic control problems may be solved by max-plus or min-plus (more generally, idempotent) numerical methods. Here, max-plus methods are appropriate for problems with maximizing controllers and vice versa. These methods include max-plus basis-expansion approaches (Akian, Gaubert, & Lakhoua, 2004, 2008; Fleming & McEneaney, 2000; McEneaney, 2004, 2006; McEneaney & Dower, 2003), as well as the more recently developed curse-of-dimensionality-free methods (McEneaney, 2006, 2007). However, previously, stochastic control problems have eluded idempotent methods.

In a recent application, a min-plus approach was useful for a problem in sensing control (McEneaney, Oran, & Cavender, 2008; Oran & McEneaney, 2010). In that example, the state process was an observation-conditioned probability vector, where the stochasticity of the state process was due to the stochastic observation inputs. This led to the realization that idempotent methods were indeed applicable to stochastic control problems. The key tool that had been missing previously was simply the idempotent distributive property.

It was also necessary to realize that idempotent linearity of the associated semigroup was not required for applicability of curse-of-dimensionality-free types of idempotent methods. Instead, it is sufficient that certain solution forms are retained through application of the semigroup operator, i.e., the dynamic programming principle operator. We will see that, under certain conditions, pointwise minima of affine and quadratic forms will pass through this operator, when we have a minimizing control problem. As the operator contains an expectation component, this will require application of the idempotent distributive property. In the case of finite sums and products, this property looks like our standard-algebra distributive property; in the infinitesimal case, it is familiar to control theorists through notions of strategies, non-anticipative mappings, and/or progressively measurable controls. Using this technology, the value function will be propagated backwards with a representation as a pointwise minimum of quadratic or affine forms.

This approach will have some similarities to the idempotent curse-of-dimensionality-free methods developed for deterministic control problems (McEneaney, 2006, 2007; McEneaney, Deshpande, & Gaubert, 2008; McEneaney, 2009b). However, the need to use the idempotent distributive property makes this approach very different from those.

We will use a very limited class of control problems to introduce this new approach. We consider discrete-time, finite time-horizon control problems. The dynamics will contain a possibly continuum-valued stochastic input. They will also contain both a possibly continuum-valued control, and a control taking values in a finite set. The running cost will also depend on these inputs. We may think of the dynamics and the running cost as being indexed by the values of the finite control input. In the actual

computational algorithms which will be obtained, each indexed dynamics will be linear, and each indexed running cost will be either quadratic or affine. A similar approach was taken with the curse-of-dimensionality-free methods for deterministic control. In that case, it was shown that any semiconvex (respectively, convex) Hamiltonian could be arbitrarily well approximated as a pointwise maximum of quadratic (respectively, affine) forms, and so this class is not so restrictive.

Note that the purpose of this paper is the announcement of this newly discovered class of algorithms. The general theory is indicated in an abstract, albeit discrete-time, setting, resulting in what we refer to here as the idempotent distributed dynamic programming principle (IDDPP). The basic algorithms will be described for two relatively simple classes of problems, where the value function will be given in terms of pointwise minima of quadratic and affine forms. Instantiation of these basic algorithms in computationally efficient forms requires approximation steps. A study of such approximations will not be included here, but the reader may refer to McEneaney, Oran et al. (2008) and Oran and McEneaney (2010) for some initial approximation methods relevant to this class of algorithms.

A few words about the expected utility of the approach are appropriate here. The paper considers a discrete-time, finite time-horizon stochastic control problem with continuum-valued state process. The basic theory is developed for continuum-valued control and noise processes. However, at the ends of Sections 5 and 6 (on quadratic and affine forms), versions of the method are given in the case of finite control and disturbance spaces. In fact, these are the cases which are most readily amenable to computation using this approach. The well-known curse of dimensionality, which is a serious impediment to grid-based methods, is due to the discretization over the state space. Here, this is avoided in the base form of the algorithm, but there is a large penalty to be paid in terms of extremely rapid growth of the solution complexity as one back-propagates. Without the pruning techniques for attenuation of this complexity growth, the approach would not be feasible for serious problems. This is similar to the case with the curse-of-dimensionality-free method of McEneaney (2006, 2007). We discuss complexity reduction for our new approach in McEneaney (2009a), Oran and McEneaney (2010), and will pursue this further in subsequent efforts. Nonetheless, it seems that a general rule of thumb will be that this approach is particularly useful for high-dimensionality/low-complexity problems. Here, the term high dimensionality refers to the class of problems where the state-space dimension is beyond that to which grid-based methods can feasibly be applied. On the other hand, problems where the Hamiltonian might be well modeled by pointwise minima of a small set of linear-quadratic forms tend to be of low complexity. More specifically, however, in the sense needed here, a problem is of low complexity if the solution is well approximated by a reasonably small number of simple forms such as affine or quadratic forms, where minima of such forms are inherited through the IDDPP operator to appear below. The simple example included for illustrative purposes falls into the class of low-complexity problems.

The layout of the paper is as follows. In Section 2, a finite time-horizon, discrete-time stochastic control problem is defined, and the standard dynamic programming principle (DPP) is given in this context. In Section 3, a continuum version of the min-plus distributive property is proved. This is the result that will allow us to obtain the main result. This main result is the IDDPP, which is obtained for this class of problems in Section 4. In Sections 5 and 6, the general IDDPP is reduced to computationally amenable forms for specific problem formulations. In Section 7, a simple example is included.

2. Problem definition and dynamic program

We begin by defining the specific class of problems which will be addressed here. In this initial foray into the domain of idempotent methods for stochastic control problems, we restrict ourselves to a rather narrow problem class. We will consider only discrete-time, finite time-horizon problems. Let the dynamics take the form

$$\xi_{t+1} = f(t, \xi_t, u_t, \mu_t, w_t) \tag{1}$$
$$\xi_s = x \in \mathbb{R}^n, \tag{2}$$

where ξ_t is the state at time t, s is the initial time, and f is measurable, with more assumptions on it to follow. The u_t and μ_t will be control inputs, and we find it helpful to differentiate these into the continuum-valued, u_t, and the discrete-valued, μ_t, components (see McEneaney, 2006, 2007, for motivation). The w_t will be the stochastic disturbance inputs. The terminal time is T, and we specifically write the time period as $]s, T[\doteq \{s, s+1, \ldots T\}$. More generally, we will use the notation $]a, b[\doteq \{a, a+1, \ldots b\}$ for integers $a < b$, throughout.

We will use the term probability triple to denote a triple consisting of a (topological) sample space, the collection of Borel sets on the sample space, and a probability measure on the Borel sets; the generic notation will be $(\Omega, \mathcal{B}, P)$, where Ω denotes the sample space, $\mathcal{B}$ denotes the collection of Borel sets, and P denotes the probability measure, with various subscripts and superscripts distinguishing such triples. We suppose that the range of each random w_t is separable metric space (W, d_w) (where, for simplicity, we skip the additional generality of letting the range space depend on time). We also suppose that each w_t is a random variable with respect to the time-indexed probability triple $(\Omega_t^w, \mathcal{B}_t^w, P_t^w)$; that is, $w_t : \Omega_t^w \to W$ is measurable with respect to $\mathcal{B}_t^w$. For each $t \in]s, T[$, let $\overline{\Omega}_t^w \doteq \prod_{r-s}^t \Omega_t^w$, where the product notation, indicating outer product here, will be used both for standard product and outer product; the meaning will be clear from the context. Let $\overline{\mathcal{B}}_t^w$ be the Borel sets on $\overline{\Omega}_t^w$ (which we note are generated by the rectangular outer products of the $\mathcal{B}_r^w$; see Folland (1984)). Let $\overline{P}_t^w$ denote the resulting probability measure on $\overline{\mathcal{B}}_t^w$.

We suppose that each $u_t : \overline{\Omega}_t^w \to U \subseteq \mathbb{R}^k$ and each $\mu_t : \overline{\Omega}_t^w \to \mathcal{M} =]1, 2, \ldots, M[$ is measurable with respect to Borel sets $\overline{\mathcal{B}}_t^w$. We suppose that $d_m(\cdot, \cdot)$ is a metric on $\mathcal{M}$. More particularly, we let the sets of allowable controls at time t be denoted by

$$\mathcal{U}_t \doteq \{u_t : \overline{\Omega}_t^w \to U \mid \overline{\mathcal{B}}_{t-1}^w \times \{\Omega_t^w, \emptyset\} \text{ measurable}\},$$
$$\mathcal{M}_t' \doteq \{\mu_t : \overline{\Omega}_t^w \to \mathcal{M} \mid \overline{\mathcal{B}}_{t-1}^w \times \{\Omega_t^w, \emptyset\} \text{ measurable}\},$$

where the special case $t = s$ is clear (one drops the $\overline{\mathcal{B}}_{t-1}^w$ term). Let the allowable sets of control processes be denoted by $\mathcal{U}_{]s,T-1[} \doteq \prod_{t=s}^{T-1} \mathcal{U}_t$ and $\mathcal{M}_{]s,T-1[}' \doteq \prod_{t=s}^{T-1} \mathcal{M}_t'$. Let generic elements of U, $\mathcal{M}$ and W be denoted by u, m and w, respectively.

The payoff (to be minimized) will be

$$J(s, x, u_\cdot, \mu_\cdot) \doteq \mathbf{E}\left\{\left[\sum_{t=s}^{T-1} l(t, \xi_t, u_t, \mu_t, w_t)\right] + \Psi(\xi_T)\right\} \tag{3}$$

where

$$\Psi(x) \doteq \inf_{z_T \in Z_T} \{g_T(x, z_T)\}, \tag{4}$$

where $\xi_s = x$ is the initial state, l and the g_T are measurable, and (Z_T, d_{z_T}) is a separable metric space. Here, Ψ denotes the terminal cost, which we are assuming takes the given form (as an infimum of the g_T functionals, indexed by the z_T). It is important to note that the μ_t process may serve as a means of generating