Reinforcement Learning and Optimal Control

by

Dimitri P. Bertsekas

2019

Chapter 6
Aggregation

SELECTED SECTIONS

WWW site for book information and orders
http://www.athenasc.com

Athena Scientific, Belmont, Massachusetts
6

Aggregation

<table>
<thead>
<tr>
<th>Contents</th>
</tr>
</thead>
<tbody>
<tr>
<td>6.1. Aggregation with Representative States . . . . . . . . p. 308</td>
</tr>
<tr>
<td>6.1.1. Continuous Control Space Discretization . . . . . p. 314</td>
</tr>
<tr>
<td>6.1.2. Continuous State Space - POMDP Discretization . . . . . p. 315</td>
</tr>
<tr>
<td>6.2. Aggregation with Representative Features . . . . . . . . p. 317</td>
</tr>
<tr>
<td>6.2.1. Hard Aggregation and Error Bounds . . . . . . . . . . p. 320</td>
</tr>
<tr>
<td>6.2.2. Aggregation Using Features . . . . . . . . . . . . . . . . p. 322</td>
</tr>
<tr>
<td>6.3. Methods for Solving the Aggregate Problem . . . . . . . . p. 328</td>
</tr>
<tr>
<td>6.3.1. Simulation-Based Policy Iteration . . . . . . . . . . p. 328</td>
</tr>
<tr>
<td>6.3.2. Simulation-Based Value Iteration and Q-Learning . . . p. 331</td>
</tr>
<tr>
<td>6.4. Feature-Based Aggregation with a Neural Network . . . . . p. 332</td>
</tr>
<tr>
<td>6.5. Biased Aggregation . . . . . . . . . . . . . . . . . . . . . p. 334</td>
</tr>
<tr>
<td>6.6. Notes and Sources . . . . . . . . . . . . . . . . . . . p. 337</td>
</tr>
<tr>
<td>6.7. Appendix: Mathematical Analysis . . . . . . . . . . . . . p. 340</td>
</tr>
</tbody>
</table>
In this chapter we consider approximation in value space using a problem approximation approach that is based on aggregation. In particular, we construct a simpler and more tractable “aggregate” problem by creating special “groups” of multiple states, which we view as “aggregate states.” We solve the aggregate problem exactly by DP, and we use its optimal cost-to-go function in a one-step or multistep lookahead approximation scheme for the original problem.

In addition to problem approximation, aggregation is related to feature-based parametric approximation. For example, it often produces a piecewise constant cost function approximation, which may be viewed as a linear feature-based parametrization, where the features are 0-1 membership functions; see Example 3.1.1. Aggregation can also be combined with other approximation schemes, to add a local correction to a cost function approximation $\tilde{J}$, which is already available, possibly through the use of a neural network.

Aggregation can be applied to both finite horizon and infinite horizon problems. In this chapter, we will focus primarily on the discounted infinite horizon problem. We will introduce aggregation in a simple intuitive form in Section 6.1, and then generalize to a more sophisticated form of aggregation in Section 6.2.

### 6.1 AGGREGATION WITH REPRESENTATIVE STATES

In this section we focus on a relatively simple form of aggregation, which involves a special subset of states, called representative. Our approach is to view these states as the states of a smaller optimal control problem, called aggregate problem, which we will formulate and solve exactly in place of the original. We will then use the optimal aggregate costs of the representative states to approximate the optimal costs of the original problem states by interpolation. Let us describe a classical example.

**Example 6.1.1 (Coarse Grid Approximation)**

Consider a discounted problem where the state space is a grid of points $i = 1, \ldots, n$ on the plane. We introduce a coarser grid that consists of a subset $A$ of the states/points, which we call representative and denote by $x$; see Fig. 6.1.1. We now wish to formulate a lower-dimensional DP problem just on the coarse grid of states. The difficulty here is that there may be positive transition probabilities $p_{xj}(u)$ from some representative states $x$ to some non-representative states $j$. To deal with this difficulty, we introduce artificial transition probabilities $\phi_{jy}$ from non-representative states $j$ to representative states $y$, which we call aggregation probabilities. In particular, a transition from representative state $x$ to a nonrepresentative state $j$, is followed by a transition from $j$ to some other representative state $y$ with probability $\phi_{jy}$; see Fig. 6.1.2.
Sec. 6.1 Aggregation with Representative States

Representative states \( x \) (coarse grid)

States \( i \) (fine grid)

Figure 6.1.1 Illustration of aggregation with representative states; cf. Example 6.1.1. A relatively small number of states are viewed as representative. We define transition probabilities between pairs of aggregate states and we also define the associated expected transition costs; cf. Eq. (6.1). These specify a smaller DP problem, called the aggregate problem, which is solved exactly. The optimal cost function \( J^* \) of the original problem is approximated by interpolation from the optimal costs of the representative states \( r_y^* \) in the aggregate problem:

\[
\tilde{J}(j) = \sum_{y \in A} \phi_{jy} r_y^*, \quad j = 1, \ldots, n, \tag{6.2}
\]

and is used in a one-step or multistep lookahead scheme.

This process involves approximation but constructs a transition mechanism for an aggregate problem whose states are just the representative ones. The transition probabilities between representative states \( x, y \) under control \( u \in U(x) \) and the corresponding expected transition costs are

\[
\hat{p}_{xy}(u) = \sum_{j=1}^{n} p_{xj}(u) \phi_{jy}, \quad \hat{g}(x, u) = \sum_{j=1}^{n} p_{xj}(u) g(x, u, j). \tag{6.1}
\]

We can solve the aggregate problem by any suitable exact DP method. Let \( r_y^* \) denote the corresponding optimal cost of representative state \( x \). We can then approximate the optimal cost function of the original problem with the interpolation formula

\[
\tilde{J}(j) = \sum_{y \in A} \phi_{jy} r_y^*, \quad j = 1, \ldots, n. \tag{6.2}
\]

This function may in turn be used in a one-step or multistep lookahead scheme for approximation in value space of the original problem.

Note that there is a lot of freedom in selecting the aggregation probabilities \( \phi_{jy} \). Intuitively, \( \phi_{jy} \) should express a measure of proximity between \( j \) and \( y \), e.g., \( \phi_{jy} \) should be relatively large when \( y \) is geometrically close to \( j \). For example, we could set \( \phi_{jy_j} = 1 \) for the representative state \( y_j \) that is “closest” to \( j \), and \( \phi_{jy_{\neq j}} = 0 \) for all other representative states \( y \neq y_j \). In this case, Eq. (6.2) yields a piecewise constant cost function approximation \( \tilde{J} \) (the constant values are the scalars \( r_y^* \) of the representative states \( y \)).
Representative States

Original State Space

Representative States

Aggregation Probabilities

φ₁y

Relate

Original States to

Representative States

Figure 6.1.2 Illustration of the use of aggregation probabilities φ₁y from non-representative states j to representative states y. A transition from a state x to a nonrepresentative state j is followed by a transition to aggregate state y with probability φ₁y. In this figure, from representative state x, there are three possible transitions, to states j₁, j₂, and j₃, according to \( p_{xj₁}(u) \), \( p_{xj₂}(u) \), and \( p_{xj₃}(u) \), and each of these states is associated with a convex combination of representative states using the aggregation probabilities. For example, the state j₁ is associated with

\[
\phi₁y₁y₁ + \phi₁y₂y₂ + \phi₁y₃y₃.
\]

We will now formalize our framework for aggregation with representative states by straightforward generalization of the preceding example; see Fig. 6.1.3.

### Aggregation Framework with Representative States

We introduce a finite subset of representative states \( \mathcal{A} \subset \{1, \ldots, n\} \), and we denote them by symbols such as \( x \) and \( y \). We construct an aggregate problem, with state space \( \mathcal{A} \), and transition probabilities and transition costs defined as follows:

(a) We relate the original system states \( j \) to representative states \( y \in \mathcal{A} \) with aggregation probabilities \( \phi_{jy} \); these are scalar “weights” satisfying \( \phi_{jy} \geq 0 \) for all \( y \in \mathcal{A} \), and \( \sum_{y \in \mathcal{A}} \phi_{jy} = 1 \).

(b) We define the transition probabilities between representative states \( x \) and \( y \) under control \( u \in U(x) \) by

\[
\hat{p}_{xy}(u) = \sum_{j=1}^{n} p_{xj}(u)\phi_{jy}.
\]

(c) We define the expected transition costs at representative states \( x \) under control \( u \in U(x) \) by
Sec. 6.1 Aggregation with Representative States

\[ \hat{g}(x,u) = \sum_{j=1}^{n} p_{xj}(u) g(x,u,j). \]  

(6.4)

The optimal costs of the representative states \( y \in A \) in the aggregate problem are denoted by \( r_y^* \), and they define approximate costs for the original problem through the interpolation formula

\[ \tilde{J}(j) = \sum_{y \in A} \phi_{jy} r_y^*, \quad j = 1, \ldots, n. \]

(6.5)

Aside from the selection of representative states, an important consideration is the choice of the aggregation probabilities. These probabilities express “similarity” or “proximity” of original to representative states (as in the case of the coarse grid Example 6.1.1), but in principle they can be arbitrary. Intuitively, \( \phi_{jy} \) may be interpreted as some measure of “strength of relation” of \( j \) to \( y \). The vectors \( \{\phi_{jy} \mid j = 1, \ldots, n\} \) may also be viewed as basis functions for a linear cost function approximation via Eq. (6.5).

**Hard Aggregation and Error Bound**

A special case of interest, called hard aggregation, is when for every state \( j \), we have \( \phi_{jy} = 0 \) for all representative states \( y \), except a single one, denoted \( y_j \), for which we have \( \phi_{jy_j} = 1 \). In this case, the one step lookahead approximation

\[ \tilde{J}(j) = \sum_{y \in A} \phi_{jy} r_y^*, \quad j = 1, \ldots, n, \]

is piecewise constant; it is constant and equal to \( r_y^* \) for all \( j \) in the set

\[ S_y = \{j \mid \phi_{jy} = 1\}, \quad y \in A, \]

called the footprint of representative state \( y \); see Fig. 6.1.4. Moreover the footprints of all the representative states are disjoint and form a partition of the state space, i.e.,

\[ \cup_{x \in A} S_x = \{1, \ldots, n\}. \]

The footprint sets can be used to define a bound for the error \((J^* - \tilde{J})\). Basically this bound indicates that the error is small if \( J^* \) varies little within each \( S_y \). In particular, it can be shown that

\[ |J^*(j) - \tilde{J}(j)| \leq \frac{\epsilon}{1 - \alpha}, \quad j = 1, \ldots, n, \]
One-step Lookahead with
\[ \hat{J}(j) = \sum_{y \in A} \phi_{jy} r^*_y \]

Aggregation Probabilities \( \phi_{jy} \)

Figure 6.1.3 Illustration of the aggregate problem in the representative states framework. The transition probabilities \( \hat{p}_{xy}(u) \) and transition costs \( \hat{g}(x, u) \) are shown in the bottom part of the figure. Once the aggregate problem is solved (exactly) for its optimal costs \( r^*_y \), we define approximate costs

\[ \tilde{J}(j) = \sum_{y \in A} \phi_{jy} r^*_y, \quad j = 1, \ldots, n, \]

which are used for one-step lookahead approximation of the original problem.

where
\[ \epsilon = \max_{y \in A} \max_{i,j \in S_y} |J^*(i) - J^*(j)| \]
is the maximum variation of \( J^* \) within the footprint sets \( S_y \). We will show this fact in the context of an error bound result for the more general aggregation framework, which will be given in the next section.

For a special hard aggregation case of interest, consider the geometrical context of Example 6.1.1. There, aggregation probabilities are often based on a nearest neighbor approximation scheme, whereby each non-representative state \( j \) takes the cost value of the “closest” representative state \( y \), i.e.,

\[ \phi_{jy_j} = 1 \quad \text{if } y_j \text{ is the closest representative state to } j. \]

Then all states \( j \) for which a given representative state \( y \) is the closest to \( j \) (the footprint of \( y \)) are assigned equal approximate cost \( \hat{J}(j) = r^*_y \).
Sec. 6.1 Aggregation with Representative States

The most straightforward way to solve the aggregate problem is to compute the aggregate problem transition probabilities \( \hat{p}_{xy}(u) \) [cf. Eq. (6.3)] and transition costs \( \hat{g}(x, u) \) [cf. Eq. (6.4)] by either an algebraic calculation or by simulation. The aggregate problem may then be solved by any one of the standard methods, such as VI, PI, or linear programming (cf. Chapter 4). This exact calculation is plausible if the number of representative states is relatively small.

Another possibility is to use a simulation-based VI or PI method. We postpone the discussion of these methods to Section 6.3, where we will consider them in the context of a more general aggregation framework.

An important observation is that if the original problem is deterministic and hard aggregation is used, the aggregate problem is also deterministic, and can be solved by shortest-path like methods. This is true for both discounted problems and for undiscounted shortest path-type problems. In the latter case, the termination state of the original problem must
Aggregation with representative states extends without difficulty to problems with a continuous state space, as long as the control space is finite. Then once the representative states and the aggregation probabilities have been defined, the corresponding aggregate problem is a finite spaces discounted problem, which can be solved with the standard methods. The only potential difficulty arises when the disturbance space is also infinite, in which case the calculation of the transition probabilities and expected stage costs of the aggregate problem must be obtained by some form of integration process.

The case where both the state and the control spaces are continuous is somewhat more complicated, because both of these spaces must be discretized using representative state-control pairs, instead of just representative states. The following example illustrates what may happen if we use representative state discretization only.

**Example 6.1.2 (Continuous Shortest Path Discretization)**

Suppose that we want to find the fastest route for a car to travel between two points A and B located at the opposite ends of a square with side 1000 meters, while avoiding some known obstacles. We assume a constant car speed of 1 meter per second and that the car can drive in any direction; cf. Fig. 6.1.5.
Let us consider discretizing the space with a square grid (a set of representative states), and restrict the directions of motion to horizontal and vertical, so that at each stage the car moves from a grid point to one of the four closest grid points. Thus in the discretized version of the problem the car travels with a sequence of horizontal and vertical moves as indicated in the right side of Fig. 6.1.5. Is it possible to approximate the fastest route arbitrarily closely with the optimal solution of the discretized problem, assuming a sufficiently fine grid?

The answer is no! To see this note that in the discretized problem the optimal travel time is 2000 secs, regardless of how fine the discretization is. On the other hand, in the continuous space/nondiscretized problem the optimal travel time can be as little as $\sqrt{2} \cdot 1000$ secs (this corresponds to the favorable case where the straight line from A to B does not meet an obstacle).

The difficulty in the preceding example is that the state space is discretized finely but the control space is not. What is needed is to introduce a fine discretization of the control space as well, through some set of “representative controls.” We can deal with this situation with a suitable form of discretized aggregate problem, which when solved provides an appropriate form of cost function approximation for use with one-step lookahead. The discretized problem is a stochastic infinite horizon problem, even if the original problem is deterministic. Further discussion of this approach is outside our scope in this book, and we refer to the sources cited at the end of the chapter. Under reasonable assumptions it is possible to show consistency, i.e., that the optimal cost function of the discretized problem converges to the optimal cost function of the original continuous spaces problem as the discretization of both the state and the control spaces becomes increasingly fine.

The type of difficulty illustrated in Example 6.1.2 does not arise if the state space is continuous but the control space is finite. In particular, this is true in partially observed finite state and control spaces Markov decision problems (POMDP), which are defined over their belief space (the space of probability distributions over their states). We briefly discuss this case next.

### 6.1.2 Continuous State Space - POMDP Discretization

Let us consider any $\alpha$-discounted DP problem, where the state space is a bounded convex subset $C$ of a Euclidean space, such as the unit simplex, but the control space $U$ is finite. We use $z$ to denote the states, to distinguish them from $x$, which we will use to denote representative states. Bellman’s equation is $J = TJ$ with the Bellman operator $T$ defined by

$$(TJ)(z) = \min_{u \in U} E \{ g(z, u, w) + \alpha J(f(z, u, w)) \}, \quad z \in C.$$ 

We introduce a set of representative states $\{x_1, \ldots, x_m\} \subset C$. We assume that the convex hull of $\{x_1, \ldots, x_m\}$ is equal to $C$, so each state
$z \in C$ can be expressed as

$$z = \sum_{i=1}^{m} \phi_{zz_i} x_i,$$

where $\{\phi_{zz_i} \mid i = 1, \ldots, m\}$ is a probability distribution:

$$\phi_{zz_i} \geq 0, \; i = 1, \ldots, m, \quad \sum_{i=1}^{m} \phi_{zz_i} = 1, \quad \text{for all } z \in C.$$

We view $\phi_{zz_i}$ as the aggregation probabilities.

Consider the operator $\hat{T}$ that transforms a function

$$r = \{r_z \mid z \in C\}$$

into the function

$$\hat{T}r = \{(\hat{T}r)(z) \mid z \in C\}$$

defined by

$$(\hat{T}r)(z) = \min_{u \in U} \mathbb{E}_w \left\{ g(z, u, w) + \alpha \sum_{j=1}^{m} \phi_{f(z,u,w)x_j} r_{x_j} \right\}, \quad z \in C,$$

where $\phi_{f(z,u,w)x_j}$ are the aggregation probabilities of the next state $f(z, u, w)$.

It can then be shown that $\hat{T}$ is a contraction mapping with respect to the maximum norm (we give the proof for a similar result in the next section). Let $\hat{r}$ denote the unique fixed point of $\hat{T}$, so that for the representative states we have

$$\hat{r}_{x_i} = (\hat{T}\hat{r})(x_i), \quad i = 1, \ldots, m.$$

This is Bellman’s equation for an aggregate finite-state discounted DP problem whose states are $x_1, \ldots, x_m$.

The transitions in this problem occur as follows: from state $x_i$ under control $u$, we first move to $f(x_i, u, w)$ at cost $g(x_i, u, w)$, and then we move to a state $x_j$, $j = 1, \ldots, m$, according to the probabilities $\phi_{f(z,u,w)x_j}$. The optimal costs $r^*_{x_i}, i = 1, \ldots, m$, of this problem can often be obtained by standard VI and PI methods that may or may not use simulation. We may then approximate the optimal cost function of the original problem by

$$\tilde{J}(z) = \sum_{i=1}^{m} \phi_{zz_i} r^*_{x_i}, \quad \text{for all } z \in C,$$

and prove under reasonable conditions that the optimal discretized solution converges to the optimal as the number of representative states increases.
In the case where \( C \) is the belief space of an \( \alpha \)-discounted POMDP, the representative states/beliefs and the aggregation probabilities define an aggregate problem, which is a finite-state \( \alpha \)-discounted problem with a perfect state information structure. This problem can be solved with the exact methods of Chapter 4 if either the aggregate transition probabilities and transition costs can be obtained analytically (in favorable cases) or if the number of representative states is small enough to allow their calculation by simulation. The aggregate problem can also be addressed with the approximate methods of Chapter 5, such as problem approximation/certainty equivalence approaches. It can also be addressed by a rollout method, which is suitable for an on-line implementation.