Throughout this and the next chapters, we will abbreviate a belief by the symbol b, instead of the more elaborate bel used in previous chapters.

POMDPs compute a value function over belief space:

$$V_T(b) \quad = \quad \gamma \max_u \left[r(b,u) + \int V_{T-1}(b')\; p(b' \mid u, b)\; db' \right] \tag{15.2}$$

with $V_1(b) \;=\; \gamma \max_u E_x[\, r\, (x,u)]$. The induced control policy is as follows:

$$\pi_T(b) \quad = \quad \underset{u}{\text{argmax}} \left[r(b,u) + \int V_{T-1}(b')\; p(b' \mid u, b)\; db' \right] \tag{15.3}$$

A belief is a probability distribution; thus, each value in a POMDP is a function of an entire probability distribution. This is problematic. If the state space is finite, the belief space is continuous, since it is the space of all distributions over the state space. Thus, there is a continuum of different values; whereas there was only a finite number of different values in the MDP case. The situation is even more delicate for continuous state spaces, where the belief space is an infinitely-dimensional continuum.

An additional complication arises from the computational properties of the value function calculation. Equations (15.2) and (15.3) integrate over all beliefs b'. Given the complex nature of the belief space, it is not at all obvious that the integration can be carried out exactly, or that effective approximations can be found. It should therefore come at no surprise that calculating the value function V_T is more complicated in belief space than it is in state space.

Luckily, an exact solution exists for the interesting special case of finite worlds, in which the state space, the action space, the space of observations, and the planning horizon are all finite. This solution represents value functions by *piecewise linear functions* over the belief space. As we shall see, the linearity of this representation arises directly from the fact that the expectation is a linear operator. The piecewise nature is the result of the fact that the robot has the ability to select controls, and in different parts of the belief space it will select different controls. All these statements will be proven in this chapter.

PIECEWISE LINEAR FUNCTION

This chapter discusses the general POMDP algorithm for calculating policies defined over the space of all belief distributions. This algorithm is computationally cumbersome but correct for finite POMDPs; although a variant will be discussed that is highly tractable. The subsequent chapter will discuss a number of more efficient POMDP algorithms, which are approximate but scale to actual robotics problems.