

state  $s$ . This means that we have a simplified version of the Bellman equation (17.5) relating the utility of  $s$  (under  $\pi_i$ ) to the utilities of its neighbors:

$$U_i(s) = R(s) + \gamma \sum_{s'} P(s' | s, \pi_i(s)) U_i(s'). \quad (17.10)$$

For example, suppose  $\pi_i$  is the policy shown in Figure 17.2(a). Then we have  $\pi_i(1, 1) = U_p$ ,  $\pi_i(1, 2) = U_p$ , and so on, and the simplified Bellman equations are

$$\begin{aligned} U_i(1, 1) &= -0.04 + 0.8U_i(1, 2) + 0.1U_i(1, 1) + 0.1U_i(2, 1), \\ U_i(1, 2) &= -0.04 + 0.8U_i(1, 3) + 0.2U_i(1, 2), \\ &\vdots \end{aligned}$$

The important point is that these equations are *linear*, because the “max” operator has been removed. For  $n$  states, we have  $n$  linear equations with  $n$  unknowns, which can be solved exactly in time  $O(n^3)$  by standard linear algebra methods.

For small state spaces, policy evaluation using exact solution methods is often the most efficient approach. For large state spaces,  $O(n^3)$  time might be prohibitive. Fortunately, it is not necessary to do *exact* policy evaluation. Instead, we can perform some number of simplified value iteration steps (simplified because the policy is fixed) to give a reasonably good approximation of the utilities. The simplified Bellman update for this process is

$$U_{i+1}(s) \leftarrow R(s) + \gamma \sum_{s'} P(s' | s, \pi_i(s)) U_i(s'),$$

