How to Make Reactive Planners Risk-Sensitive*

(Without Altering Anything)

Sven Koenig and Reid G. Simmons
School of Computer Science
Carnegie Mellon University
Pittsburgh, PA 15213-3891
skoenig@cs.cmu.edu, reids@cs.cmu.edu

Abstract

Probabilistic planners can have various planning objectives: usually they either maximize the probability of goal achievement or minimize the expected execution cost of the plan. Researchers have largely ignored the problem how to incorporate risk-sensitive attitudes into their planning mechanisms. We discuss a risk-sensitive planning approach that is based on utility theory. Our key result is that this approach can, at least for risk-seeking attitudes, be implemented with *any* reactive planner that maximizes (or satisfices) the probability of goal achievement. First, the risk-sensitive planning problem is transformed into a different planning problem, that is then solved by the planner. The larger the probability of goal achievement of the resulting plan, the better its expected utility is for the original (risk-sensitive) planning problem. This approach extends the functionality of reactive planners that maximize the probability of goal achievement, since it allows one to use them (unchanged) for risk-sensitive planning.

Introduction

In the last several years, numerous reactive planning methods have been developed that are able to deal with probabilistic domains. Examples include (Bresina & Drummond 1990), (Koenig 1991), (Dean *et al.* 1993), and others. Given a planning problem, reactive planners determine state-action mappings (either implicitly or explicitly). Such closed-loop plans specify for every state the action that the agent has to execute when it is in that state. Not all reactive planning approaches have the same objective: different planners consider different plans to be optimal for the same planning problem. In this paper, we concentrate on three planning objectives: to maximize the probability of goal achievement, to minimize the expected execution cost, and to maximize the expected utility of plan execution.

In some domains, it is impossible to determine state-action mappings that are guaranteed to achieve a given goal. Even if a plan exists that achieves the goal with probability one, time might not permit to find it. When planning in such domains, one is usually satisfied with finding plans that maximize the probability of goal achievement. However, if one is willing and able to determine more than one plan that is always successful, one needs a criterion for choosing among these plans. A common metric is the execution cost of a plan.[1] Since probabilistic plans are "lotteries" in that their execution costs can vary from plan execution to plan execution, planners usually choose the plan for execution that minimizes the *expected* execution cost (when optimizing) or, at least, one whose expected execution cost is smaller than a given threshold (when satisficing).

Such planners are called risk-neutral, because they consider two plans with the same expected execution cost to be equally good, even if their variances are not equal. In contrast, a risk-seeking planner ("gambler") is willing to accept a plan with a larger expected execution cost if the uncertainty is sufficiently increased, and a risk-averse planner ("insurance holder") accepts a plan with a larger expected execution cost only if the uncertainty is sufficiently decreased. Since human decision makers are usually not risk-neutral for non-repetitive planning tasks, the plans produced by planning methods should reflect the risk-attitudes of the people that depend on them.

Utility theory (von Neumann & Morgenstern 1947), a part of decision theory, provides a normative framework for making decisions according to a given risk attitude, provided that the decision maker accepts a few simple axioms and has unlimited planning resources available. Its key result is that, for every risk attitude, there exists a utility function that transforms costs c into real values $u(c)$ ("utilities") such that it is rational to maximize expected utility. A planner that is based on utility theory chooses the plan for execution that maximizes the expected utility of the cost of plan execution.

The utility-theoretic approach encompasses the other two objectives: Maximizing the probability of goal achievement is rational according to utility theory if the agent prefers a smaller execution cost over a larger one and incurs total execution cost c_{goal} for achieving a goal state and a larger total execution cost $c_{non-goal}$ otherwise. Minimizing expected execution cost is

*This research was supported in part by NASA under contract NAGW-1175. The views and conclusions contained in this document are those of the authors and should not be interpreted as representing the official policies, either expressed or implied, of NASA or the U.S. government.

[1] In some domains it is even easy to construct plans that always succeed. An example is travel planning. When planning how to get to a given conference, one can easily devise plans that are always successful under normal circumstances. Since human decision makers usually ignore exceptional circumstances (such as sudden illnesses), it appears reasonable for a planner to do so as well. Possible evaluation metrics for travel plans include travel cost and travel time.

rational according to utility theory if the agent is risk-neutral, i.e. has utility function $u(c) = c$.

Although the application of utility theory to planning tasks has been studied by (Etzioni 1991), (Russell & Wefald 1991), (Haddawy & Hanks 1992), (Wellman & Doyle 1992) and (Goodwin & Simmons 1992), most researchers have ignored the problem how to incorporate risk-sensitive attitudes into their planning mechanisms. Notable exceptions include (Karakoulas 1993) and (Heger 1994).

In the following, we will describe an approach for utility-based planning. We will show, perhaps surprisingly, that *any* reactive planner that maximizes (or satisfices) the probability of goal achievement can be used to implement our approach, at least in the risk-seeking case. First, the risk-sensitive planning problem is transformed into a different planning problem, that is then solved by the planner. The larger the probability of goal achievement of the resulting plan, the better its expected utility is for the original (risk-sensitive) planning problem. Thus, planners that maximize or satisfice the probability of goal achievement can also be used for risk-sensitive planning, and our approach extends their functionality. It also extends our previous work by showing that these planners, that do not consider costs, can – perhaps surprisingly – be used to take execution costs into account.

The Planning Framework

The probabilistic planning framework that we use in this paper is similar to those used by (Bresina & Drummond 1990), (Dean *et al.* 1993), and most table-based reinforcement learning approaches: S is a finite set of states, $s_0 \in S$ the start state, and $G \subseteq S$ a set of goal states. When the agent reaches a goal state, it has solved the task successfully and execution stops. $A(s)$ denotes the finite set of actions that are available in non-goal state s. After the agent executes an action $a \in A(s)$ in s, nature determines the outcome of the action with a coin flip: with transition probability $p^a[s, s']$, the agent incurs an action cost $c^a[s, s'] < 0$ and is in successor state s'. This assumes that the outcomes of all action executions are mutually independent given the current state of the agent (Markov property). The action costs reflect the prices of the consumed resources, for example time needed or energy spent. We assume that the transition probabilities and transition costs are completely known to the planner and do not change over time. We do not assume, however, that the planner uses a planning approach that operates in the state space (instead of, say, the space of partial plans).

A planning domain with the above properties can for example be described with a STRIPS-like notation. Although the original STRIPS-notation (Fikes & Nilsson 1971) applies to deterministic domains only, it can easily be augmented for probabilistic domains (Koenig 1991). As an illustration, consider a probabilistic blocks-world. In every blocks-world state, one can move a block that has a clear top onto either the table or a different block with a clear top. With probability 0.1, the moved block ends up at its intended destination and the move action takes two minutes to complete. With probability 0.9, however, the gripper loses the block and it ends up directly on the table. If the block slips, it does so after only one minute. Thus, the execution time of the move action is less if slipping occurs. This domain can be modeled with three augmented STRIPS-rules: "move block X from the top of block Y on top of block Z", "stack block X on top of block Y", "unstack block X from block Y". The first move operator can for example be expressed as follows:

```
move(X,Y,Z)
  precond:   on(X,Y), clear(X), clear(Z), block(X),
             block(Y), block(Z), unequal(X,Z)
  outcome:   /* the primary outcome */
    prob:    0.1
    cost:    -2
    delete:  on(X,Y), clear(Z)
    add:     on(X,Z), clear(Y)
  outcome:   /* failure: block X falls down */
    prob:    0.9
    cost:    -1
    delete:  on(X,Y)
    add:     clear(Y), on(X, table)
```

A plan is a state-action mapping that assigns an action $a[s] \in A(s)$ to each non-goal state s. Given the above assumptions, plans can be restricted to state-action mappings without losing optimality. For a given plan, we define the probability of goal achievement $g[s]$ of state s as the probability with which the agent eventually reaches a goal state if it is started in s and obeys the plan. Iff this probability equals one, we say that the plan solves s. The expected cost of state s for the given plan is the expected sum of the costs of the actions that the agent executes from the time at which it starts in s until it stops in a goal state. Similarly, the expected utility $u[s]$ of state s is the expected utility of the sum of the costs of the executed actions.

In the following, we make the assumption that all states are solvable, because it simplifies the description of our approach. However, the approach can be extended to produce plans that allow the agent to stop execution without having reached a goal state, for example because a goal state cannot be reached at all or it is to costly to do so (Koenig & Simmons 1993).

Utility-Based Planning

A utility-based planner has to solve planning task PT1: given a utility function, find a plan for which the start state has the largest expected utility.

In order to come up with a good planning approach, we need to be concerned about its complexity. If the agent were *risk-neutral*, then the planning task could be solved, for example, with dynamic programming methods from Markov decision theory in a time that is polynomial in the size of the state space (Bertsekas 1987). Unfortunately, it is not possible to solve the *risk-sensitive* planning task PT1 by first replacing all action costs with their respective utilities and then using a risk-neutral planner on the resulting planning task, because in general $u(c_1 + c_2) \neq u(c_1) + u(c_2)$ for two costs c_1 and c_2. Even worse, dynamic programming methods can no longer be used in any way without considering the action costs that the agent has already accumulated when deciding on an action, because the best action in a state is no longer guaranteed to be

independent of the already accumulated cost.[2]

Our approach for overcoming this problem is to limit the utility functions to those that maintain the Markov property. Ultimately, there is a trade-off between closeness to reality and efficiency. If finding plans that are optimal for a given risk-attitude is inefficient, then one can use approximations: one can, for example, use approximate planning algorithms on exact utility functions or exact planning algorithms on approximate utility functions. Our approach belongs into the latter category and was partly motivated by the fact that assessing the utility function of a decision maker will usually not be possible with total accuracy.

The only utility functions that maintain the Markov property are the identity function, convex exponential functions $u(c) = \gamma^c$ for $\gamma > 1$, concave exponential functions $u(c) = -\gamma^c$ for $0 < \gamma < 1$, and their positively linear transformations (Watson & Buede 1987). Since the utility functions are parameterized with a parameter γ, one can express various degrees of risk-sensitivity ranging from being strongly risk-averse over being risk-neutral to being strongly risk-seeking. The larger γ, the more risk-seeking the agent is, and vice versa. Although utility functions of this kind can approximate a wide range of risk-attitudes, they have their limitations. They cannot be used to explain, for example, the behavior of agents that are risk-seeking and risk-averse at the same time ("insurance holders that buy lottery tickets").

(Howard & Matheson 1972) apply these utility functions to Markov decision problems for which every state-action mapping determines an irreducible (that is, strongly connected) Markov chain. Unfortunately, planning task PT1 does not possess these properties, and thus we cannot use their methods and proofs unchanged.

Planning for Risk-Seeking Agents

In this section, we deal with risk-seeking agents that have utility function $u(c) = \gamma^c$ for $\gamma > 1$. This class of utility functions approximates well-studied risk-attitudes as γ approaches one or infinity. To simplify the explanation, we use two terms from utility theory. A "lottery" is recursively defined to be either a cost that is received for sure (that is, with probability one) or a probability distribution over lotteries. If the expected utility of a lottery is x, then $u^{-1}(x)$ is called its "certainty monetary equivalent."

For γ approaching one, the certainty monetary equivalent of any state for any plan approaches the expected cost of that state. Therefore, the optimal plan for a risk-neutral agent is also best for a risk-seeking agent if γ approaches one.

Proof: Assume that the execution of the plan leads with probability p_i to execution cost c_i if the agent is started in state s. Then, the expected cost of s equals $\sum_i p_i c_i$ and its certainty monetary equivalent is $u^{-1}(\sum_i p_i u(c_i))$. Thus, $\lim_{\gamma\to 1} u^{-1}(\sum_i p_i u(c_i)) = \lim_{\gamma\to 1} \log_\gamma(\sum_i p_i \gamma^{c_i}) = \lim_{\gamma\to 1} \frac{\ln(\sum_i p_i \gamma^{c_i})}{\ln \gamma} \overset{L'Hôpital}{=} \lim_{\gamma\to 1} \frac{(\sum_i p_i c_i \gamma^{c_i - 1})/\sum_i p_i \gamma^{c_i}}{1/\gamma} = \lim_{\gamma\to 1} \frac{\sum_i p_i c_i \gamma^{c_i}}{\sum_i p_i \gamma^{c_i}} = \frac{\lim_{\gamma\to 1}\sum_i p_i c_i \gamma^{c_i}}{\lim_{\gamma\to 1}\sum_i p_i \gamma^{c_i}} = \frac{\sum_i p_i c_i}{1} = \sum_i p_i c_i.$

In contrast, for γ approaching infinity, the certainty monetary equivalent of any state for any plan approaches the cost of that state if nature acts like a friend (that is, chooses the action outcomes not with a coin flip, but deliberately so that it is best for the agent). Of course, according to our assumptions, nature does flip coins. We call an agent that assumes (wrongly) that nature helps it as much as it can and calculates it utilities accordingly "extremely risk-seeking." Thus, max-max-ing ("both the agent and nature maximize the reward for the agent"), which calculates the utility of a non-goal state s for a given plan as $u[s] = \max_{s' \in S}(c^{a[s]}[s, s'] + u[s'])$, determines the plan that is best for a risk-seeking agent if γ approaches infinity.

Proof: Assume again that the execution of the plan leads with probability p_i to execution cost c_i if the agent is started in state s. Then, $\max_i c_i = \log_\gamma \gamma^{\max_i c_i} = \log_\gamma \sum_i p_i \gamma^{\max_i c_i} \geq \log_\gamma \sum_i p_i \gamma^{c_i} \geq \max_i \log_\gamma(p_i \gamma^{c_i}) = \max_i(\log_\gamma p_i + c_i)$. It follows that $\max_i c_i = \lim_{\gamma\to\infty} \max_i c_i \geq \lim_{\gamma\to\infty} \log_\gamma \sum_i p_i \gamma^{c_i} \geq \lim_{\gamma\to\infty} \max_i(\log_\gamma p_i + c_i) = \max_i c_i$, and thus $\lim_{\gamma\to\infty} u^{-1}(\sum_i p_i u(c_i)) = \lim_{\gamma\to\infty} \log_\gamma \sum_i p_i \gamma^{c_i} = \max_i c_i$.

In the following, we will first show how to calculate the expected utility of a given plan. Then, we will transform planning task PT1 into one for an agent that maximizes the probability of goal achievement. Finally, we will apply our approach to a simple probabilistic navigation task.

Calculating the Expected Utility of a Plan Assume that, for some planning problem, a plan (that is, a state-action mapping) is given that assigns action $a[s]$ to non-goal state s. The expected utility $u[s_0]$ of this plan can recursively be calculated as follows: The utility of a goal state s is $u[s] = u(0) = \gamma^0 = 1$. After the agent has executed action $a[s]$ in a non-goal state s, it incurs action cost $c^{a[s]}[s, s']$ and is in successor state s' with probability $p^{a[s]}[s, s']$. In state s', it faces a lottery again. This lottery has expected utility $u[s']$ and certainty monetary equivalent $u^{-1}(u[s'])$. According to the axioms of utility theory, the lottery can be replaced with its certainty monetary equivalent. Then, the agent incurs total cost $c^{a[s]}[s, s'] + u^{-1}(u[s'])$ with probability $p^{a[s]}[s, s']$. Thus, the expected utility of s can be calculated as follows:[3]

$$\begin{aligned} u[s] &= \sum_{s' \in S} p^{a[s]}[s, s'] u(c^{a[s]}[s, s'] + u^{-1}(u[s'])) \\ &= \sum_{s' \in S} p^{a[s]}[s, s'] \gamma^{c^{a[s]}[s,s'] + u^{-1}(u[s'])} \\ &= \sum_{s' \in S} p^{a[s]}[s, s'] \gamma^{c^{a[s]}[s,s']} \gamma^{u^{-1}(u[s'])} \\ &= \sum_{s' \in S} p^{a[s]}[s, s'] \gamma^{c^{a[s]}[s,s']} u[s'] \end{aligned}$$

[2] For a detailed explanation and an example see (Koenig & Simmons 1993).

[3] This corresponds to the policy-evaluation step in (Howard & Matheson 1972) with the "certain equivalent gain" $\bar{g} = 0$.

$$= \sum_{s' \in S \setminus G} p^{a[s]}[s,s']\gamma^{c^{a[s]}[s,s']}u[s'] + \sum_{s' \in G} p^{a[s]}[s,s']\gamma^{c^{a[s]}[s,s']}$$

Transforming the Planning Problem Using the results of the previous section, we can now show how every planning task PT1 for a risk-seeking agent can be transformed into an equivalent planning task PT2 for an agent that maximizes the probability of goal achievement.

Assume again that a state-action mapping is given and let $\bar{p}^{a[s]}[s,s']$ denote the transition probabilities for planning task PT2. The probability of goal achievement $g[s]$ of a goal state s is one. For a non-goal state s, it is

$$g[s] = \sum_{s' \in S} \bar{p}^{a[s]}[s,s']g[s'] = \sum_{s' \in S \setminus G} \bar{p}^{a[s]}[s,s']g[s'] + \sum_{s' \in G} \bar{p}^{a[s]}[s,s']$$

Comparing these results to the ones in the previous section shows that $g[s] = u[s]$ for $s \in S$ if $\bar{p}^{a[s]}[s,s'] = p^{a[s]}[s,s']\gamma^{c^{a[s]}[s,s']}$ for $s \in S \setminus G$ and $s' \in S$.

Thus, planning task PT1 for a risk-seeking agent with utility function $u(c) = \gamma^c$ is equivalent to the following planning task PT2 for an agent that maximizes the probability of goal achievement: The state space, action space, start state, and goal states remain unchanged. When the agent executes action $a \in A(s)$ in any non-goal state s, it is in successor state s' with transition probability $p^{a[s]}[s,s']\gamma^{c^{a[s]}[s,s']}$. The action costs of the actions do not matter.

This transformation is trivial and can be done in linear time. It changes only the transition probabilities, but neither the state space, action space, nor which states are goal states. The probabilities do not add up to one – we could remedy this by introducing a single non-goal state in which only one action is applicable, which has action cost zero and does not change state. When the agent executes an action $a \in A(s)$ in any other non-goal state s, it reaches this new non-goal state with transition probability $1 - \sum_{s' \in S} p^{a[s]}[s,s']\gamma^{c^{a[s]}[s,s']}$. Since the probability of goal achievement for the new non-goal state is zero, it does not affect the calculations and all we need to do is to recalculate the probabilities in the manner described above.

For example, consider again the augmented STRIPS-rule for the probabilistic blocks-world and assume that $\gamma = 2$. The transformed STRIPS-rule looks as follows:

```
move(X,Y,Z)
  precond:   on(X,Y), clear(X), clear(Z), block(X),
             block(Y), block(Z), unequal(X,Z)
  outcome:   /* the primary outcome */
    prob:    0.025
    delete:  on(X,Y), clear(Z)
    add:     on(X,Z), clear(Y)
  outcome:   /* failure: block X falls down */
    prob:    0.450
    delete:  on(X,Y)
    add:     clear(Y), on(X, table)
```

With the complementary probability (0.525), the action execution results in the agent no longer being able to reach a goal state.

Determining Optimal Plans The last section demonstrates that the expected utility of a plan for planning task PT1 equals the probability of goal achievement of the same plan for planning task PT2. Thus, a plan is optimal for planning task PT1 if it is optimal for planning task PT2 as well (and vice versa).

Any planning method that determines plans that maximize the probability of goal achievement (or plans whose probability of goal achievement exceeds a given threshold) can be used to maximize (or satisfice) expected utility for planning task PT1. One can apply the planning method by first transforming the utility-based planning task into a strictly probability-based planning task, as described above. The optimal plan for the transformed planning task is optimal for the original planning task as well. Thus, perhaps surprisingly, planners that maximize the probability of goal achievement *can* be used to take execution costs into account.

Planning for Risk-Averse Agents

For risk-averse agents, one has to use a utility function from the family $u(c) = -\gamma^c$ (or any positively linear transformation thereof) for $0 < \gamma < 1$. For γ approaching zero, the certainty monetary equivalent of any state for any plan approaches the cost of that state if nature acts like an enemy (that is, chooses action outcomes not with a coin flip, but deliberately so that it is worst for the agent). We call an agent that assumes (wrongly) that nature hurts it as much as it can and calculates its utilities accordingly "extremely risk-averse." Planning for extremely risk-averse agents has recently been studied by (Moore & Atkeson 1993) and (Heger 1994).

Although the values $p^{a[s]}[s,s']\gamma^{c^{a[s]}[s,s']}$ can no longer be interpreted as probabilities (since $\sum_{s' \in S} p^{a[s]}[s,s']\gamma^{c^{a[s]}[s,s']} > 1$), one can proceed as outlined for risk-seeking agents in the previous section if one addresses the following problem: The solution $u[s_0]$ of the system of linear equations from Section "Calculating the Expected Utility of a Plan" can now be finite even for plans that have expected utility minus infinity. The planning methods can then erroneously return such plans as optimal solutions. Since such plans are easy to characterize, one can modify planning algorithms to make sure that these plans get ignored during planning (Koenig & Simmons 1994). Thus, while we cannot claim that *any* probabilistic planner can be used unchanged for the risk-averse case (as *is* the case for risk-seeking behavior), we believe that for many of the existing algorithms only slight modifications would be needed to enable them to handle the risk-averse case as well.

An Example

In the following, we demonstrate our approach to risk-sensitive planning on a probabilistic navigation task and show how the best plan changes as the agent becomes more and more risk-seeking. To implement the approach, we used a simple dynamic programming algorithm that maximizes the probability of goal achievement (Koenig & Simmons 1993).

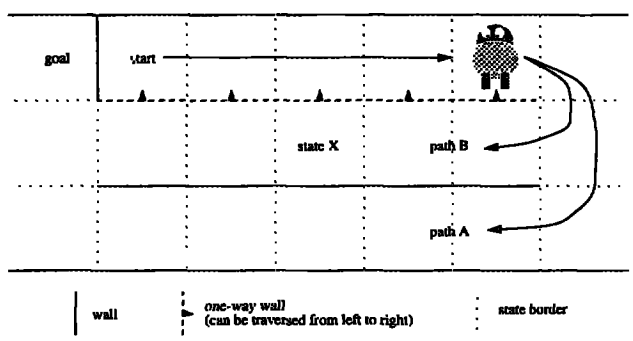


Figure 1: A Path-Planning Problem

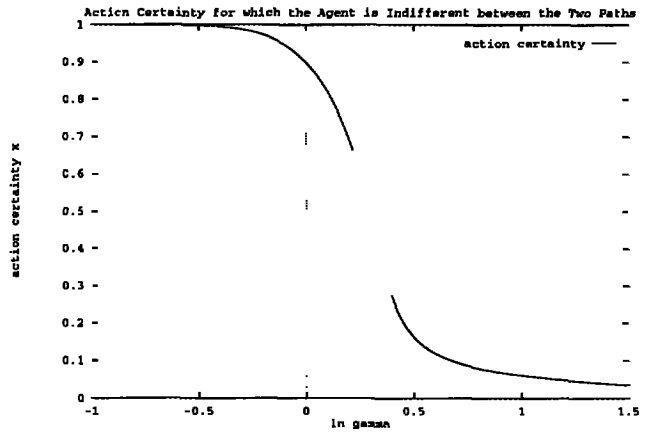


Figure 2: Solution of the Path-Planning Problem

Consider the simple path-planning domain shown in Figure 1. The states of this grid-world correspond to locations. In every state, the agent has at most four actions available, namely to go up, right, down, or left. All actions take the agent one minute to execute, but they are not necessarily deterministic. They succeed with probability $\frac{1+x}{3-x}$, but their outcomes deviate ninety degrees to the left or right of the intended direction with probability $\frac{1-x}{3-x}$ each. Thus, $x \in [0, 1]$ is a parameter for the certainty of the actions: the larger the value of x is, the more certain their outcomes are. Actions have deterministic outcomes if $x = 1$; their intended outcome and its two deviations are equally likely for the other extreme, $x = 0$.

In every state, the agent can execute all of the actions whose intended direction is not immediately blocked by a wall. Besides standard walls, the grid-world also contains "one-way walls," that can be traversed from left to right, but not in the opposite direction. (They might for example be steep slopes that the agent can slide down, but not climb up.) If the agent executes an action and it has the intended outcome, the agent cannot bump into a wall. However, running into a wall is possible for unintended outcomes, in which case the agent does not change its location. As an example, consider state X in figure 1 and assume that $x = 0.5$. In this state, the agent can go left, up, or right. If it tries to go left, it will succeed with probability 0.6, unintentionally go up with probability 0.2, and unintentionally remain in state X with the same probability (since it cannot go down).

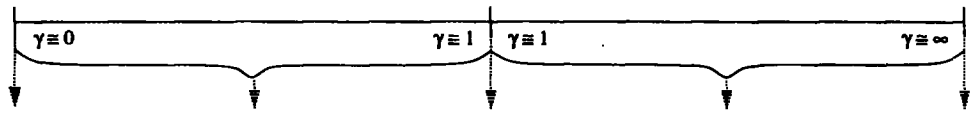
The agent can reach the goal state from the start state on two different paths. If the actions have deterministic effects (that is, if $x = 1$), the traversal of path B takes 13 minutes and the one of path A 15 minutes. Thus, the agent prefers path B over path A, independently of its risk-attitude. If the actions do not have deterministic effects, however, the agent risks traversing a one-way wall unintentionally when following path B, in which case it has to retrace parts of its path.

We use $\bar{x}_\gamma$ to denote the value of the action certainty parameter that makes an agent with risk parameter γ indifferent between the two paths. If $\bar{x}_\gamma < x$, then the agent chooses path B, otherwise it chooses path A. Figure 2 shows how the value of $\bar{x}_\gamma$ varies with the natural logarithm of γ. The smaller $\ln\gamma$, the more risk-averse the agent is. The figure comprises risk-seeking ($\ln\gamma > 0$), risk-neutral ($\ln\gamma = 0$), and risk-averse behavior ($\ln\gamma < 0$). The graph shows that $\bar{x}_\gamma$ decreases the more risk-seeking the agent becomes. It approaches zero in the limit: an extremely risk-seeking agent prefers path B over path A, since path B can be traversed in 13 minutes in the best case, whereas path A cannot be traversed in less than 15 minutes.

Conclusion

This paper concerns probabilistic planning for *risk-sensitive* agents, since there are many situations where it is not appropriate to determine plans that maximize the probability of goal achievement or minimize expected execution cost. Our approach to risk-sensitive planning fills the gap between approaches previously studied in the AI planning literature, namely the approach of minimizing expected execution cost (risk-neutral attitude) and the approach of assuming that nature acts like a friend (extremely risk-seeking attitude) or enemy (extremely risk-averse attitude) which we can asymptotically approximate as shown in Figure 3.

Building on previous work by (Howard & Matheson 1972), we demonstrated that *any* reactive planner that maximizes the probability of goal achievement can be used to determine optimal plans for risk-seeking agents and, perhaps surprisingly, can therefore be used to take execution costs into account. First, the risk-seeking planning problem is transformed into a different planning problem for which the planner then determines the plan with the largest (or a good) probability of goal achievement. This plan has the largest (or a good) expected utility for the original planning problem. The transformation is not complicated. Only the transition probabilities have to be changed according to the following simple rule: if an action leads with probability p and action cost c to a certain outcome, then its new transition probability is $p\gamma^c$ for a risk-seeking agent with degree of risk-sensitivity γ. For



risk attitude	extremely risk-averse	risk-averse	risk-neutral	risk-seeking	extremely risk-seeking
formalization	max-min-ing	utility theory	utility theory	utility theory	max-max-ing
utility of lottery	$u(L) = \min_i u(c_i)$	$u(L) = \sum_i p_i \cdot u(c_i)$	$u(L) = \sum_i p_i \cdot u(c_i)$	$u(L) = \sum_i p_i \cdot u(c_i)$	$u(L) = \max_i u(c_i)$
utility function		$u(c) = -m \cdot \gamma^c + n$ $(m>0, 0<\gamma<1)$	$u(c) = m \cdot c + n$ $(m>0)$	$u(c) = m \cdot \gamma^c + n$ $(m>0, \gamma>1)$	

(L is a lottery: cost c_i is won with probability p_i for all i)

Figure 3: Continuum of Risk-Sensitive Behavior

risk-averse agents, the problem is a bit more complex, but we believe that many existing planners can be extended in simple ways to deal with the risk-averse case.

Our approach can therefore be used to extend the functionality of reactive planners that maximize the probability of goal achievement, since they can now also be used to maximize expected utility for a particular class of utility functions.

Acknowledgements

Thanks to Justin Boyan, Lonnie Chrisman, Matthias Heger, Andrew Moore, Joseph O'Sullivan, Stuart Russell, Jiří Sgall, and Mike Wellman for helpful discussions and comments.

References

Bertsekas, D. 1987. *Dynamic Programming, Deterministic and Stochastic Models*. Englewood Cliffs, NJ: Prentice-Hall.

Bresina, J., and Drummond, M. 1990. Anytime synthetic projection: Maximizing the probability of goal satisfaction. In *Proceedings of the AAAI*, 138–144.

Dean, T.; Kaelbling, L.; Kirman, J.: and Nicholson, A. 1993. Planning with deadlines in stochastic domains. In *Proceedings of the AAAI*, 574–579.

Etzioni, O. 1991. Embedding decision-analytic control in a learning architecture. *Artificial Intelligence* (1-3):129–159.

Fikes, R., and Nilsson, N. 1971. Strips: A new approach to the application of theorem proving to problem solving. *Artificial Intelligence* 2:189–208.

Goodwin, R., and Simmons, R. 1992. Rational handling of multiple goals for mobile robots. In *Proceedings of the First International Conference on AI Planning Systems*, 70–77.

Haddawy, P., and Hanks, S. 1992. Representation for decision-theoretic planning: Utility functions for deadline goals. In *Proceedings of the International Conference on Principles of Knowledge Representation and Reasoning*.

Heger, M. 1994. Risk and reinforcement learning. Technical report, Computer Science Department, University of Bremen, Bremen, Germany. (a shorter version has been accepted for publication in the Proceedings of the Eleventh International Machine Learning Conference, 1994).

Howard, R., and Matheson, J. 1972. Risk-sensitive Markov decision processes. *Management Science* 18(7):356–369.

Karakoulas, G. 1993. A machine learning approach to planning for economic systems. In *Proceedings of the Third International Workshop on Artificial Intelligence in Economics and Management*.

Koenig, S., and Simmons, R. 1993. Utility-based planning. Technical Report CMU–CS–93–222, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA.

Koenig, S., and Simmons, R. 1994. Risk-sensitive game-playing, any-time planning, and reinforcement learning. Technical report, School of Computer Science, Carnegie Mellon University, Pittsburgh, PA. (forthcoming).

Koenig, S. 1991. Optimal probabilistic and decision-theoretic planning using Markovian decision theory. Master's thesis, Computer Science Department, University of California at Berkeley, Berkeley, CA. (available as Technical Report UCB/CSD 92/685).

Moore, A., and Atkeson, C. 1993. The parti-game algorithm for variable resolution reinforcement learning in multidimensional state-spaces. In *Proceedings of the NIPS*.

Russell, S., and Wefald, E. 1991. *Do the Right Thing – Studies in Limited Rationality*. Cambridge, MA: The MIT Press.

von Neumann, J., and Morgenstern, O. 1947. *Theory of games and economic behavior*. Princeton, NJ: Princeton University Press, second edition.

Watson, S., and Buede, D. 1987. *Decision Synthesis*. Cambridge (Great Britain): Cambridge University Press.

Wellman, M., and Doyle, J. 1992. Modular utility representation for decision theoretic planning. In *Proceedings of the First International Conference on AI Planning Systems*, 236–242.