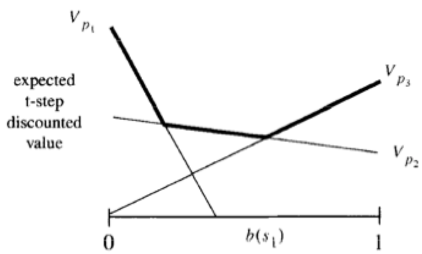
Structure of Optimal Policies

Each policy tree p induces a value function V_p that is linear in b. V_t is the upper surface of this collection of functions. Therefore, V_t is piecewise linear and convex.

Example: suppose there are just two states in the world. Then the belief state is completely determined by the probability of being in state 1. Then, for example:



When there are three states, belief state is determined by two parameters, and the optimal value function is a bowl shape composed of planar facets.

To choose an optimal action in belief state b, choose the action at the root of the policy tree p that maximizes $b \cdot \alpha_p$ (can store the regions in which each tree is best when constructing the value function)

In the example above, p_1, p_2, and p_3 are optimal in the three regions where they are bolded.

Computational Issues

First, note that, while it is theoretically possible for all trees of depth $t-1$ to contribute to the value function at some point in the belief space, this is not typical. So we restrict our attention to *useful* or undominated policy trees of depth $t-1$ when building trees of depth t.

Two steps, then: generation of trees of depth t and then pruning the set to get useful trees.

Generation: Generate superset $\mathcal{V}_t^+$ of the useful $t-1$ step policy trees $\mathcal{V}_{t_1}$. How many elements are there in this superset? Well, there are $|A|$ ways of choosing the action. Each can lead to $|\Omega|$ observations, and we can select any of the subtrees for any of the observations, so:

$$|A||\mathcal{V}_{t-1}|^{|\Omega|}$$

This is bad, and makes it very hard to solve POMDPs with more than a few states. So there has been a lot of work on approximate solutions.

Pruning? Can be done using linear programming. One linear program needed for each element of the set of policy trees, so doesn't add to asymptotic complexity. There are some relatively efficient methods for doing this.

Representation: basically just need to maintain one vector (linear in number of states) plus one action for each useful policy tree, so parsimonious.