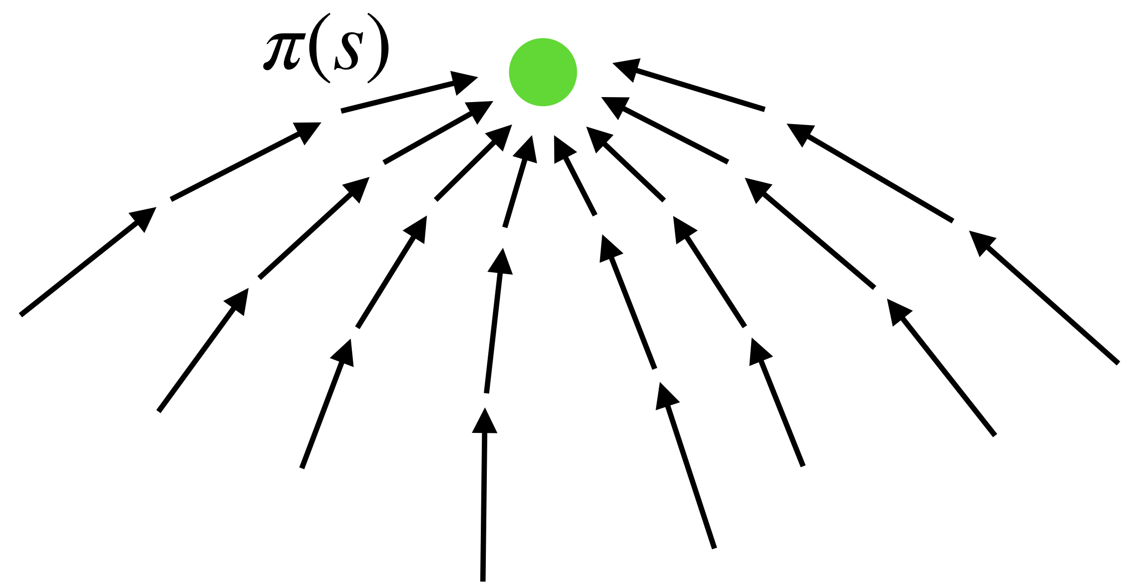
Model assumptions

- Discrete time
- A stochastic dynamics with Markov property:
  $\mathbf{s}_{t+1} = \mathbf{f}(\mathbf{s}_t, \mathbf{a}_t, \omega_t)$ with $\omega_t = \omega_{t-1}(\mathbf{s}_t, \mathbf{a}_t)$
- Later ω_t is normally distributed
- In stochastic settings optimise for an expected reward

Solution 1: Dynamic programming



- Dynamic Programming (Principle of Optimality - sufficient condition)
  - ➡ Compositionality of optimal paths
  - ➡ Closed-loop solutions: find a solution for all states at all times
- Solvable via Bellman equation in a backward recursive fashion
- Algorithms as e.g. Value iteration, Policy iteration (see Sutton and Barto)
- <u>No direct notion of constraints for states or actions!</u>