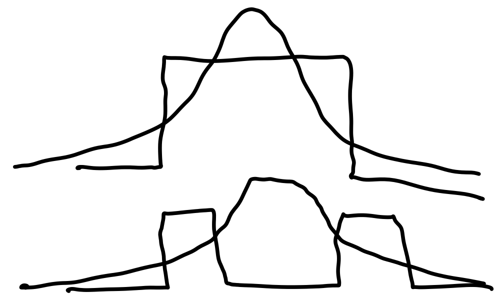
Last time

Exact
- Value Iteration ← converge asymptotically
  Discrete state and Actions

- LQR
  $S = \mathbb{R}^n \quad A = \mathbb{R}^m$
  $s_{t+1} \sim \mathcal{N}(A s_t + B a_t, \Sigma)$
  $R(s,a) = s^T Q s + a^T R a$



$V^* \rightarrow \pi^*$ Easy $\quad \pi^*(s) = \arg\max_a R(s,a) + \gamma \mathop{E}_{s' \sim T(s,a)}[V(s')]$

$\pi^* \rightarrow V^*$ Easy

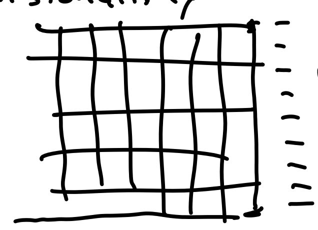
MDP $\rightarrow V^*$ Hard

MDP $\rightarrow \pi^*$ Hard

X_i binary R.V.

$$\lim_{N \to \infty} \sum_{i=1}^{N} x_i - \frac{N}{2} \sim \mathcal{N}(0, \;)$$

What if S is continuous
- Discretize ← Curse of Dimensionality
- Function Approximation $\quad |S| = k^d$



$\boxed{\text{Approximate Dynamic Programming}}$

Value Function Approx

$$V(s) = \sum_i \lambda_i \beta_i(s) = \lambda^T \beta(s)$$

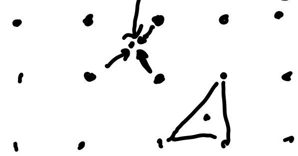
Local $\quad\quad$ Global

Local

- calculate $V(s)$ for $s \in \tilde{S}$
- Interpolate between

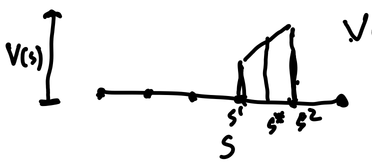
$S = \mathbb{R}^2$

$V(s)$ = weighted sum with more weight on nearer states

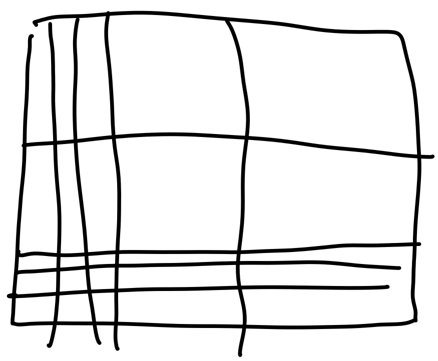


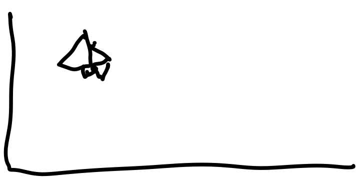
Multilinear 2^d

Simplex $d+1$



$V(s^*) =$ Look up in book





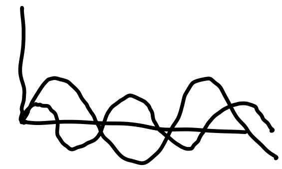
Global

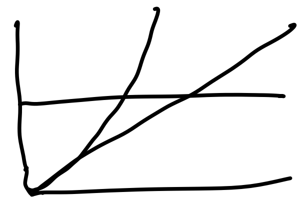
(I have never seen work)

m - Basis function

- Fourier
- Polynomial Basis

$\beta_i(s) = \cos(w_i s)$





$\beta_0(x) = 1$

$\beta_1(x) = x$

$\beta_2(x) = x^2$

Legendre Polynomials

$\lambda = \text{regress}(\beta, x, y)$

$\lambda = \arg\min_{\lambda} \left(\sum_j \lambda^T \beta(x_j) - y_i \right)^2$

$Y = 2x + w$

$w \sim N$

k-fold

Global/Local Approximation Value Iteration

$\lambda = 0$ / random

loop until convergence

  for i in 1 to n

    s_i = sample from S

    $v_i = \max_a \left(R(s,a) + \gamma \underset{s' \sim T(s,a)}{E} \left[\tilde{V}(s') \right] \right)$

  $\lambda = \text{regress}(\beta, s_{1:n}, v_{1:n})$

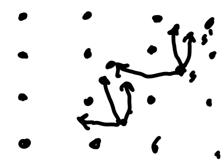
$\tilde{V}(s) = \lambda^T \beta(s)$

$\lambda^T \beta(s')$

$\sum_{s'} T(s' \mid s,a) \tilde{V}(s')$

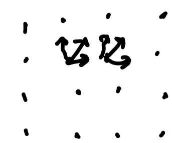
$\frac{1}{m} \sum_{k=1}^{m} \lambda \beta(G(s_i, a, w_k))$

Monte Carlo



Things to Worry about

- Gibb's Phenomenon ← Fix this



$\lambda \leftarrow 0$

loop

  for s_i in grid points

    $v_i = \max R(s_i, a) + \gamma E\left[\lambda^T \beta(s')\right]$

interpolation weights

  $\lambda = v_i$

$Q(s,a) = \lambda \beta^{s,a}(s,a)$

- $\pi(s) = \underset{a}{\text{argmax}}\ Q(s,a)$

- Features have to approximate V well at every iteration

Direct Policy Search

$$\pi(s) = f_\theta(s) \qquad \text{e.g. } \pi(s) = Ks$$

loop
$\theta' = \text{improvement}(\theta)$
evaluate $f_{\theta'}$

Derivative-free / zeroth order optimization

- Genetic Alg
- Simulated Annealing
- Cross-Entropy