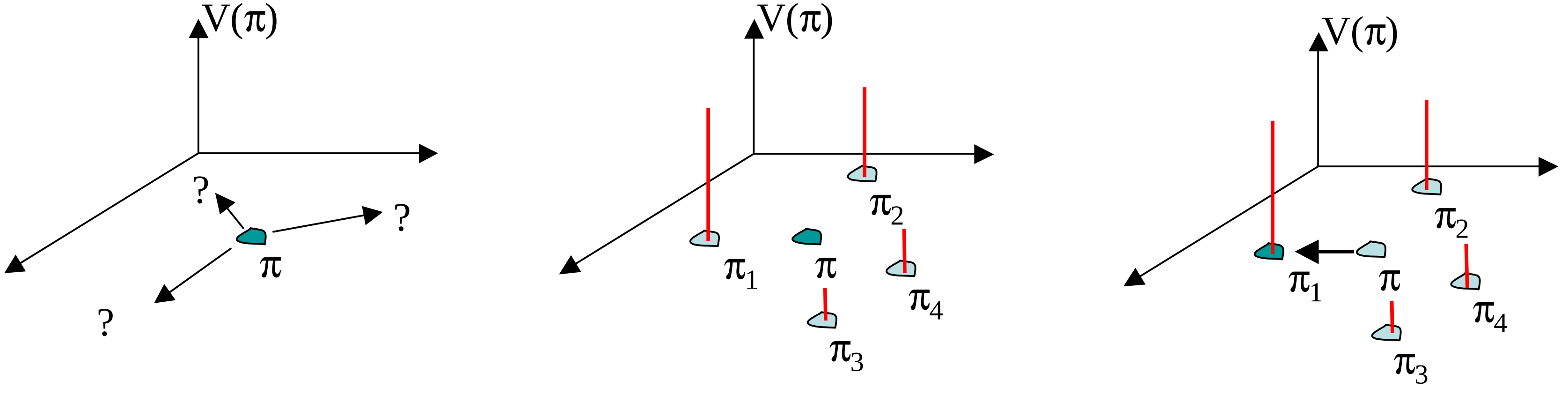
Hill Climbing Algorithm

- Policy $\pi = \{\theta_1, \ldots, \theta_{12}\}$, $V(\pi)$ = **walk speed** when using π
- Evaluate t (15) policies in the neighborhood of π
- From π, move towards the best neighboring policy

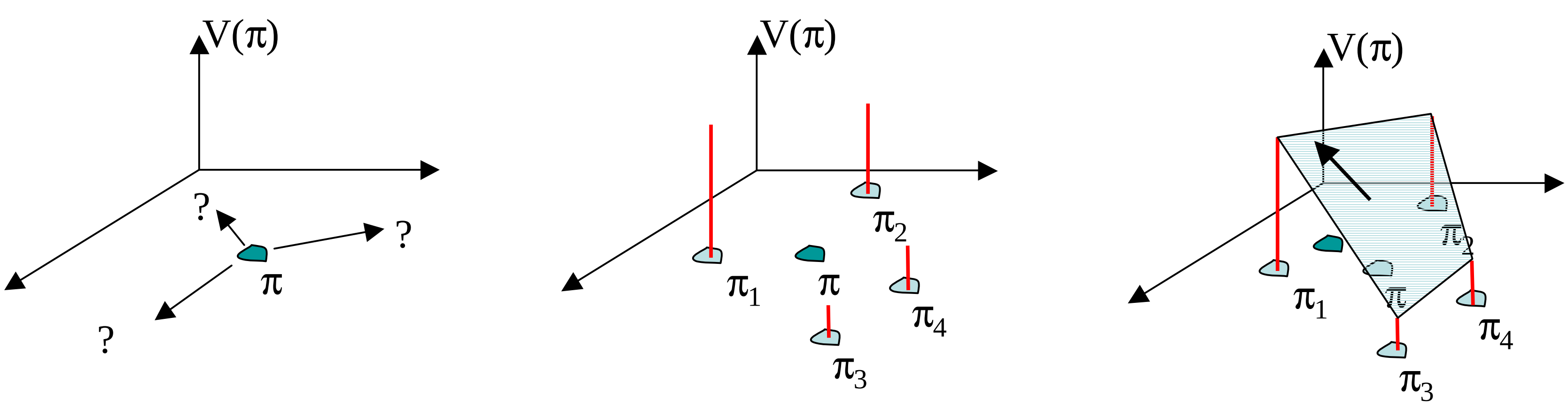


Policy Gradient RL

- Policy $\pi = \{\theta_1, \ldots, \theta_{12}\}$, $V(\pi)$ = **walk speed** when using π

- From π, move in the direction of the gradient of $V(\pi)$
  - Can't compute gradient directly: **estimate** empirically

- Evaluate neighboring policies to estimate gradient



Nate Kohl

Policy Gradient RL

- Determine **3 average values** for each dimension

- Compute an adjustment vector *A*:

$$A_i = \begin{cases} 0 & \text{If } Avg_{+0,\,i} > Avg_{+\varepsilon,\,i} \text{ and } Avg_{+0,\,i} > Avg_{-\varepsilon,\,i} \\ Avg_{+\varepsilon,\,i} - Avg_{-\varepsilon,\,i} & \text{otherwise} \end{cases}$$

- **Normalize** *A*, multiply by a scalar step size η

- $\pi = \pi + \eta A$