Hill Climbing Algorithm

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

- Evaluate $t$ (15) policies in the neighborhood of $\pi$

- From $\pi$, move towards the best neighboring policy
Policy Gradient RL

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

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

- Evaluate neighboring policies to estimate gradient
Policy Gradient RL

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

- Compute an adjustment vector $A_i$:

\[
A_i = \begin{cases} 
0 & \text{If } \text{Avg}_{+0,i} > \text{Avg}_{+\varepsilon,i} \text{ and } \text{Avg}_{+0,i} > \text{Avg}_{-\varepsilon,i} \\
\text{Avg}_{+\varepsilon,i} - \text{Avg}_{-\varepsilon,i} & \text{otherwise}
\end{cases}
\]

- **Normalize** $A$, multiply by a scalar step size $\eta$

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

Nate Kohl