Reward, observation and action shapes considered in the training environments

1 Observation

OBS1

$$[Ex, Ey, Ez, A1, A2, A3, A4, A5, A6] \tag{1}$$

OBS2

$$[Gx, Gy, Gz, A1, A2, A3, A4, A5, A6] \tag{2}$$

OBS3

$$[ETx, ETy, ETz, EGx, EGy, EGz, A1, A2, A3, A4, A5, A6] \tag{3}$$

OBS4

$$[EGx, EGy, EGz, A1, A2, A3, A4, A5, A6] \tag{4}$$

OBS5

$$[ETx, ETy, ETz, EGx, EGy, EGz, Gx, Gy, Gz, A1, A2, A3, A4, A5, A6] \tag{5}$$

Ei = End effector coordinate along the i axis
Gi = Goal coordinate along the i axis
EGi = Vector End effector - Goal along the i axis
ETx = Vector End effector - Torso along the i axis
Ai = Angular position of joint i

2 Action

3 Reward

r = reward
d_t = distance at time t
a = action
s = state
G = set of goals

3.1 Dense rewards

$$r = -d_t^2 \tag{6}$$

$$r = -d_t \tag{7}$$

$$r = -\alpha d_t - \beta a^T a \tag{8}$$

$$r = -\alpha d_{t-1}^p - d_t^p \tag{9}$$

$\alpha = 0$ or 1
$p = 1$ or 2
but don't work well...

$$r = -d_t - \|a_{t-1}\| \tag{10}$$

Penalise large torque

$$r = -d_t^2 + \frac{d_{t-1} - d_t}{d_t} \tag{11}$$

3.2 Sparse rewards

$$r = \begin{cases} -1, & \text{if } d \geq \epsilon \\ 0, & \text{if } d < \epsilon \end{cases} \tag{12}$$

$$r = \begin{cases} 1, & \text{if } s \in G \\ 0, & \text{otherwise} \end{cases} \tag{13}$$

3.3 Dense + sparse rewards

$$r = \begin{cases} -d_t, & \text{if no collision and } d \geq 3 \\ -d_t - 20\beta, & \text{if collision and } d \geq 3 \\ -d_t + 2, & \text{if no collision and } d < 3 \\ -d_t - 20\beta + 2, & \text{if collision and } d < 3 \end{cases} \tag{14}$$

$$r = \begin{cases} -1 - \beta \|a_{t-1}\|^2, & \text{if } d \geq \epsilon \\ 1 - \beta \|a_{t-1}\|^2, & \text{if } d < \epsilon \end{cases} \tag{15}$$

where $\beta \|a_{t-1}\|^2 \ll 1$ (penalise large actions)

$$r = \begin{cases} -d_t, & \text{if } d \geq \epsilon \\ 1, & \text{if } d < \epsilon \end{cases} \tag{16}$$

$$r = \begin{cases} -0.02, & \text{if } d \geq \epsilon \\ 1, & \text{if } d < \epsilon \end{cases} \tag{17}$$

$$r = \begin{cases} \alpha(d_{t-1} - d_t), & \text{if } d \geq \epsilon \\ \alpha(d_{t-1} - d_t) + 10, & \text{if } d < \epsilon \end{cases} \tag{18}$$

$$r = \begin{cases} -0.001, & \text{if } d \geq \epsilon \\ 10, & \text{if } d < \epsilon \end{cases} \tag{19}$$

$$r = \begin{cases} -0.001, & \text{if } d \geq \epsilon \\ 10, & \text{if } d < \epsilon \end{cases} \tag{20}$$

My attempts:

$$r = -d_t - \beta \|a_{t-1}\| \tag{21}$$

$$r = -d_t^2 - \beta \|a_{t-1}\| \tag{22}$$