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

where

- 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 Reward

REW1
$$r = -d_t^2 \tag{6}$$

REW2
$$r = -d_t^2 - \alpha \left\| a_t \right\| \tag{7}$$

REW3
$$r = d_{t-1} - d_t \tag{8}$$

REW4

$$r = -d_t^2 - \alpha \frac{d_{t-1} - d_t}{d_t} \tag{9}$$

REW5

$$r = \begin{cases} -1, & \text{if } d \geq \epsilon \\ 0, & \text{if } d < \epsilon \end{cases} \tag{10}$$

REW6

$$r = \begin{cases} 1, & \text{if } d \geq \epsilon \\ 0, & \text{if } d < \epsilon \end{cases} \tag{11}$$

where

- r : Reward

- d_t : Distance at time t

- a_t : Action at time t

- α : Scaling coefficient (1)

- ϵ : Threshold for sparse reward (0.001)

2.1 From the literature

2.1.1 Dense rewards

$$r = -d_t^2 \tag{12}$$

$$r = -d_t \tag{13}$$

$$r = -\alpha d_t - \beta a^T a \tag{14}$$

$$r = -\alpha d_{t-1}^p - d_t^p \tag{15}$$

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

$$r = -d_t - \|a_{t-1}\| \tag{16}$$

Penalise large torque

$$r = -d_t^2 + \frac{d_{t-1} - d_t}{d_t} \tag{17}$$

2.1.2 Sparse rewards

$$r = \begin{cases} -1, & \text{if } d \geq \epsilon \\ 0, & \text{if } d < \epsilon \end{cases} \tag{18}$$

$$r = \begin{cases} 1, & \text{if } s \in G \\ 0, & \text{otherwise} \end{cases} \tag{19}$$

2.1.3 Dense + sparse rewards

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

$$r = \begin{cases} -1 - \beta \|a_{t-1}\|^2, & \text{if } d \geq \epsilon \\ 1 - \beta \|a_{t-1}\|^2, & \text{if } d < \epsilon \end{cases} \tag{21}$$

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

$$r = \begin{cases} -d_t, & \text{if } d \geq \epsilon \\ 1, & \text{if } d < \epsilon \end{cases} \tag{22}$$

$$r = \begin{cases} -0.02, & \text{if } d \geq \epsilon \\ 1, & \text{if } d < \epsilon \end{cases} \tag{23}$$

$$r = \begin{cases} \alpha(d_{t-1} - d_t), & \text{if } d \geq \epsilon \\ \alpha(d_{t-1} - d_t) + 10, & \text{if } d < \epsilon \end{cases} \tag{24}$$

$$r = \begin{cases} -0.001, & \text{if } d \geq \epsilon \\ 10, & \text{if } d < \epsilon \end{cases} \tag{25}$$

$$r = \begin{cases} -0.001, & \text{if } d \geq \epsilon \\ 10, & \text{if } d < \epsilon \end{cases} \tag{26}$$

Where s = state
G = set of goals

3 Action

3.1 Relative joint position

3.2 Absolute joint position

3.3 Relative joint torque

3.4 Absolute joint torque