LIDOSS Appendix

1 Size limit of g_{SDM}

Size of g_{SDM} is simply the number of subgoals contained in this set. This size is mostly regulated by pruning according to the *saliency check* (Algorithm 1, lines 19 to 21). A limit on this size is also used in Algorithm 1 as an additional criteria for regulation. The limit used in the experiments is 50, but it is rarely reached.

2 Generation of intermediate subgoal g^2

g^2 is a vector generated using a DDPG actor-critic network (details in sec. 4). The DDPG actor takes the environment state and the primary subgoal g^{HL} as inputs and provides $g^2 \in \mathbb{X}$ (same as $\mathbb{X}_{continuous}$) as output. The output layer of DDPG actor network is basically the subgoal vector g^2.

The primitive policy π can directly take g^{HL}. The intermediate subgoal g^2 is added as an option if multi-level decomposition is needed. In the experiments, 3-level hierarchy largely outperformed 2-level (without g^2) empirically. The authors of HAC also reported the same in their paper. So, using a 3-level hierarchy allows fare comparison with 3-level HAC.

The benefit of 3-level hierarchy may be because g^2 subgoals are intermediate landmarks to reach g^{HL}, therefore they can be achieved by π in shorter time-horizon. This means that the subgoal achievement rewards are more densely available for π. This is in contrast to a scenario where π tries to achieve the longer-horizon subgoal g^{HL} directly, which could be harder. Alternatively, we could use 2-level hierarchy and generate g^{HL} more frequently so that they are shorter-horizon subgoals for π. But this also doesn't work well because it reduces the advantage of temporal abstraction and longer-term commitment to a subgoal.

3 Subgoal achievement thresholds and rewards

The rewards for the intermediate and primitive levels are based on achieving the respective subgoals. For intermediate-level, this reward is equal to +1 if $\forall j : |s[j] - g^{HL}[j]| < subgoal\_threshold[j]$. Here, $s[j]$ is the value of state feature along dimension j and $g^{HL}[j]$ is the value of the subgoal state feature along

dimension j. Basically, if the state s reached by the agent is within a bounding box defined by the subgoal thresholds, then reward is +1, otherwise 0. Similarly, for the primitive level, reward is +1 if $\forall j : |s[j] - g^2[j]| < subgoal\_threshold[j]$, otherwise 0.

The subgoal thresholds are same for both intermediate and primitive levels. Using the definitions of *qpos* and *qvel* from section 4 of the paper: the subgoal thresholds for the Ant Four Rooms and Ant Maze domains are 0.4 along *qpos[1,2]*, 0.2 along *qpos[3]*, and 0.5 along *qvel[1,2]*. For Ant Reacher domain, the subgoal thresholds are 0.5 along *qpos[1,2]*, 0.2 along *qpos[3]*, and 0.5 along *qvel[1,2]*.