Supplementary Material for Dynamics-Aware Metric Embedding: Metric Learning in a Latent Space for Visual Planning

Mineui Hong[1], Kyungjae Lee[2], Minjae Kang[1], Wonsuhk Jung[3], and Songhwai Oh[1]

APPENDIX

In this manuscript, we present detailed proofs of the theorems in the paper.

I. Probabilistic reachability

In this section, we show that the proposed reachability measure can be estimated by training a conditional discriminator. Before we present the proofs, we first summarize the definition of the reachability measure we proposed. We first define a random walk as a stochastic process $\{\mathbf{X}_1^{\text{RW}}, \mathbf{X}_2^{\text{RW}}, \cdots\}$, such that,

$$\mathbf{X}_{t+1}^{\text{RW}} = f^{\text{trans}}(\mathbf{X}_t^{\text{RW}}, \mathbf{u}_t), \text{ where } \mathbf{u}_t \sim \text{Unif}(\mathcal{U}).$$

We then define the probabilistic reachability, $\rho_K^{\epsilon}(x'|x)$ for $x', x \in \mathcal{X}$ as,

$$\rho_K^{\epsilon}(x^j|x^i) = \sum_{k=1}^{K} P(\mathbf{X}_k^{\text{RW}} \in B_\epsilon(x^j) | \mathbf{X}_1^{\text{RW}} = x^i),$$
$$\text{where } B_\epsilon(x^j) := \{x \in \mathcal{X} \,\big|\, \|x - x^j\|_2 < \epsilon\}.$$

The proposed reachability measure can be estimated by training a conditional discriminator, $D(x'|x) : \mathcal{X} \times \mathcal{X} \to [0, 1]$, to minimize the following loss function:

$$\begin{aligned} L(x; D) := &\sum_{k=1}^{K} \mathbb{E}_{x' \sim p_{\mathbf{X}_k^{\text{RW}}|x}}\big[-\log(D(x'|x))\big] \\ &+ \mathbb{E}_{x' \sim p_X}\big[-\log(1 - D(x'|x)))\big], \end{aligned} \tag{1}$$

where $x' \sim p_{\mathbf{X}_k^{\text{RW}}|x}$ indicates that x' is sampled at the k-th step of a random walk starting from x and p_X denotes the prior distribution. As we denote the optimal discriminator which minimizes (1) by $D^*(x'|x)$, we can show the following proposition.

Proposition 1. *The optimal discriminator $D^*(x'|x)$, which minimizes (1), satisfies,*

$$D^*(x'|x) = \frac{\sum_{k=1}^{K} p_{\mathbf{X}_k^{RW}|x}(x')}{\sum_{k=1}^{K} p_{\mathbf{X}_k^{RW}|x}(x') + p_X(x')}.$$

Proof. By taking derivative of (1) with respect to $D(x'|x)$,

$$\frac{\partial L(x; D)}{\partial D(x'|x)} = -\sum_{k=1}^{K} \frac{p_{\mathbf{X}_k^{\text{RW}}|x}(x')}{D(x'|x)} + \frac{p_X(x')}{1 - D(x'|x)} \tag{2}$$

At $D(x'|x) = D^*(x'|x)$, the derivative in (2) becomes 0. Therefore,

$$D^*(x'|x) = \frac{\sum_{k=1}^{K} p_{\mathbf{X}_k^{\text{RW}}|x}(x')}{\sum_{k=1}^{K} p_{\mathbf{X}_k^{\text{RW}}|x}(x') + p_X(x')}.$$

□

Then as we let $\epsilon \ll 1$, the reachability $\rho_K^{\epsilon}(x'|x)$ is approximately proportional to $\sum_{k=1}^{K} p_{\mathbf{X}_k^{\text{RW}}|x}(x')$. Consequently, we can estimate the reachability as:

$$\rho_K^{\epsilon}(x'|x) \approx \frac{D^*(x'|x)}{1 - D^*(x'|x)} p_X(x').$$

II. Completeness proof of DAME planning

Algorithm 1 DAME Planning

> **Input:** current state, x_0, goal state, x_{goal}.
> **Initialize:** $t \leftarrow 0$
> **while** $x_t \notin B_\epsilon(x_{\text{goal}})$ **do**
> $\quad u_t \leftarrow \text{argmax}_{u \in \mathcal{U}}\ \rho_K^{\epsilon}(x_{\text{goal}} | f^{\text{trans}}(x_t, u))$
> $\quad x_{t+1} \leftarrow f^{\text{trans}}(x_t, u_t)$
> $\quad t \leftarrow t + 1$
> **end while**
> **Return:** $\{u_0, u_1, \cdots, u_{t-1}\}$

In this section, we theoretically verify the intuition that the proposed reachability represents the dynamical closeness between states, by showing that a planning algorithm (Algorithm 1) which maximizes reachability to the goal can always find a trajectory to reach the goal.

First, we define a goal-conditioned greedy policy $\pi^* : \mathcal{X} \times \mathcal{X} \to \mathcal{U}$, which takes a control signal maximizing the reachability to the goal at each time step:

$$\begin{aligned} &\pi^*(x_t | x_{\text{goal}}) := \underset{u_t \in \mathcal{U}}{\text{argmax}}\ \rho_K^{\epsilon}(x_{\text{goal}} | x_{t+1}), \\ &\text{subject to } x_{t+1} = f^{\text{trans}}(x_t, u_t). \end{aligned} \tag{3}$$

We can now start our investigation with two assumptions on the smoothness and controllability of f^{trans}:

[1]M. Hong, M. Kang, and S. Oh are with the Department of Electrical and Computer Engineering and ASRI, Seoul National University, Seoul 08826, Republic of Korea (e-mails: {mineui.hong, minjae.kang}@rllab.snu.ac.kr, songhwai@snu.ac.kr, corresponding author: S. Oh).

[2]K. Lee is with the Department of Artificial Intelligence, Chung-Ang University, Seoul 06974, Republic of Korea (e-mail: kyungjae.lee@ai.cau.ac.kr).

[3]W. Jung is with the Department of Mechanical Engineering and Interdisciplinary Major in Artificial Intelligence, Seoul National University, Seoul 08826, Republic of Korea (e-mail: smw04017@snu.ac.kr).

Assumption 1. *The forward dynamics function, $f^{trans}: \mathcal{X} \times \mathcal{U} \to \mathcal{X}$, is Lipschitz continuous on $\mathcal{X}$ and $\mathcal{U}$, i.e.,*

$$\begin{aligned}
&\exists_{\kappa_x,\kappa_u>0}, \text{ such that } \forall_{x,\tilde{x}\in\mathcal{X},\ u,\tilde{u}\in\mathcal{U}},\\
&||f^{\text{trans}}(\tilde{x},u) - f^{\text{trans}}(x,u)||_2 < \kappa_x||\tilde{x}-x||_2,\\
&||f^{\text{trans}}(x,\tilde{u}) - f^{\text{trans}}(x,u)||_2 < \kappa_u||\tilde{u}-u||_2.
\end{aligned}$$

Assumption 2. *The given goal state x_{goal} is reachable in K-step from all $x \in \mathcal{X}$, and an agent stays in x_{goal} after it reaches x_{goal}, i.e.,*

$$\begin{aligned}
&\forall_{x_0\in\mathcal{X}}, \exists_{u_0,\cdots,u_{K-1}\in\mathcal{U}}, \text{ such that } x_K = x_{\text{goal}},\\
&\text{where } x_{k+1} = f^{\text{trans}}(x_k,u_k),\ \forall_{k=0,\cdots,K-1}.
\end{aligned}$$

Using the definition of π^* and Assumption 1 and 2, we can now show the following lemma:

Lemma 1. *Consider π^* defined in (3) and let $x_{t+1} = f^{trans}(x_t, \pi^*(x_t|x_{goal}))$, where f^{trans} is a forward dynamics model and satisfies Assumption 1 and 2. Then the following inequality is satisfied, unless $x \in B_\epsilon(x_{goal})$.*

$$\begin{aligned}
&\rho_K^\epsilon(x_{goal}|x_{t+1})\\
&> \rho_K^\epsilon(x_{goal}|x_t) + \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1},
\end{aligned}$$

where $V_d(r)$ denotes the volume of a d-dimensional ball with radius r.

Proof. We first show that,

$$\begin{aligned}
&\forall x\in\mathcal{X},\ P(\mathbf{X}_K^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_1^{\text{RW}} = x)\\
&> \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1}.
\end{aligned}$$

With Assumption 2, we let $u^*_{1:K-1} := \{u^*_1, \cdots, u^*_{K-1}\}$ be a sequence of controls, such that, $x^*_K = x_{\text{goal}}$, where $x^*_1 = x$ and $x^*_{k+1} = f^{\text{trans}}(x^*_k, u^*_k)$, for $k = 1, \cdots, K-1$. Then consider a new sequence of controls $\{\tilde{u}_1, \cdots, \tilde{u}_{K-1}\}$, such that,

$$||\tilde{u}_k - u^*_k||_2 < \frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)},$$

and the corresponding sequence of states, $\{\tilde{x}_1 = x, \cdots, \tilde{x}_K\}$. Then by Assumption 1,

$$\begin{aligned}
&||\tilde{x}_2 - x^*_2||_2 < \kappa_u||\tilde{u}_1 - u^*_1||_2 < \frac{(\kappa_x-1)\epsilon}{\kappa_x^{K-1}-1},\\
&||\tilde{x}_3 - x^*_3||_2 < \kappa_u||\tilde{u}_2 - u^*_2||_2 + \kappa_x||\tilde{x}_2 - x^*_2||_2\\
&< \frac{(\kappa_x-1)\epsilon}{\kappa_x^{K-1}-1}(1+\kappa_x)\\
&\cdots\\
&||\tilde{x}_K - x^*_K||_2 = ||\tilde{x}_K - x_{\text{goal}}||_2 < \frac{(\kappa_x-1)\epsilon}{\kappa_x^{K-1}-1}\sum_{h=0}^{K-2}\kappa_x^h = \epsilon,
\end{aligned}$$

which indicates that, by sampling each $\tilde{u}_k$ from $\left\{u\in\mathcal{U}\,\middle|\, ||u-u^*_k||_2 < \frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right\}$, we can let $\tilde{x}_K \in B_\epsilon(x_{\text{goal}})$. Therefore, we find that,

$$\begin{aligned}
&P(\mathbf{X}_K^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_1^{\text{RW}} = x)\\
&> \int_{B_\delta(u^*_{K-1})}\cdots\int_{B_\delta(u^*_1)} \frac{1}{|\mathcal{U}|^K} du_1\cdots du_{K-1}\\
&\geq \frac{1}{2^{d_u(K-1)}}\left(\frac{V_{d_u}(\delta)}{2^{d_u}}\right)^{K-1}\\
&= \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1},
\end{aligned}$$

where

$$\delta = \frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)},\ B_\delta(u^*_k) := \{u\in\mathcal{U}\ |\ ||u-u^*_k|| < \delta\}.$$

Now, considering $x \in \mathcal{X} - B_\epsilon(x_{\text{goal}})$ and $x' = f^{\text{trans}}(x, \pi^*(x|x_{\text{goal}}))$, we show that,

$$\begin{aligned}
&\rho_K^\epsilon(x_{\text{goal}}|x')\\
&> \rho_K^\epsilon(x_{\text{goal}}|x) + \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1},
\end{aligned}$$

as following:

$$\begin{aligned}
&\rho_K^\epsilon(x_{\text{goal}}|x) + \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1}\\
&= \sum_{k=2}^{K} P(\mathbf{X}_k^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_1^{\text{RW}} = x)\\
&\quad + \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1}\\
&\Big(\because P(\mathbf{X}_1^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_1^{\text{RW}} = x) = 0\Big)\\
&= \int_{\mathcal{U}} \frac{1}{|\mathcal{U}|}\Big[\sum_{k=2}^{K} P(\mathbf{X}_k^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_2^{\text{RW}} = f^{\text{trans}}(x,u))\\
&\quad + \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1}\Big]du\\
&< \int_{\mathcal{U}} \frac{1}{|\mathcal{U}|}\left(\sum_{k=2}^{K+1} P(\mathbf{X}_k^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_2^{\text{RW}} = f^{\text{trans}}(x,u))\right)du\\
&\Big(\because \forall u\in\mathcal{U},\ P(\mathbf{X}_{K+1}^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_2^{\text{RW}} = f^{\text{trans}}(x,u))\\
&\quad > \frac{1}{2^{2d_u(K-1)}}\left\{V_{d_u}\left(\frac{(\kappa_x-1)\epsilon}{\kappa_u(\kappa_x^{K-1}-1)}\right)\right\}^{K-1}\Big)\\
&\leq \sum_{k=1}^{K} P\Big(\mathbf{X}_k^{\text{RW}} \in B_\epsilon(x_{\text{goal}})|\mathbf{X}_1^{\text{RW}} = f^{\text{trans}}(x,\pi^*(x|x_{\text{goal}}))\Big)\\
&= \rho_K^\epsilon(x_{\text{goal}}|x')
\end{aligned}$$

□

From Lemma 1, we can finally show that Algorithm 1 is complete, i.e., greedy planning based on the reachability reaches the goal state within a finite time.

Theorem 2. *(Completeness of Algorithm 1) Consider a forward dynamics model f^{trans} and a given goal state $x_{goal} \in \mathcal{X}$, which satisfy Assumption 1 and 2, respectively. Then an agent controlled by π^* will reach $B_\epsilon(x_{goal})$ in at most t_{max} of time steps, where*

$$t_{max} = 2^{2d_u(K-1)} K \left\{ V_{d_u} \left(\frac{(\kappa_x - 1)\epsilon}{\kappa_u(\kappa_x^{K-1} - 1)} \right) \right\}^{-(K-1)}.$$

Proof. Note that,

$$\begin{aligned} &\forall x \in \mathcal{X} - B_\epsilon(x_{\text{goal}}), \\ &\rho_K^\epsilon(x_{\text{goal}}|x) = \sum_{k=1}^{K} P(\mathbf{X}_k^{\text{RW}} \in B_\epsilon(x_{\text{goal}}) | \mathbf{X}_1^{\text{RW}} = x) \leq K, \end{aligned} \tag{4}$$

since, $P(\mathbf{X}_k^{\text{RW}} \in B_\epsilon(x_{\text{goal}}) | \mathbf{X}_1^{\text{RW}} = x) \leq 1$.

Assuming that, Algorithm 1 is not finished until $t = 2^{2d_u(K-1)} K \left\{ V_{d_u} \left(\frac{(\kappa_x - 1)\epsilon}{\kappa_u(\kappa_x^{K-1} - 1)} \right) \right\}^{-(K-1)}$, then by Lemma 1,

$$\begin{aligned} &\rho_K^\epsilon(x_{\text{goal}}|x_t) \\ &> t \times \frac{1}{2^{2d_u(K-1)}} \left\{ V_{d_u} \left(\frac{(\kappa_x - 1)\epsilon}{\kappa_u(\kappa_x^{K-1} - 1)} \right) \right\}^{K-1} = K, \end{aligned}$$

which is a contradiction to (4). Therefore, Algorithm 1 must be finished, before the maximum time step,

$$t_{\text{max}} = 2^{2d_u(K-1)} K \left\{ V_{d_u} \left(\frac{(\kappa_x - 1)\epsilon}{\kappa_u(\kappa_x^{K-1} - 1)} \right) \right\}^{-(K-1)}.$$

$\square$