Truncated Diffusion Process for Temporal Consistent Inference

Qingyuan Jiang[1] and Volkan Isler[1,2]

I. APPENDIX

We provide extra details on the theoretical analysis of the truncated diffusion process in Sec. IV-B, including the definition of the Markov kernel for the truncated diffusion process, and the proof of the KL contractivity.

A. Definition details

1) Markov Kernel Definition: In Sec. IV-B, we define the k-step truncated diffusion process as a Markov kernel $K_{\mathbf{c}}^{(k)}$. Here we provide the detailed math process. Given a sample $\mathbf{x}$ and a new condition $\mathbf{c}$, the probability of generating a new sample $\mathbf{x}'$ can be calculated by:

$$\begin{aligned} K_{\mathbf{c}}^{(k)} &:= p_\theta(\mathbf{x}'|\mathbf{x},\mathbf{c}) \\ &= \int_{\mathbf{x}^k} p_\theta(\mathbf{x}',\mathbf{x}^k|\mathbf{x},\mathbf{c})\,\mathrm{d}\mathbf{x}^k \\ &= \int_{\mathbf{x}^k} p_\theta(\mathbf{x}'|\mathbf{x},\mathbf{x}^k,\mathbf{c})p(\mathbf{x}^k|\mathbf{x},\mathbf{c})\,\mathrm{d}\mathbf{x}^k \\ &= \int_{\mathbf{x}^k} p_\theta(\mathbf{x}'|\mathbf{x}^k,\mathbf{c})q(\mathbf{x}^k|\mathbf{x})\,\mathrm{d}\mathbf{x}^k \end{aligned} \tag{1}$$

In the forth row, we remove the dependency of $\mathbf{x}$ in $p_\theta(\mathbf{x}'|\mathbf{x},\mathbf{x}^k,\mathbf{c})$ because the denoising process only depends on the noised sample $\mathbf{x}^k$ and the condition $\mathbf{c}$. Similarly, we remove the dependency of $\mathbf{c}$ in $p(\mathbf{x}^k|\mathbf{x},\mathbf{c})$ because the forward diffusion process is independent of the condition, and we replace the notation with $q(\mathbf{x}^k|\mathbf{x})$ to follow the convention of diffusion models [5].

2) Output Distribution Definition: Given the Markov kernel defined above, we can calculate the output distribution from the truncated diffusion process by applying this kernel to the existing arbitrary distribution $P(\mathbf{x})$.

$$\begin{aligned} Q_\theta(\mathbf{x}'|\mathbf{c}) &:= \int_{\mathbf{x}} p(\mathbf{x}',\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \quad \text{(marginalization)} \\ &= \int_{\mathbf{x}} p(\mathbf{x}'|\mathbf{x},\mathbf{c})p(\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \quad \text{(chain rule)} \\ &= \int_{\mathbf{x}} p(\mathbf{x}'|\mathbf{x},\mathbf{c})P(\mathbf{x})\,\mathrm{d}\mathbf{x} \quad \text{(Definition of } P(\mathbf{x})\text{)} \end{aligned} \tag{2}$$

Note that in the third row, we replace the notation of $p(\mathbf{x}|\mathbf{c})$ with $P(\mathbf{x})$ because the input distribution is independent of the condition $\mathbf{c}$, and we use $P(\mathbf{x})$ to follow the convention of probability distributions.

[1] Department of Computer Science & Engineering, University of Minnesota, 4-192 Keller Hall, 200 Union Street SE, Minneapolis, MN 55455, US

[2] Department of Computer Science, University of Texas at Austin, 2317 Speedway, Austin, Texas 78712, US

3) Identity Transition: Here we prove the identity transition property of the Markov kernel is an identity mapping when sampling under the same condition, i.e., $K_{\mathbf{c}}^{(k)} \circ p(\mathbf{x}|\mathbf{c}) = p(\mathbf{x}|\mathbf{c})$.

$$\begin{aligned} K_{\mathbf{c}}^{(k)} \circ p(\mathbf{x}|\mathbf{c}) &:= \int_{\mathbf{x}} p(\mathbf{x}'|\mathbf{x},\mathbf{c})p(\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \\ &= \int_{\mathbf{x}} p(\mathbf{x}',\mathbf{x}|\mathbf{c})\,\mathrm{d}\mathbf{x} \quad \text{(chain rule)} \\ &= p(\mathbf{x}'|\mathbf{c}) \quad \text{(marginalization)} \end{aligned} \tag{3}$$

B. KL Contractivity Proof

In this section, we provide a unconditional proof of the KL contractivity that's used in **Lemma 3** in Sec. IV-B.

Proposition. (Data Processing Inequality for KL Divergence [1]): Let $K(x'|x)$ be any Markov kernel, then for any distributions $p(x)$ and $q(x)$, we have:

$$D_{\mathrm{KL}}\left(K \circ p \;||\; K \circ q\right) \le D_{\mathrm{KL}}\left(p \;||\; q\right) \tag{4}$$

Proof. Define the push-forward distribution through the kernel on x' as, $\tilde{p}(x') = \int K(x'|x)p(x)\,\mathrm{d}x$, and similarly $\tilde{q}(x') = \int K(x'|x)q(x)\,\mathrm{d}x$. we introduce the joint distributions $P(x,x') = p(x)K(x'|x)$ and $Q(x,x') = q(x)K(x'|x)$.

$$\begin{aligned} P(x,x') &= p(x)K(x'|x) = \tilde{p}(x')p(x|x') \\ Q(x,x') &= q(x)K(x'|x) = \tilde{q}(x')q(x|x') \end{aligned} \tag{5}$$

The KL divergence between the joint distributions is:

$$\begin{aligned} &D_{\mathrm{KL}}\left(P(x,x') \;||\; Q(x,x')\right) \\ &= \int_x \int_{x'} P(x,x') \log \frac{P(x,x')}{Q(x,x')}\,\mathrm{d}x'\,\mathrm{d}x \\ &= \int_x \int_{x'} P(x,x') \log \frac{\tilde{p}(x')p(x|x')}{\tilde{q}(x')q(x|x')}\,\mathrm{d}x'\,\mathrm{d}x \\ &= \int_x \int_{x'} P(x,x') \log \frac{\tilde{p}(x')}{\tilde{q}(x')}\,\mathrm{d}x'\,\mathrm{d}x \quad + \\ &\quad \int_x \int_{x'} P(x,x') \log \frac{p(x|x')}{q(x|x')}\,\mathrm{d}x'\,\mathrm{d}x \\ &= T_1 + T_2 \end{aligned} \tag{6}$$

$$\begin{aligned} T_1 &= \int_x \int_{x'} P(x,x') \log \frac{\tilde{p}(x')}{\tilde{q}(x')}\,\mathrm{d}x'\,\mathrm{d}x \\ &= \int_{x'} \left(\int_x P(x,x')\,\mathrm{d}x \right) \log \frac{\tilde{p}(x')}{\tilde{q}(x')}\,\mathrm{d}x' \\ &= \int_{x'} \tilde{p}(x') \log \frac{\tilde{p}(x')}{\tilde{q}(x')}\,\mathrm{d}x' \\ &= D_{\mathrm{KL}}\left(\tilde{p} \;||\; \tilde{q}\right) \end{aligned} \tag{7}$$

$$
\begin{aligned}
T_2 &= \int_x \int_{x'} P(x, x') \log \frac{p(x|x')}{q(x|x')} \, \mathrm{d}x' \, \mathrm{d}x \\
&= \int_{x'} \left(\int_x P(x, x') \log \frac{p(x|x')}{q(x|x')} \, \mathrm{d}x \right) \mathrm{d}x' \\
&= \int_{x'} \left(\int_x \tilde{p}(x') p(x|x') \log \frac{p(x|x')}{q(x|x')} \, \mathrm{d}x \right) \mathrm{d}x' \\
&= \int_{x'} \tilde{p}(x') \left(\int_x p(x|x') \log \frac{p(x|x')}{q(x|x')} \, \mathrm{d}x \right) \mathrm{d}x' \\
&= \int_{x'} \tilde{p}(x') D_{\mathrm{KL}} \left(p(x|x') \,||\, q(x|x') \right) \mathrm{d}x' \\
&= \mathbb{E}_{x' \sim \tilde{p}} \left[D_{\mathrm{KL}} \left(p(x|x') \,||\, q(x|x') \right) \right]
\end{aligned}
\tag{8}
$$

Combining the two terms, we have:

$$
\begin{aligned}
&D_{\mathrm{KL}} \left(P(x, x') \,||\, Q(x, x') \right) = \\
&\quad D_{\mathrm{KL}} \left(\tilde{p} \,||\, \tilde{q} \right) + \mathbb{E}_{x' \sim \tilde{p}} \left[D_{\mathrm{KL}} \left(p(x|x') \,||\, q(x|x') \right) \right]
\end{aligned}
\tag{9}
$$

Since the KL divergence term $D_{\mathrm{KL}} \left(p(x|x') \,||\, q(x|x') \right)$ is non-negative, its expectation is also non-negative, i.e. $\mathbb{E}_{x' \sim \tilde{p}} \left[D_{\mathrm{KL}} \left(p(x|x') \,||\, q(x|x') \right) \right] \geq 0$, we have:

$$
D_{\mathrm{KL}} \left(\tilde{p} \,||\, \tilde{q} \right) \leq D_{\mathrm{KL}} \left(P(x, x') \,||\, Q(x, x') \right) \tag{10}
$$

Note that by definition of the joint distributions, we have:

$$
\begin{aligned}
&D_{\mathrm{KL}} \left(P(x, x') \,||\, Q(x, x') \right) \\
&= \int_x \int_{x'} p(x) K(x'|x) \log \frac{p(x) K(x'|x)}{q(x) K(x'|x)} \, \mathrm{d}x' \, \mathrm{d}x \\
&= \int_x \int_{x'} p(x) K(x'|x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x' \, \mathrm{d}x \\
&= \int_x \left(\int_{x'} K(x'|x) \, \mathrm{d}x' \right) p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x \\
&= \int_x p(x) \log \frac{p(x)}{q(x)} \, \mathrm{d}x \\
&= D_{\mathrm{KL}} \left(p \,||\, q \right)
\end{aligned}
\tag{11}
$$

Therefore:

$$
D_{\mathrm{KL}} \left(K \circ p \,||\, K \circ q \right) \leq D_{\mathrm{KL}} \left(p \,||\, q \right) \quad \blacksquare \tag{12}
$$

C. Qualitative Results

Here we provide more qualitative results of the truncated diffusion process for human motion prediction.

REFERENCES

[1] Thomas M. Cover and Joy A. Thomas. *Elements of Information Theory 2nd Edition (Wiley Series in Telecommunications and Signal Processing)*. Wiley-Interscience, July 2006. ISBN 0-471-24195-4. Published: Hardcover.

[2] Siyuan Huang, Zan Wang, Puhao Li, Baoxiong Jia, Tengyu Liu, Yixin Zhu, Wei Liang, and Song-Chun Zhu. Diffusion-based generation, optimization, and planning in 3d scenes. In *Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition*, pages 16750–16761, 2023.

[3] Qingyuan Jiang, Burak Susam, Jun-Jee Chao, and Volkan Isler. Map-Aware Human Pose Prediction for Robot Follow-Ahead, March 2024. URL http://arxiv.org/abs/2403.13294. arXiv:2403.13294 [cs].

[4] Mohammad Mahdavian, Payam Nikdel, Mahdi TaherAhmadi, and Mo Chen. STPOTR: Simultaneous Human Trajectory and Pose Prediction Using a Non-Autoregressive Transformer for Robot Following Ahead, September 2022. URL http://arxiv.org/abs/2209.07600. arXiv:2209.07600 [cs].

[5] Kihyuk Sohn, Honglak Lee, and Xinchen Yan. Learning Structured Output Representation using Deep Conditional Generative Models. In *Advances in Neural Information Processing Systems*, volume 28. Curran Associates, Inc., 2015. URL https://papers.nips.cc/paper_files/paper/2015/hash/8d55a249e6baa5c06772297520da2051-Abstract.html.

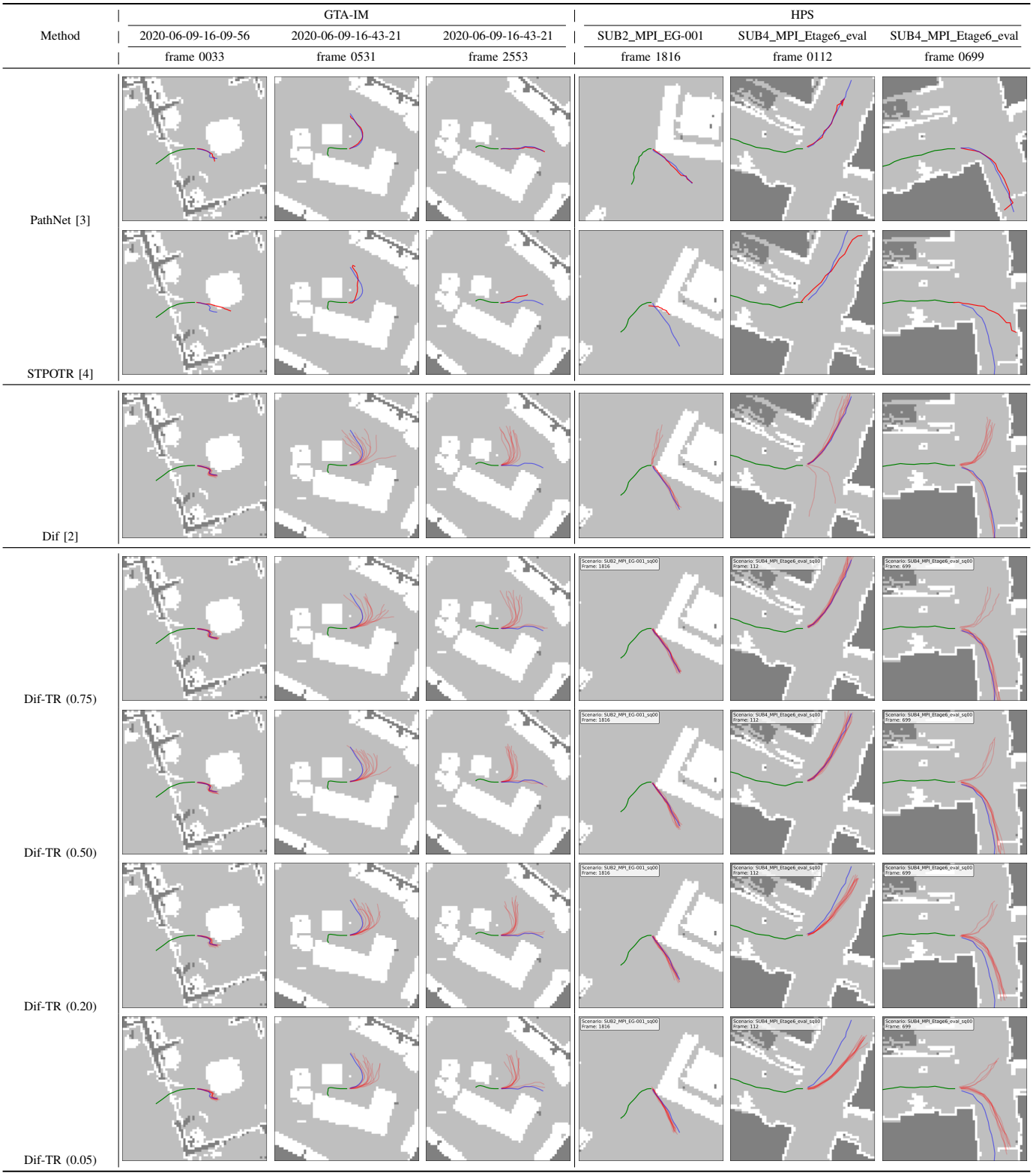

Fig. 1: **Qualitative results for human motion prediction**. We plot the human motion history in green, the predicted human motion in red, and the ground truth in blue. The local occupancy map is visualized by: white - obstacles, grey - free space, dark - unknown. Each row is a method, and each column is a selected sample.