Visual Geometric Skill Inference by Watching Human Demonstration: Supplementary Materials

Jun Jin[1], Masood Dehghan[1], Martin Jagersand[1]

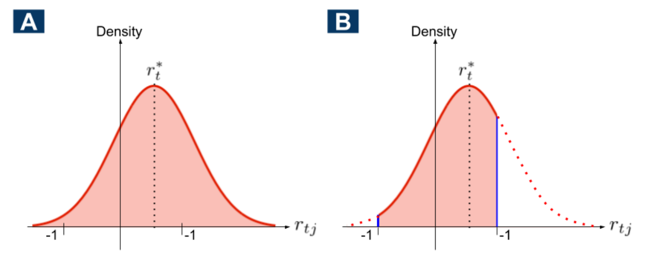

Fig. 1: The partition function $\mathcal{Z}_t$ is the expectation of all possible $\{r_{tj}\}$ when at state s_t. r_{tj} follows a normal distribution parameterized by r_t^*. **A** shows a regular normal distribution. **B** shows a truncated normal distribution.

A. Conditions when the cost function is a constant

We prove that when $p(r_{tj})$ is a regular normal distribution with doman $[-\infty, \infty]$, the cost function Eq. (11) in our paper is a constant which is related to human factor[1] σ_0^2.

Firstly, let's review the cost function in Eq. (11):

$$\mathcal{L} = \arg\max_{\theta} \sum r_t^* - log\mathcal{Z}_t \tag{1}$$

, where $\mathcal{Z}_t$ is the partition function that integrates the exponential reward r_{tj} of all possible actions $\{a_{tj}\}$ when human demonstrator is at state s_t. Since human demonstrator makes selections only from promising actions instead of any uniform actions, we assume $r_{tj} \sim \mathcal{N}(r_t^*, \sigma_0)$, where r_t^* is reward from the selected action a_t^* that is observed in the demonstration. So $\mathcal{Z}_t$ can be written as:

$$\mathcal{Z}_t = \mathbb{E}_{p(r_{tj};r_t^*)}[exp(r_{tj})] \tag{2}$$

Considering a $[-\infty, \infty]$ domain of r_{tj}, we have:

$$\mathcal{Z}_t = \int_{-\infty}^{\infty} exp(r_{tj})p(r_{tj})dr_{tj} \tag{3}$$

where $p(r_{tj}) = \mathcal{N}(r_{tj}|r_t^*, \sigma_0)$, Eq. (3) can be rewritten as:

$$\begin{aligned}\mathcal{Z}_t = \frac{1}{\sqrt{2\pi}\sigma_0} \int_{-\infty}^{\infty} exp(&- \frac{1}{2\sigma_0^2} r_{tj}^2 + (\frac{r_t^*}{\sigma_0^2} + 1)r_{tj} \\ &- \frac{1}{2\sigma_0^2} r_t^{*2})dr_{tj}\end{aligned} \tag{4}$$

Now $\mathcal{Z}_t$ has a standard form as a Gaussian integral, which is tractable in practice[1]:

$$\int_{-\infty}^{\infty} k\, exp(-fx^2 + gx + h)dx = k\sqrt{\frac{\pi}{f}} exp(\frac{g^2}{4f} + h) \tag{5}$$

So, we have:

$$\mathcal{Z}_t = exp(r_t^* + \frac{\sigma_0^2}{2}) \tag{6}$$

As a result, r_t^* is neutralized in the cost function, Eq. (1) can be rewritten as:

$$\mathcal{L} = \arg\max_{\theta} \sum -\frac{\sigma_0^2}{2} \tag{7}$$

which is now a constant related to human factor σ_0.

B. Cost function with truncated normal distribution

We empirically calculate the cost values given different r_t* and σ_0 (Fig. 2.). A Monte Carlo estimator with a sampling size=2000 is used for computation. Results show the cost value overall increases as r_t^* grows, however the slope is different. Lower σ_0 outputs a smaller gradient for learning the reward function while higher σ_0 outputs a larger one.

Intuitively, a lower σ_0 means human demonstrator is more confident in selecting actions, which will result the learned reward function easily over-fit to observed demonstrations. On the other side, a higher σ_0 means human demonstrator is not so confident in demonstration. So the demonstration samples have more randomness compared to smaller σ_0 demonstrations. Any updates in the resulting r_t^* should have more value in learning.

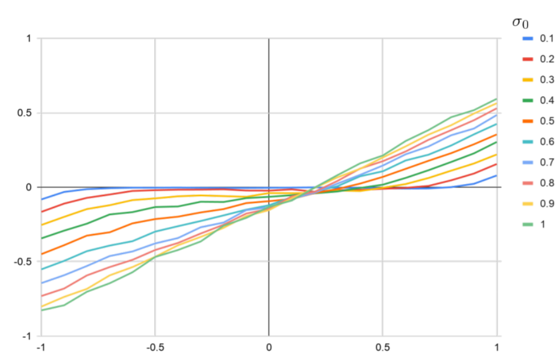

Fig. 2: Cost function values with different σ_0 and r_t^*.

[1]Authors are with Department of Computing Science, University of Alberta, Edmonton AB., Canada, T6G 2E8 `jjin5, masood1, martin.jagersand @ualberta.ca`

[1]σ_0^2 is determined by human demonstrator's confidence level α.

REFERENCES

[1] Wikipedia contributors, "Gaussian functions - integral of a gaussian function," 2019, [Online; accessed 5-Sept-2019]. [Online]. Available: https://en.wikipedia.org/wiki/Gaussian_function