A Proof of Proposition 4.1

Proof. Taking $h(t) = f''(t)$, $a = f'(0)$ and $b = f(0)$ in Eq (5), we have

$$
(f'(0)t + f(0)) + \int_0^{\infty} (t - \mu)_+ h(\mu) d\mu
$$

= $(f'(0)t + f(0)) + \int_0^t (t - \mu) f''(\mu) d\mu$
= $(f'(0)t + f(0)) + (t - \mu) f'(\mu) \Big|_{\mu=0}^t + \int_0^t f'(\mu) d\mu$ (integration by parts)
= $f(0) + \int_0^t f'(\mu) d\mu$
= $f(t)$.

Conversely, if $f(t) = (at + b) + \int_0^\infty (t - \mu)_+ h(\mu) d\mu$, calculation shows

$$
f'(t) = a + \int_0^t h(\mu)d\mu
$$
, $f''(t) = h(t)$.

Therefore, f is convex if h is non-negative.

To prove Eq. (6), we substitute $f(t) = f'(0)t + f(0) + \int_0^\infty (t - \mu)_+ f''(\mu) d\mu$ into the definition of f-divergence,

$$
D_f(p || q) = \mathbb{E}_q \left[f\left(\frac{p(x)}{q(x)}\right) - f(1) \right]
$$

= $\mathbb{E}_q \left[f'(0) \frac{p(x)}{q(x)} + f(0) + \int_0^\infty (p(x)/q(x) - \mu)_+ f''(\mu) d\mu - f(1) \right]$
= $[f'(0) + f(0) - f(1)] + \int_0^\infty \mathbb{E}_q \left[\left(\frac{p(x)}{q(x)} - \mu\right)_+ \right] f''(\mu) d\mu.$

This completes the proof.

$$
\Box
$$

B Proof of Proposition 4.2

Proof. By chain rule and the "score-function trick" $\nabla_{\theta} q_{\theta}(x) = q_{\theta}(x) \nabla_{\theta} \log q_{\theta}(x)$, we have

$$
\nabla_{\theta} D_{f}(p \mid q_{\theta}) = \mathbb{E}_{q_{\theta}} \left[\nabla_{\theta} f \left(\frac{p(x)}{q_{\theta}(x)} \right) + f \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} \log q_{\theta}(x) \right]
$$

\n
$$
= \mathbb{E}_{q_{\theta}} \left[f' \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} \left(\frac{p(x)}{q_{\theta}(x)} \right) + f \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} \log q_{\theta}(x) \right]
$$

\n
$$
= \mathbb{E}_{q_{\theta}} \left[-f' \left(\frac{p(x)}{q_{\theta}(x)} \right) \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} \log q_{\theta}(x) + f \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} \log q_{\theta}(x) \right]
$$

\n
$$
= -\mathbb{E}_{q_{\theta}} \left[\rho_{f} \left(\frac{p(x)}{q_{\theta}(x)} \right) \log q_{\theta}(x) \right],
$$

where $\rho_f(t) = f'(t)t - f(t)$. This proves Eq. (7).

To prove Eq. (8), we note that for any function ϕ , we have by the *reparamertization trick*:

$$
\nabla_{\theta} \mathbb{E}_{q_{\theta}}[\phi(x)] = \mathbb{E}_{x \sim q_{\theta}}[\phi(x) \nabla_{\theta} \log q_{\theta}(x)] \quad \text{(score function)}
$$

=
$$
\mathbb{E}_{\xi \sim q_{0}}[\nabla_{x} \phi(x) \nabla_{\theta} g_{\theta}(\xi)] \quad \text{(reparameterization trick)},
$$

where we assume $x \sim q_\theta$ is generated by $x = g_\theta(\xi)$, $\xi \sim q_0$.

Taking $\phi(x) = \rho_f(p(x)/q_\theta(x))$ in Eq. (7), we have

$$
\nabla_{\theta} D_{f}(p || q_{\theta}) = -\mathbb{E}_{x \sim q_{\theta}} \left[\rho_{f} \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} \log q_{\theta}(x) \right]
$$

\n
$$
= -\mathbb{E}_{\xi \sim q_{0}} \left[\nabla_{x} \rho_{f} \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} g_{\theta}(\xi) \right]
$$

\n
$$
= -\mathbb{E}_{\xi \sim q_{0}} \left[\rho_{f}' \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{x} \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} g_{\theta}(\xi) \right]
$$

\n
$$
= -\mathbb{E}_{\xi \sim q_{0}} \left[\rho_{f}' \left(\frac{p(x)}{q_{\theta}(x)} \right) \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{x} \log \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} g_{\theta}(\xi) \right]
$$

\n
$$
= -\mathbb{E}_{\xi \sim q_{0}} \left[\gamma_{f} \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{x} \log \left(\frac{p(x)}{q_{\theta}(x)} \right) \nabla_{\theta} g_{\theta}(\xi) \right],
$$

where $\gamma_f(t) = \rho'_f(t)t$.

 \Box

C Tail-adaptive f-divergence with Score-Function Gradient Estimator

Algorithm [2](#page-1-0) summarizes our method using the score-function gradient estimator (7).

Algorithm 2 Variational Inference with Tail-adaptive f-Divergence (with Score Function Gradient) Goal: Find the best approximation of $p(x)$ from $\{q_\theta : \theta \in \Theta\}.$ Initialize θ , set an index β (e.g., $\beta = -1$). for iteration do Draw $\{x_i\}_{i=1}^n \sim q_\theta$. Set $\hat{F}(t) = \sum_{j=1}^n \mathbb{I}(p(x_j)/q(x_j) \ge t)/n$, and $\rho_i = \hat{F}(p(x_i)/q(x_i))^{\beta}$. Update $\theta \leftarrow \theta + \epsilon \Delta \theta$, where ϵ is stepsize, and $\frac{1}{\sqrt{n}}$

$$
\Delta \theta = \frac{1}{z_{\rho}} \sum_{i=1} \left[\rho_i \nabla_{\theta} \log q_{\theta}(x_i) \right],
$$

where $z_{\rho} = \sum_{i=1}^{n} \rho_i$. end for

D More Results for Bayesian Neural Network

Table [2](#page-2-0) shows more results in Bayesian networks with more choices of α in α -divergence. We can see that our approach achieves the best performance in most of the cases.

Table 2: Test RMSE and LL results for Bayesian neural network regression.

E Reinforcement Learning

In this section, we provide more information and results of the Reinforcement learning experiments, including comparisons of algorithms using score-function gradient estimators (Algorithm [2\)](#page-1-0).

E.1 MuJoCo Environments

We test six MuJoCo environments in this paper: *HalfCheetah*, *Hopper*, *Swimmer(rllab)*, *Humanoid(rllab)*, *Walker*, and *Ant*, for which the dimensions of the action space are 6, 3, 2, 21, 6, 8, respectively. Figure [4](#page-2-1) shows examples of the environment used in our experiments.

Figure 4: MuJoCo environments used in our reinforcement learning experiments. From left to right: HalfCheetah, Hopper, Swimmer(rllab), Humanoid(rllab), Walker, and Ant.

E.2 Different Choices of α

In this section, we present the average reward of α -divergences with different choices of α on Hopper and Walker with both score-function and reparameterization gradient estimators. We can see from Figure [5](#page-3-0) that $\alpha = 0.5$ and $\alpha = +\infty$ (denoted by $\alpha = \max$ in the legends) perform consistently better than standard KL divergence ($\alpha = 0$), which is used the original SAC paper.

E.3 Tail-adaptive f -divergence with score function estimation

In this section, we investigate optimization with score function gradient estimators (Algorithm [2\)](#page-1-0). The results in Figure [6](#page-3-1) show that our tail-adaptive f-divergence tends to yield better performance across all environments tested.

Figure 6: Results of average rewards with score function gradients.