PathFlow: A Normalizing Flow Generator that Finds Transition Paths
Supplementary Material

Tianyi Liu[1] **Weihao Gao**[1] **Zhirui Wang**[1] **Chong Wang**[1]

[1]ByteDance Inc.

A PROOF OF THEOREM 4.1

For any function $F \in \{M_{ij}, \nabla_i U, \nabla_l M_{ij}, \nabla^2_{ij} U\}$, the error of the estimator can be decomposed by triangle inequality as follows,

$$|F^{(T,k)}(\boldsymbol{z}) - F(\boldsymbol{z})| \le |F^{(k)}(\boldsymbol{z}) - F(\boldsymbol{z})| + |F^{(T,k)}(\boldsymbol{z}) - F^{(k)}(\boldsymbol{z})|.$$

The first term is introduced by the finiteness of k. The second term is introduced by finiteness of the time T of restrained dynamics. The following two lemmas provide upper bounds for finite-k error and finite-T error respectively.

Lemma A.1 (Error by finite k)**.** *For any function $g(\boldsymbol{r})$, consider the two functionals $\mathcal{I}[g]$ and $\mathcal{I}^{(k)}[g]$ defined as follows*

$$\mathcal{I}[g](\boldsymbol{z}) = \int_{\mathbb{R}^{3D}} g(\boldsymbol{r}) e^{-\beta V(\boldsymbol{r})} \prod_{j=1}^{N} \delta(x_j(\boldsymbol{r}) - z_j) d\boldsymbol{r},$$

$$\mathcal{I}^{(k)}[g](\boldsymbol{z}) = (\frac{2\pi}{\beta k})^{N/2} \int_{\mathbb{R}^{3D}} g(\boldsymbol{r}) e^{-\beta (V(\boldsymbol{r}) + \frac{k}{2}\sum_{j=1}^{N}(x_j(\boldsymbol{r}) - z_j)^2)} d\boldsymbol{r}.$$

Their difference are bounded by

$$|\mathcal{I}[g](\boldsymbol{z}) - \mathcal{I}^{(k)}[g](\boldsymbol{z})| \le \frac{1}{2\beta k} Tr[\nabla^2_z \mathcal{I}[g](\boldsymbol{z})].$$

Moreover, the difference of the derivatives of $\mathcal{I}[g](\boldsymbol{z})$ and $\mathcal{I}^{(k)}[g](\boldsymbol{z})$ are bounded by

$$|\frac{\partial^p}{\partial z_{i_1}, \dots, \partial z_{i_p}} \mathcal{I}[g](\boldsymbol{z}) - \frac{\partial^p}{\partial z_{i_1}, \dots, \partial z_{i_p}} \mathcal{I}^{(k)}[g](\boldsymbol{z})| \le \frac{1}{2\beta k} \frac{\partial^p}{\partial z_{i_1}, \dots, \partial z_{i_p}} Tr[\nabla^2_z \mathcal{I}[g](\boldsymbol{z})].$$

Proof. The proof of the upper bound of $|\mathcal{I}[g](\boldsymbol{z}) - \mathcal{I}^{(k)}[g](\boldsymbol{z})|$ follows Maragliano et al. (2006). We generate the proof to the upper bound of the derivatives. Consider the Fourier transform of $\mathcal{I}[g](\boldsymbol{z})$ as follows

$$\hat{G}(\zeta) = \int_{\mathbb{C}^N} e^{-i\zeta \cdot \boldsymbol{z}} \int_{\mathbb{R}^{3D}} g(\boldsymbol{r}) e^{-\beta V(\boldsymbol{r})} \prod_{j=1}^{N} \delta(x_j(\boldsymbol{r}) - z_j) d\boldsymbol{r} d\boldsymbol{z} = \int_{\mathbb{R}^{3D}} g(\boldsymbol{r}) e^{-\beta V(\boldsymbol{r})} \int_{\mathbb{C}^N} e^{-i\zeta \cdot \boldsymbol{z}} \prod_{j=1}^{N} \delta(x_j(\boldsymbol{r}) - z_j) d\boldsymbol{z} d\boldsymbol{r}$$

$$= \int_{\mathbb{R}^{3D}} g(\boldsymbol{r}) e^{-\beta V(\boldsymbol{r})} e^{-i\zeta \cdot \boldsymbol{x}(\boldsymbol{r})} d\boldsymbol{r}.$$

Accepted for the 38th Conference on Uncertainty in Artificial Intelligence (UAI 2022).

The Fourier transform of $\mathcal{I}^{(k)}[g](\boldsymbol{z})$ is

$$\begin{aligned}
\hat{G^{(k)}}(\zeta) &= (\frac{2\pi}{\beta k})^{N/2}\int_{\mathbb{C}^N} e^{-i\zeta\cdot\boldsymbol{z}}\int_{\mathbb{R}^{3D}} g(\boldsymbol{r})e^{-\beta(V(\boldsymbol{r})+\frac{k}{2}\sum_{j=1}^N(x_j(\boldsymbol{r})-z_j)^2)}d\boldsymbol{r}d\boldsymbol{z}\\
&= (\frac{2\pi}{\beta k})^{N/2}\int_{\mathbb{R}^{3D}} g(\boldsymbol{r})e^{-\beta V(\boldsymbol{r})}\int_{\mathbb{C}^N}\exp\{-i\zeta\cdot\boldsymbol{z}-\frac{\beta k}{2}\sum_{j=1}^N(x_j(\boldsymbol{r})-z_j)^2\}d\boldsymbol{z}d\boldsymbol{r}\\
&= (\frac{2\pi}{\beta k})^{N/2}\int_{\mathbb{R}^{3D}} g(\boldsymbol{r})e^{-\beta V(\boldsymbol{r})}\int_{\mathbb{C}^N}\exp\{-\frac{\beta k}{2}\sum_{j=1}^N(z_j-x_j(\boldsymbol{r})+\frac{i}{\beta k}\zeta_j)^2-i\zeta\cdot\boldsymbol{x}(\boldsymbol{r})-\frac{|\zeta|^2}{2\beta k}\}d\boldsymbol{z}d\boldsymbol{r}\\
&= \int_{\mathbb{R}^{3D}} g(\boldsymbol{r})e^{-\beta V(\boldsymbol{r})}e^{-i\zeta\cdot\boldsymbol{x}(\boldsymbol{r})-\frac{|\zeta|^2}{2\beta k}}d\boldsymbol{r} = e^{-\frac{|\zeta|^2}{2\beta k}}\hat{G}(\zeta).
\end{aligned}$$

By applying reverse Fourier transformation, we have

$$\begin{aligned}
&|\mathcal{I}[g](\boldsymbol{z})-\mathcal{I}^{(k)}[g](\boldsymbol{z})| = \Big|\int_{\mathbb{C}^N} e^{i\zeta\cdot\boldsymbol{z}}(\hat{G}(\zeta)-\hat{G^{(k)}}(\zeta))d\zeta\Big| \leq \int_{\mathbb{C}^N} e^{i\zeta\cdot\boldsymbol{z}}\hat{G}(\zeta)|1-e^{-\frac{|\zeta|^2}{2\beta k}}|d\zeta\\
&\leq \int_{\mathbb{C}^N} e^{i\zeta\cdot\boldsymbol{z}}\hat{G}(\zeta)\frac{|\zeta|^2}{2\beta k}d\zeta = \frac{1}{2\beta k}Tr[\nabla_z\mathcal{I}^{(k)}[g](\boldsymbol{z})].
\end{aligned}$$

To generalize the upper bound to the derivatives, notice the Fourier transform of the derivatives $\frac{\partial^p}{\partial z_{i_1},\cdots,\partial z_{i_p}}\mathcal{I}[g](\boldsymbol{z})$ and $\frac{\partial^p}{\partial z_{i_1},\ldots,\partial z_{i_p}}\mathcal{I}^{(k)}[g](\boldsymbol{z})$ are $i^p\zeta_{i_1}\ldots\zeta_{i_p}\hat{G}(\zeta)$ and $i^p\zeta_{i_1}\cdots\zeta_{i_p}\hat{G}_k(\zeta)$, respectively. Similarly, applying reverse Fourier transformation

$$\begin{aligned}
&|\frac{\partial^p}{\partial z_{i_1},\ldots,\partial z_{i_p}}\mathcal{I}[g](\boldsymbol{z})-\frac{\partial^p}{\partial z_{i_1},\ldots,\partial z_{i_p}}\mathcal{I}^{(k)}[g](\boldsymbol{z})| \leq \int_{\mathbb{C}^N} e^{i\zeta\cdot\boldsymbol{z}}i^p\zeta_{i1}\cdots\zeta_{i_p}\hat{G}(\zeta)|1-e^{-\frac{|\zeta|^2}{2\beta k}}|d\zeta\\
&\leq \int_{\mathbb{C}^N} e^{i\zeta\cdot\boldsymbol{z}}i^p\zeta_{i1}\cdots\zeta_{i_p}\hat{G}(\zeta)\frac{|\zeta|^2}{2\beta k}d\zeta = \frac{1}{2\beta k}\frac{\partial^p}{\partial z_{i_1},\ldots,\partial z_{i_p}}Tr[\nabla_z^2\mathcal{I}[g](\boldsymbol{z})].
\end{aligned}$$

□

Lemma A.2 (Error by finite T (Maragliano et al., 2006)). *For any function $f(\boldsymbol{r},\boldsymbol{z})$, consider the true average functional and the time average estimator defined as follows*

$$\mathcal{A}^{(k)}[f](\boldsymbol{z}) = \int_{\mathbb{R}^{3N}} f(\boldsymbol{r},\boldsymbol{z})p_k(\boldsymbol{r},\boldsymbol{z})d\boldsymbol{r},$$
$$\mathcal{A}^{(T,k)}[f](\boldsymbol{z}) = \frac{1}{T}\int_0^T f(\boldsymbol{r}(t),\boldsymbol{z})dt.$$

As $T\to\infty$, their difference is given by

$$\mathcal{A}^{(T,k)}[f](\boldsymbol{z})-\mathcal{A}^{(k)}[f](\boldsymbol{z})->\sqrt{\frac{\tau_k[f](\boldsymbol{z})}{T}}\xi_k[f](\boldsymbol{z}),$$

where $\xi_k[f](\boldsymbol{z})$ is a Gaussian variable with mean zero and variance

$$Var[\xi_k[f](\boldsymbol{z})] = \int_{\mathbb{R}^{3N}}(f(\boldsymbol{r},\boldsymbol{z})-\mathcal{A}^{(k)}[f](\boldsymbol{z}))^2p_k(\boldsymbol{r},\boldsymbol{z})d\boldsymbol{r},$$

and $\tau_k[f](\boldsymbol{z})$ is given by

$$\tau_k[f](\boldsymbol{z}) = \frac{1}{Var[\xi_k[f](\boldsymbol{z})]}\int_{t=0}^T\int_{\mathbb{R}^{3N}}\mathbb{E}[f(\boldsymbol{r}(t),\boldsymbol{z})-\mathcal{A}^{(k)}[f](\boldsymbol{z})](f(\boldsymbol{r},\boldsymbol{z})-\mathcal{A}^{(k)}[f](\boldsymbol{z}))^2p_k(\boldsymbol{r},\boldsymbol{z})d\boldsymbol{r}dt.$$

Moreover, as k goes to infinity, $Var[\xi_k[f](\boldsymbol{z})] = Var[\xi[f](\boldsymbol{z})]+O(1/k)$ and $\tau_k[f](\boldsymbol{z}) = \tau[f](\boldsymbol{z})+O(1/\sqrt{k})$, where $\xi[f]$ and $\tau[f]$ are defined by replacing $\mathcal{A}^{(k)}[f]$ by its limiting functional

$$\mathcal{A}[f](\boldsymbol{z}) = Z^{-1}e^{\beta U(\boldsymbol{z})}\int_{\mathbb{R}^{3N}} f(\boldsymbol{r},\boldsymbol{z})e^{-\beta V(\boldsymbol{r})}\prod_{j=1}^N\delta(z_j-x_j(\boldsymbol{r}))d\boldsymbol{r}.$$

Now we are ready to given upper bounds for the errors of estimators of $\nabla U(\boldsymbol{z})$, $M(\boldsymbol{z})$, $\nabla M(\boldsymbol{z})$ and $\nabla^2U(\boldsymbol{z})$ by applying Lemma 1.1 and 1.2 respectively.

A.1 ERROR OF $M(\boldsymbol{z})$

Define $f_{ij}(\boldsymbol{r}) = \sum_k \frac{\partial x_i(r(\alpha))}{\partial r_k}\frac{\partial x_j(r(\alpha))}{\partial r_k}$, and $\mathbf{1}(\boldsymbol{r}) = 1$. Then $M_{ij}(\boldsymbol{z})$ can be written as

$$\begin{aligned} M_{ij}(\boldsymbol{z}) &= Z^{-1}e^{\beta U(\boldsymbol{z})}\int_{\mathbb{R}^{3D}}\sum_k \frac{\partial x_i(\boldsymbol{r}(\alpha))}{\partial r_k}\frac{\partial x_j(\boldsymbol{r}(\alpha))}{\partial r_k}e^{-\beta V(\boldsymbol{r})}\prod_{i=1}^{N}(z_i - x_i(\boldsymbol{r}))d\boldsymbol{r} \\ &= \frac{\int_{\mathbb{R}^{3D}} f_{ij}(\boldsymbol{r})e^{-\beta V(\boldsymbol{r})}\prod_{i=1}^{N}(z_i - x_i(\boldsymbol{r}))d\boldsymbol{r}}{\int_{\mathbb{R}^{3D}} e^{-\beta V(\boldsymbol{r})}\prod_{i=1}^{N}(z_i - x_i(\boldsymbol{r}))d\boldsymbol{r}} = \frac{\mathcal{I}[f_{ij}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})}. \end{aligned}$$

Therefore, the finite-k error $|M_{ij}(\boldsymbol{z}) - M_{ij}^{(k)}(\boldsymbol{z})|$ is bounded by

$$\begin{aligned} |M_{ij}(\boldsymbol{z}) - M_{ij}^{(k)}(\boldsymbol{z})| &= |\frac{\mathcal{I}[f_{ij}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})} - \frac{\mathcal{I}^{(k)}[f_{ij}](\boldsymbol{z})}{\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}| \\ &\leq \frac{1}{2\beta k}\frac{\mathcal{I}[f_{ij}](\boldsymbol{z})Tr[\nabla^2\mathcal{I}[\mathbf{1}](\boldsymbol{z})] + Tr[\nabla^2\mathcal{I}[f_{ij}](\boldsymbol{z})]\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} + O(\frac{1}{k^2}). \end{aligned}$$

The finite-T error is bounded by

$$\begin{aligned} |M_{ij}^{(T,k)}(\boldsymbol{z}) - M_{ij}^{(k)}(\boldsymbol{z})| &= |\frac{1}{T}\int_0^T f_{ij}(\boldsymbol{z})dt - \int_{\mathbb{R}^{3D}} f_{ij}(\boldsymbol{z})p_k(\boldsymbol{r},\boldsymbol{z})d\boldsymbol{r}| \\ &= |\mathcal{A}^{(T,k)}[f_{ij}](\boldsymbol{z}) - \mathcal{A}_k[f_{ij}](\boldsymbol{z})| \to \sqrt{\frac{\tau[f_{ij}](\boldsymbol{z})}{T}}\xi_k[f_{ij}](\boldsymbol{z}). \end{aligned}$$

A.2 ERROR OF $\nabla U(\boldsymbol{z})$

$\nabla_i U(\boldsymbol{z})$ can be written as

$$\begin{aligned} \nabla_i U(\boldsymbol{z}) &= -\beta^{-1}\nabla_i \ln\left(Z^{-1}\int_{\mathbb{R}^{3D}}\exp(-\beta V(\boldsymbol{r}))\prod_{j=1}^{N}\delta(x_j(\boldsymbol{r}) - z_j)d\boldsymbol{r}\right) \\ &= -\beta^{-1}\nabla_i \ln \mathcal{I}[\mathbf{1}](\boldsymbol{z}) = -\beta^{-1}\frac{\nabla_i\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})}. \end{aligned}$$

The finite-k error is bounded by

$$\begin{aligned} |\nabla_i U(\boldsymbol{z}) - \nabla_i U^{(k)}(\boldsymbol{z})| &= \beta^{-1}|\frac{\nabla_i\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})} - \frac{\nabla_i\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}{\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}| \\ &\leq \frac{1}{2\beta^2 k}\frac{|\nabla_i\mathcal{I}[\mathbf{1}](\boldsymbol{z})|Tr[\nabla^2\mathcal{I}[\mathbf{1}](\boldsymbol{z})] + |\nabla_i Tr[\nabla^2\mathcal{I}[\mathbf{1}](\boldsymbol{z})]|\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} + O(\frac{1}{k^2}). \end{aligned}$$

The finite-T error is bounded by

$$\begin{aligned} |\nabla_i U^{(T,k)}(\boldsymbol{z}) - \nabla_i U^{(k)}(\boldsymbol{z})| &= |\frac{k}{T}\int_0^T (z_i - x_i(\boldsymbol{r}(t)))dt - \int_{\mathbb{R}^{3D}} k(z_i - x_i(\boldsymbol{r}))p_k(\boldsymbol{r},\boldsymbol{z})d\boldsymbol{r}| \\ &= k|\mathcal{A}^{(T,k)}[z_i - x_i(\boldsymbol{r})](\boldsymbol{z}) - \mathcal{A}_k[z_i - x_i(\boldsymbol{r})](\boldsymbol{z})| \to k\sqrt{\frac{\tau[z_i - x_i(\boldsymbol{r})](\boldsymbol{z})}{T}}\xi_k[z_i - x_i(\boldsymbol{r})](\boldsymbol{z}). \end{aligned}$$

A.3 ERROR OF $\nabla M(\boldsymbol{z})$

$\nabla_l M_{ij}(\boldsymbol{z})$ can be written as

$$\nabla_l M_{ij}(\boldsymbol{z}) = \nabla_l\left(\frac{\mathcal{I}[f_{ij}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})}\right) = \frac{\nabla_l\mathcal{I}[f_{ij}](\boldsymbol{z})\mathcal{I}[\mathbf{1}](\boldsymbol{z}) - \mathcal{I}[f_{ij}](\boldsymbol{z})\nabla_l\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2}.$$

Therefore, the finite-k error is bounded by

$$
\begin{aligned}
&|\nabla_l M_{ij}(\boldsymbol{z}) - \nabla_l M_{ij}^{(k)}(\boldsymbol{z})| \\
&= |\frac{\nabla_l \mathcal{I}[f_{ij}](\boldsymbol{z})\mathcal{I}[\mathbf{1}](\boldsymbol{z}) - \mathcal{I}[f_{ij}](\boldsymbol{z})\nabla_l \mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} - \frac{\nabla_l \mathcal{I}^{(k)}[f_{ij}](\boldsymbol{z})\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z}) - \mathcal{I}^{(k)}[f_{ij}](\boldsymbol{z})\nabla_l \mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z}))^2}| \\
&\leq |\frac{\nabla_l \mathcal{I}[f_{ij}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})} - \frac{\nabla_l \mathcal{I}^{(k)}[f_{ij}](\boldsymbol{z})}{\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}| + |\frac{\mathcal{I}[f_{ij}](\boldsymbol{z})\nabla_l \mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} - \frac{\mathcal{I}^{(k)}[f_{ij}](\boldsymbol{z})\nabla_l \mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z}))^2}| \\
&\leq \frac{1}{2\beta k}\Big(\frac{|\nabla_l \mathcal{I}[f_{ij}](\boldsymbol{z})|Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})] + |\nabla_l Tr[\nabla^2 \mathcal{I}[f_{ij}](\boldsymbol{z})]|\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} \\
&+ \frac{\mathcal{I}[f_{ij}](\boldsymbol{z})|\nabla_l Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})]| + Tr[\nabla^2 \mathcal{I}[f_{ij}](\boldsymbol{z})]|\nabla_l \mathcal{I}[\mathbf{1}](\boldsymbol{z})|}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} + \frac{\mathcal{I}[f_{ij}](\boldsymbol{z})|\nabla_l \mathcal{I}[\mathbf{1}](\boldsymbol{z})|Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})]}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^3}\Big) + O(\frac{1}{k^2}).
\end{aligned}
$$

The finite-T error is bounded by

$$
\begin{aligned}
&|\nabla_l M_{ij}^{(T,k)}(\boldsymbol{z}) - \nabla_l M_{ij}^{(k)}(\boldsymbol{z})| \leq |\mathcal{A}^{(T,k)}[\frac{\partial f_{ij}(\boldsymbol{z})}{z_l}](\boldsymbol{z}) - \mathcal{A}^{(k)}[\frac{\partial f_{ij}(\boldsymbol{z})}{z_l}](\boldsymbol{z})| \\
&+ \beta k|\mathcal{A}^{(T,k)}[f_{ij}(\boldsymbol{z})(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z}) - \mathcal{A}^{(k)}[f_{ij}(\boldsymbol{z})(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z})| \\
&+ \beta k|\mathcal{A}^{(T,k)}[f_{ij}(\boldsymbol{z})](\boldsymbol{z})\mathcal{A}^{(T,k)}[(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z}) - \mathcal{A}^{(k)}[f_{ij}](\boldsymbol{z})\mathcal{A}^{(k)}[(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z})| \\
\rightarrow &\sqrt{\frac{\tau[\frac{\partial f_{ij}(\boldsymbol{z})}{z_l}](\boldsymbol{z})}{T}}\xi[\frac{\partial f_{ij}(\boldsymbol{z})}{z_l}](\boldsymbol{z}) + \beta k\sqrt{\frac{\tau[f_{ij}(\boldsymbol{z})(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z})}{T}}\xi[f_{ij}(\boldsymbol{z})(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z}) \\
&+ \beta k\sqrt{\frac{\tau[f_{ij}(\boldsymbol{z})]}{T}}\mathcal{A}^{(k)}[(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z})\xi[f_{ij}(\boldsymbol{z})](\boldsymbol{z}) \\
&+ \beta k\sqrt{\frac{\tau[(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z})}{T}}\mathcal{A}^{(k)}[f_{ij}](\boldsymbol{z})\xi[(z_l - x_l(\boldsymbol{r}))](\boldsymbol{z}) + O(\frac{1}{T}).
\end{aligned}
$$

A.4 ERROR OF $\nabla^2 U(\boldsymbol{z})$

$\nabla^2_{ij} U(\boldsymbol{z})$ can be written as

$$
\nabla^2_{ij} U(\boldsymbol{z}) = \nabla_j \left(-\beta^{-1}\frac{\nabla_i \mathcal{I}[\mathbf{1}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})}\right) = -\beta^{-1}\frac{\nabla^2_{ij}\mathcal{I}[\mathbf{1}](\boldsymbol{z})\mathcal{I}[\mathbf{1}](\boldsymbol{z}) - \nabla_i \mathcal{I}[\mathbf{1}](\boldsymbol{z})\nabla_j \mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2}.
$$

Therefore, the finite-k error is bounded by

$$
\begin{aligned}
&|\nabla^2_{ij} U(\boldsymbol{z}) - \nabla^2_{ij} U^{(k)}(\boldsymbol{z})| \\
&= \beta^{-1}|\frac{\nabla^2_{ij}\mathcal{I}[\mathbf{1}](\boldsymbol{z})\mathcal{I}[\mathbf{1}](\boldsymbol{z}) - \nabla_i \mathcal{I}[\mathbf{1}](\boldsymbol{z})\nabla_j \mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} - \frac{\nabla^2_{ij}\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z}) - \nabla_i \mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})\nabla_j \mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z}))^2}| \\
&\leq \beta^{-1}\left(|\frac{\nabla^2_{ij}\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{\mathcal{I}[\mathbf{1}](\boldsymbol{z})} - \frac{\nabla^2_{ij}\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}{\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}| + |\frac{\nabla_i \mathcal{I}[\mathbf{1}](\boldsymbol{z})\nabla_j \mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} - \frac{\nabla_i \mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})\nabla_j \mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}^{(k)}[\mathbf{1}](\boldsymbol{z}))^2}|\right) \\
&\leq \frac{1}{2\beta^2 k}\Big(\frac{|\nabla^2_{ij}\mathcal{I}[\mathbf{1}](\boldsymbol{z})|Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})] + |\nabla^2_{ij} Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})]|\mathcal{I}[\mathbf{1}](\boldsymbol{z})}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} \\
&+ \frac{|\nabla_i \mathcal{I}[\mathbf{1}](\boldsymbol{z})\nabla_j Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})]| + |\nabla_i Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})]\nabla_j \mathcal{I}[\mathbf{1}](\boldsymbol{z})|}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^2} + \frac{|\nabla_i \mathcal{I}[\mathbf{1}](\boldsymbol{z})\nabla_j \mathcal{I}[\mathbf{1}](\boldsymbol{z})|Tr[\nabla^2 \mathcal{I}[\mathbf{1}](\boldsymbol{z})]}{(\mathcal{I}[\mathbf{1}](\boldsymbol{z}))^3}\Big) + O(\frac{1}{k^2}).
\end{aligned}
$$

The finite-T error is bounded by

$$
\begin{aligned}
&|\nabla^2_{ij}U^{(T,k)}(\boldsymbol{z}) - \nabla^2_{ij}U^{(k)}(\boldsymbol{z})| \leq k|\mathcal{A}^{(T,k)}[\frac{\partial(z_j - x_j(\boldsymbol{r}))}{z_i}](\boldsymbol{z}) - \mathcal{A}^{(k)}[\frac{\partial(z_j - x_j(\boldsymbol{r}))}{z_i}](\boldsymbol{z})| \\
&+ \beta k^2|\mathcal{A}^{(T,k)}[(z_j - x_j(\boldsymbol{r}))(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z}) - \mathcal{A}^{(k)}[(z_j - x_j(\boldsymbol{r}))(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z})| \\
&+ \beta k^2|\mathcal{A}^{(T,k)}[(z_j - x_j(\boldsymbol{r}))](\boldsymbol{z})\mathcal{A}^{(T,k)}[(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z}) - \mathcal{A}^{(k)}[(z_j - x_j(\boldsymbol{r}))](\boldsymbol{z})\mathcal{A}^{(k)}[(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z})| \\
&\to k\sqrt{\frac{\tau[\frac{\partial(z_j - x_j(\boldsymbol{r}))}{z_i}](\boldsymbol{z})}{T}}\xi[\frac{\partial(z_j - x_j(\boldsymbol{r}))}{z_i}](\boldsymbol{z}) + \beta k^2\sqrt{\frac{\tau[(z_j - x_j(\boldsymbol{r}))(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z})}{T}}\xi[(z_j - x_j(\boldsymbol{r}))(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z}) \\
&+ \beta k^2\sqrt{\frac{\tau[(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z})}{T}}\mathcal{A}^{(k)}[(z_j - x_j(\boldsymbol{r}))](\boldsymbol{z})\xi[(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z}) \\
&+ \beta k^2\sqrt{\frac{\tau[(z_j - x_j(\boldsymbol{r}))](\boldsymbol{z})}{T}}\mathcal{A}^{(k)}[(z_i - x_i(\boldsymbol{r}))](\boldsymbol{z})\xi[(z_j - x_j(\boldsymbol{r}))](\boldsymbol{z}) + O(\frac{1}{T}).
\end{aligned}
$$

B CONVERGENCE RATE OF HESSIAN-VECTOR PRODUCT ESTIMATOR

Here we discuss the convergence rate of using Hessian vector product estimator

$$
\frac{\nabla U(\boldsymbol{z} + \delta v) - \nabla U(\boldsymbol{z} - \delta v)}{2\delta},
$$

compared to the direct estimator $\nabla^2 Uv$. Here $v = \frac{M^T M \nabla U}{\|M\nabla U\|}$. By Theorem 4.1, we know that

- For any $\boldsymbol{z}$, the error of estimating ∇U is bounded by $|\nabla_i U(\boldsymbol{z}) - \nabla_i U^{(T,k)}(\boldsymbol{z})| \leq O(\frac{1}{k} + \frac{k}{\sqrt{T}})$,
- For any $\boldsymbol{z}$, the error of estimating $v = \frac{M^T M \nabla U}{\|M\nabla U\|}$ is bounded by $|v_i(\boldsymbol{z}) - v_i^{(T,k)}(\boldsymbol{z})| \leq O(\frac{1}{k} + \frac{k}{\sqrt{T}})$,
- For any $\boldsymbol{z}$, the error of estimating $\nabla^2 U$ is bounded by $|\nabla^2_{ij}U(\boldsymbol{z}) - \nabla^2_{ij}U^{(T,k)}(\boldsymbol{z})| \leq O(\frac{1}{k} + \frac{k^2}{\sqrt{T}})$.

First we consider the error of directly estimating $\nabla^2 Uv$.

$$
\begin{aligned}
&|(\nabla^2 Uv)_i - (\nabla^2 U^{(T,k)}v^{(T,k)})_i| = |\sum_j \nabla^2_{ij}U(\boldsymbol{z})v_j(\boldsymbol{z}) - \nabla^2_{ij}U^{(T,k)}(\boldsymbol{z})v_j^{(T,k)}(\boldsymbol{z})| \\
&\leq \sum_j |\nabla^2_{ij}U(\boldsymbol{z})v_j(\boldsymbol{z}) - \nabla^2_{ij}U^{(T,k)}(\boldsymbol{z})v_j^{(T,k)}(\boldsymbol{z})| \\
&\leq \sum_j \Big(|\nabla^2_{ij}U(\boldsymbol{z}) - \nabla^2_{ij}U^{(T,k)}(\boldsymbol{z})|v_j(\boldsymbol{z}) + |v_j(\boldsymbol{z}) - v_j^{(T,k)}(\boldsymbol{z})|\nabla^2 U_{ij}(\boldsymbol{z}) + |\nabla^2_{ij}U(\boldsymbol{z}) - \nabla^2_{ij}U^{(T,k)}(\boldsymbol{z})||v_j(\boldsymbol{z}) - v_j^{(T,k)}(\boldsymbol{z})|\Big) \\
&\leq O(\frac{1}{k} + \frac{k^2}{\sqrt{T}}).
\end{aligned}
$$

So the convergence rate of the error is $O(\frac{1}{k} + \frac{k^2}{\sqrt{T}})$, due to the contribution of $|\nabla^2_{ij}U(\boldsymbol{z}) - \nabla^2_{ij}U^{(T,k)}(\boldsymbol{z})|$.

Now let's consider the Hessian-vector product estimator. The error can be decomposed into three terms as follows

$$
\begin{aligned}
&|\frac{\nabla_i U^{(T,k)}(\boldsymbol{z} + \delta v^{(T,k)}) - \nabla_i U^{(T,k)}(\boldsymbol{z} - \delta v^{(T,k)})}{2\delta} - (\nabla^2 Uv)_i| \\
&\leq |\frac{\nabla_i U^{(T,k)}(\boldsymbol{z} + \delta v^{(T,k)}) - \nabla_i U^{(T,k)}(\boldsymbol{z} - \delta v^{(T,k)})}{2\delta} - \frac{\nabla_i U(\boldsymbol{z} + \delta v^{(T,k)}) - \nabla_i U(\boldsymbol{z} - \delta v^{(T,k)})}{2\delta}| \\
&+ |\frac{\nabla_i U(\boldsymbol{z} + \delta v^{(T,k)}) - \nabla_i U(\boldsymbol{z} - \delta v^{(T,k)})}{2\delta} - (\nabla^2 Uv^{(T,k)})_i| \\
&+ |(\nabla^2 Uv^{(T,k)})_i - (\nabla^2 Uv)_i|.
\end{aligned}
$$

The first term comes from the error of estimating ∇U, which can be upper bounded by

$$
\begin{aligned}
&|\frac{\nabla_i U^{(T,k)}(\boldsymbol{z}+\delta v^{(T,k)}) - \nabla_i U^{(T,k)}(\boldsymbol{z}-\delta v^{(T,k)})}{2\delta} - \frac{\nabla_i U(\boldsymbol{z}+\delta v^{(T,k)}) - \nabla_i U(\boldsymbol{z}-\delta v^{(T,k)})}{2\delta}| \\
&\le \frac{1}{2\delta}\left(|\nabla_i U^{(T,k)}(\boldsymbol{z}+\delta v^{(T,k)}) - \nabla_i U(\boldsymbol{z}+\delta v^{(T,k)})| + |\nabla_i U^{(T,k)}(\boldsymbol{z}-\delta v^{(T,k)}) - \nabla_i U(\boldsymbol{z}-\delta v^{(T,k)})|\right) \\
&\le O(\frac{1}{\delta}(\frac{1}{k}+\frac{k}{\sqrt{T}})).
\end{aligned}
$$

The second term comes from finite difference estimate of $\nabla^2 U$. Using Taylor expansion,

$$
\begin{aligned}
&\nabla_i U(\boldsymbol{z}+\delta v^{(T,k)}) - \nabla_i U(\boldsymbol{z}-\delta v^{(T,k)}) \\
&= \left(U_i(\boldsymbol{z}) + \delta v^{(T,k)}\nabla\nabla_i U(\boldsymbol{z}) + \frac{\delta^2}{2}(v^{(T,k)})^T\nabla^2\nabla_i U(\boldsymbol{z})v^{(T,k)}\right) \\
&- \left(U_i(\boldsymbol{z}) - \delta v^{(T,k)}\nabla\nabla_i U(\boldsymbol{z}) + \frac{\delta^2}{2}(v^{(T,k)})^T\nabla^2\nabla_i U(\boldsymbol{z})v^{(T,k)}\right) + O(\delta^3) \\
&= 2\delta(\nabla^2 U v^{(T,k)})_i + O(\delta^3).
\end{aligned}
$$

Therefore, the error is bounded by

$$
|\frac{\nabla_i U(\boldsymbol{z}+\delta v^{(T,k)}) - \nabla_i U(\boldsymbol{z}-\delta v^{(T,k)})}{2\delta} - (\nabla^2 U v^{(T,k)})_i| = O(\delta^2).
$$

The third term comes from the error of estimating v, which can be upper bounded by

$$
|(\nabla^2 U v^{(T,k)})_i - (\nabla^2 U v)_i| \le \sum_j \nabla^2_{ij} U(\boldsymbol{z})|v_j(\boldsymbol{z}) - v_j^{(T,k)}(\boldsymbol{z})| \le O(\frac{1}{k}+\frac{k}{\sqrt{T}}).
$$

The combined convergence rate of the total estimation error is bounded by

$$
\begin{aligned}
&|\frac{\nabla_i U^{(T,k)}(\boldsymbol{z}+\delta v^{(T,k)}) - \nabla_i U^{(T,k)}(\boldsymbol{z}-\delta v^{(T,k)})}{2\delta} - (\nabla^2 U v)_i| \\
&\le O(\frac{1}{\delta}(\frac{1}{k}+\frac{k}{\sqrt{T}})) + O(\delta^2) + O(\frac{1}{k}+\frac{k}{\sqrt{T}}).
\end{aligned}
$$

The upper bound contains two terms of δ which are $O(\frac{1}{\delta}(\frac{1}{k}+\frac{k}{\sqrt{T}}))$ and $O(\delta^2)$. The optimal δ is $O((\frac{1}{k}+\frac{k}{\sqrt{T}})^{1/3})$ which makes the rate of the error be $O((\frac{1}{k}+\frac{k}{\sqrt{T}})^{2/3})$. The third term is always not the leading term.

Remark: The benifit of Hessian-vector product estimator depends on how fast T grows as k grows. Let $\alpha = \lim_{k\to\infty}\frac{\log T}{\log k}$.

- If $\alpha \le 2$, neither direct estimator nor Hessian-vector product estimator converges.
- If $2 < \alpha \le 4$, direct estimator does not converge, but Hessian-vector product estimator converges at a rate of $O(k^{-(\alpha-2)/3})$.
- If $4 < \alpha \le 6$, both estimators converge. The direct estimator converges at a rate of $O(k^{-(\alpha-4)/2})$ and the Hessian-vector product estimator converges at a rate of $O(k^{-2/3})$. If $\alpha \le 16/3$, the Hessian-vector product estimator converges faster and if $\alpha > 16/3$, the direct estimator converges faster.
- If $\alpha > 6$, both estimators converge. The direct estimator converges at a rate of $O(k^{-1})$ and the Hessian-vector product estimator converges at a rate of $O(k^{-2/3})$. The direct estimator converges faster.

The relation between the convergence rate and α are shown in Figure 1.

C MÜLLER POTENTIAL PARAMETERS

This section provides the detailed parameters in Eq.(21).

$$
\begin{aligned}
&A = (-200, -100, -170, 15),\ \ a = (-1, -1, -6.5, 0.7), \\
&b = (0, 0, 11, 0.6),\ \ c = (-10, -10, -6.5, 0.7), \\
&x^0 = (1, 0, -0.5, -1), y^0 = (0, 0.5, 1.5, 1).
\end{aligned}
$$

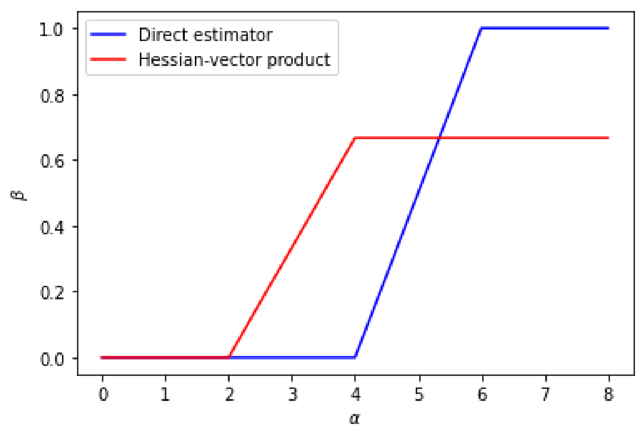

Figure 1: Relation between the convergence rate $\beta = -\lim_{k\to\infty} \frac{\log |\nabla^2 U^{(T,k)} v^{(T,k)} - \nabla^2 U v|}{\log k}$ and the rate $\alpha = \lim_{k\to\infty} \frac{\log T}{\log k}$.