Supplementary Material
CharDiff: Improving Sampling Convergence via Characteristic Function Consistency in Diffusion Models

1. Proof of Eq. 1

Remark 1 ([2]) *On a compact domain of Diameter D, 2-Wasserstein distance between two densities p and q is bounded by total variation norm,*

$$\mathcal{W}_2(p,q) = D\int |p(x) - q(x)|dx \tag{1}$$

From remark, we can further write it for ChF as,

$$\mathcal{W}_2(p,q) = \max_{|\mathbf{u}|} \mathcal{O}\Big(\frac{1}{1+|\mathbf{u}|^p}\Big) \tag{2}$$

Proof 1

$$\begin{aligned}&\int ||s(\mathbf{x},t) - s_\theta(\mathbf{x},t)||_2^2 p(\mathbf{x})d\mathbf{x}\\ &= \int ||\nabla_{\mathbf{x}_t}\log p(\mathbf{x},t) - \nabla_{\mathbf{x}_t}\log p_\theta(\mathbf{x},t)||_2^2 p(\mathbf{x})d\mathbf{x}\end{aligned} \tag{3}$$

Lets assume T be an optimal transport map such that $T_{\#}p = p_\theta$. Using change of variables, we have,

$$p_\theta(T(\mathbf{x}))det(\nabla T(\mathbf{x})) = p(\mathbf{x}) \tag{4}$$

$$\begin{aligned}\nabla_{\mathbf{x}}\log p_\theta(T(\mathbf{x})) =& \nabla_{\mathbf{x}}\log p(\mathbf{x}) + (\nabla_{\mathbf{x}}\log p_\theta(T(\mathbf{x}))\\ &- \nabla_{\mathbf{x}}\log p(\mathbf{x}))\end{aligned} \tag{5}$$

We can further rewrite the equation 3,

$$\begin{aligned}&\int ||s(\mathbf{x},t) - s_\theta(\mathbf{x},t)||_2^2 p(\mathbf{x})d\mathbf{x}\\ &= \int ||\nabla_{\mathbf{x}_t}\log p(\mathbf{x},t) - \nabla_{\mathbf{x}_t}\log p(T(\mathbf{x}),t)||_2^2 p(\mathbf{x})d\mathbf{x}\end{aligned} \tag{6}$$

Using Taylor's expansion, we have,

$$\begin{aligned}&\nabla_{\mathbf{x}}\log p(T(\mathbf{x})) = \nabla_{\mathbf{x}}\log p(\mathbf{x}) + \nabla^2_{\mathbf{x}}\log p(\mathbf{x})(T(\mathbf{x}) - \mathbf{x})\\ \implies &||\nabla_{\mathbf{x}}\log p(T(\mathbf{x})) - \nabla_{\mathbf{x}}\log p(\mathbf{x})||_2^2 \leq \frac{1}{\lambda_q}||T(\mathbf{x}) - \mathbf{x}||_2^2,\end{aligned} \tag{7}$$

where λ_q is the smallest eigenvalue of Fisher information. Finally, we write Eq. 6 as,

$$\begin{aligned}&\sqrt{\int ||s(\mathbf{x},t) - s_\theta(\mathbf{x},t)||_2^2 p(\mathbf{x})d\mathbf{x}}\\ &\leq \sqrt{\int ||T(x) - x||_2^2 p(\mathbf{x})d\mathbf{x}}\\ &= \mathcal{W}_2(p_\theta, p)\end{aligned} \tag{8}$$

At last, we use 2 in 8 to get,

$$\sqrt{\int ||s(\mathbf{x},t) - s_\theta(\mathbf{x},t)||_2^2 p(\mathbf{x})d\mathbf{x}} = \max_{|\mathbf{u}|} \mathcal{O}\Big(\frac{1}{1+|\mathbf{u}|^p}\Big) \tag{9}$$

2. Proof of Proposition 1

From [4], the PDE for $\phi(\mathbf{u})$ is given by,

$$\begin{aligned}\frac{\partial\phi(\mathbf{u})}{\partial t} &= i\sum_{k=1}^{d} u_k \mathbb{E}[e^{i\mathbf{u}^T\mathbf{x}} a_k(\mathbf{x})]\\ &- \frac{1}{2}\sum_{k,l=1}^{d} u_k u_l \mathbb{E}[exp(i\mathbf{u}^T\mathbf{x})\sum_{w=1}^{m_B} b_{kw}(\mathbf{x})b_{lw}(\mathbf{x})],\end{aligned} \tag{10}$$

where the underlying diffusion process is given by, $d\mathbf{x} = \mathbf{a}(\mathbf{x})dt + \mathbf{b}(\mathbf{x})d\mathbf{B}(t)$ with $d\mathbf{B}(t)$ being a Brownian motion. For Score-SDE, we have $\mathbf{a}(\mathbf{x}) = 0$ and $\mathbf{b}(\mathbf{x}) = c(t)\mathbf{I}$ with $c(t)$ as a function of time.

The Eq. 10 becomes,

$$\frac{\partial\phi(\mathbf{u})}{\partial t} = -\frac{d}{2}||\mathbf{u}||_2^2 c^2(t)\phi(\mathbf{u}) \tag{11}$$

3. Proof of Proposition 2

We start with applying Taylor series expansion and ignoring higher order terms of $\mathbf{u}$.

$$
\begin{aligned}
&||\phi(\mathbf{u},\mathbf{x}_0)-\phi(\mathbf{u},\hat{\mathbf{x}}_0)||^2\\
&=||e^{i\mathbf{u}^T\sqrt{\overline{\alpha}_t}\mathbf{x}_0}-e^{i\mathbf{u}^T\sqrt{\overline{\alpha}_t}\hat{\mathbf{x}}_0}||^2e^{-0.5\sigma_t^2||\mathbf{u}||_2^2}\\
&\approx\overline{\alpha}_t||\mathbf{x}_0-\hat{\mathbf{x}}_0||_2^2||\mathbf{u}||_2^2e^{-0.5\sigma_t^2||\mathbf{u}||_2^2}\\
&=\overline{\alpha}_t\Big\|\frac{1}{\sqrt{\overline{\alpha}_t}}\Big(\mathbf{x}_t-(1-\overline{\alpha}_t)\nabla_{\mathbf{x}}\log p(\mathbf{x},t)\Big)\\
&-\frac{1}{\sqrt{\overline{\alpha}_t}}\Big(\mathbf{x}_t'-(1-\overline{\alpha}_t)s_\theta(\mathbf{x},t)\Big)\Big\|_2^2e^{-0.5\sigma_t^2||\mathbf{u}||_2^2}\\
&\leq\|\mathbf{x}_t-\mathbf{x}_t'\|_2^2+(1-\overline{\alpha}_t)||\nabla_{\mathbf{x}}\log p(\mathbf{x},t)-s_\theta(\mathbf{x},t)||_2^2||\mathbf{u}||_2^2
\end{aligned}
\tag{12}
$$

For forward and reverse distributions as time t, $\mathbb{Q}_t=\mathbb{P}_{T-t}$. Assuming that x_t and x_t' are sampled from nearly identical distribution, the bound over the distance can be found by using Chernoff inequality. Substituing this inequality completes the proof.

4. Proof of Theorem 1

We first show that,

$$\mathcal{W}_2^2(p,q)\leq\mathbb{E}_{\mathbf{u}}[\|\phi_P(\mathbf{u})-\phi_Q(\mathbf{u})\|^2] \tag{13}$$

Proof 2 *From [1],*

$$
\begin{aligned}
&\mathcal{W}_2^2(p,q)\leq 4\int|m||F_{\mathbb{P}}(m)-F_{\mathbb{Q}}(m)|dm\\
&\leq 4\int|m||\int_{-\infty}^{m}(p(s)-q(s))ds|dm\\
&\leq 4\int|m|\int_{-\infty}^{\infty}|p(s)-q(s)|dsdm\\
&\leq 4\int|m|dm\int_{-\infty}^{\infty}|p(s)-q(s)|ds\\
&\leq 4\int|m|dm\int_{-\infty}^{\infty}\Big|\frac{1}{(2\pi)^d}\int_{-\infty}^{\infty}(\phi_{\mathbb{P}}(u)-\phi_{\mathbb{Q}}(u))e^{-ius}du\Big|ds\\
&\leq 4\int|m|dm\int_{-\infty}^{\infty}\Big|\frac{1}{(2\pi)^d}\int_{-\infty}^{\infty}(\phi_{\mathbb{P}}(u)-\phi_{\mathbb{Q}}(u))du\Big|ds\\
&\leq 4\int|m|dm\int_{-\infty}^{\infty}ds\Big(\frac{1}{(2\pi)^d}\int_{-\infty}^{\infty}|\phi_{\mathbb{P}}(u)-\phi_{\mathbb{Q}}(u)|du\Big)
\end{aligned}
\tag{14}
$$

Since the samples are drawn from a normal distribution, it can be assumed that for some δ, $\int|m|dm\approx\int_{(1-\delta)\mu_t}^{(1+\delta)\mu_t}|m|dm$ with probability of at least $1-2e^{-\frac{\delta^2\mu_t^2}{2\sigma_t^2}}$.

$$\mathcal{W}_2^2(p,q)\leq\frac{16\mu_t^2\delta^2}{(2\pi)^d}\mathbb{E}_u[|\phi_{\mathbb{P}}(u)-\phi_{\mathbb{Q}}(u)] \tag{15}$$

From [3], it is known that,

$$\mathcal{FID}(p,q)=\mathcal{FD}(\mathcal{I}_{\#}p,\mathcal{I}_{\#}q)\leq 2\mathcal{W}_2^2(\mathcal{I}_{\#}p,\mathcal{I}_{\#}q)\leq 2L^2\mathcal{W}_2^2(p,q), \tag{16}$$

where $\mathcal{I}$ is the Inception Network, and L is the Lipschitz constant for $\mathcal{I}$. Substituting 16 in 15 complets the proof.

5. Proof of Theorem 2

Proof 3

$$
\begin{aligned}
\nabla_{\mathbf{x}_t}\mathcal{L}=&(\hat{\phi}(\mathbf{u})-\phi(\mathbf{u}))^T\nabla_{\mathbf{x}_t}(\hat{\phi}(\mathbf{u})-\phi(\mathbf{u}))\\
=&\underbrace{(\mathbb{E}[e^{\mathbf{u}^T\hat{\mathbf{X}}}]-\mathbf{E}[e^{\mathbf{u}^T\mathbf{X}}])}_{C}(i\mathbf{u}^Te^{i\mathbf{u}^T\mathbf{x}_t}-\\
&i\mathbf{u}^Te^{j\mathbf{u}^T\sqrt{\overline{\alpha}_i}\mu-0.5\mathbf{u}^T\sigma_t\mathbf{u}}\nabla_{\mathbf{x}_t}(\sqrt{\overline{\alpha}_i}\mu))
\end{aligned}
\tag{17}
$$

Now, C can be thought of as discretization error due to finite sampling in the following way,

$$
\begin{aligned}
C\approx&i\mathbf{u}^T\mathbb{E}[\hat{\mathbf{X}}-\mathbf{X}]\\
=&i\mathbf{u}^T\delta_{discrete}
\end{aligned}
\tag{18}
$$

$$\nabla_{\mathbf{x}_t}(\sqrt{\overline{\alpha}_i}\mu)=\mathbb{I}-\frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\nabla_{\mathbf{x}_t}^2\log p_\theta \tag{19}$$

It is to be noted that $e^{ip}=cos(p)+isin(p)$. Substituting C and $\nabla_{\mathbf{x}_t}(\sqrt{\overline{\alpha}_i}\mu)$, and taking $\nabla_{\mathbf{x}_t}\mathcal{L}=Real(\nabla_{\mathbf{x}_t}\mathcal{L})+Im(\nabla_{\mathbf{x}_t}\mathcal{L})$,We get,

$$
\begin{aligned}
\nabla_{\mathbf{x}_t}\mathcal{L}=&\mathbf{u}^T\delta_{discrete}\mathbf{u}^T\Big((cos(\frac{\mathbf{u}^T\mathbf{x}_t}{2})-sin(\frac{\mathbf{u}^T\mathbf{x}_t}{2}))^2+\\
&e^{-\sigma_t^2||\mathbf{u}||_2^2}(cos(\frac{\mathbf{u}^T\mu_t}{2})-sin(\frac{\mathbf{u}^T\mu_t}{2})(\mathbb{I}\\
&-\frac{1-\alpha_t}{\sqrt{1-\overline{\alpha}_t}}\nabla_{\mathbf{x}_t}^2\log p_\theta)\Big)
\end{aligned}
\tag{20}
$$

Taking Expectation over $\mathbf{u}$, we get,

$$\nabla_{\mathbf{x}_t}\mathbb{E}_{\mathbf{u}}[\mathcal{L}]=\mathcal{A}+\mathcal{B}(\mathbb{I}-\nabla_{\mathbf{x}_t}^2\log p_\theta), \tag{21}$$

where $\mathcal{A}=\mathbb{E}\Big[\mathbf{u}^T\delta_{discrete}\mathbf{u}\Big(cos(\frac{\mathbf{u}^T\mathbf{x}_t}{2})-sin(\frac{\mathbf{u}^T\mathbf{x}_t}{2})\Big)^2\Big]$,
$\mathcal{B}=\mathbb{E}\Big[e^{-\sigma_t^2||\mathbf{u}||_2^2}\mathbf{u}^T\delta_{discrete}\mathbf{u}\Big(cos(\frac{\mathbf{u}^T\mu_t}{2})-sin(\frac{\mathbf{u}^T\mu_t}{2})\Big)^2\Big]$, where we use the identity $cosx-sinx=(cos\frac{x}{2}-sin\frac{x}{2})^2$

References

[1] Victor M. Panaretos and Yoav Zemel. Statistical aspects of wasserstein distances. *Annual Review of Statistics and Its Application*, 6(1):405–431, Mar. 2019. 2

[2] Gabriel Peyré and Marco Cuturi. Computational optimal transport, 2020. 1

[3] Litu Rout, Alexander Korotin, and Evgeny Burnaev. Generative modeling with optimal transport maps. In *International Conference on Learning Representations*, 2022. 2

[4] Wayne Isaac Tan Uy and Mircea Grigoriu. Neural network representation of the probability density function of diffusion processes, 2020. 1