A Proof of Proposition 1

Proof. Assume by induction that $\boldsymbol{x}_t^{k+1} = \boldsymbol{x}_t^\star$. Then

$$\begin{aligned}
\boldsymbol{x}_{t+1}^{k+1} &= \boldsymbol{x}_0^k + \frac{1}{T}\sum_{i=0}^{t} s(\boldsymbol{x}_i^k, \frac{i}{T}) \\
&= \left[\boldsymbol{x}_0^k + \frac{1}{T}\sum_{i=0}^{t-1} s(\boldsymbol{x}_i^k, \frac{i}{T})\right] + \frac{1}{T}s(\boldsymbol{x}_t^k, \frac{t}{T}) \\
&= \boldsymbol{x}_t^{k+1} + \frac{1}{T}s(\boldsymbol{x}_t^k, \frac{t}{T}) \\
&= \boldsymbol{x}_t^{k+1} + \frac{1}{T}s\left(h_{t-1}(\ldots h_2(h_1(\boldsymbol{x}_0))), \frac{t}{T}\right) \\
&= \boldsymbol{x}_t^\star + \frac{1}{T}s(\boldsymbol{x}_t^\star, \frac{t}{T}) = \boldsymbol{x}_{t+1}^\star
\end{aligned}$$

□

B Proof of Proposition 2

Proof. A linear convergence rate with factor ≥ 2 ensures our error from the solution $\boldsymbol{x}_t^\star$ given by sequential sampling at each timestep t is bounded by the chosen tolerance.

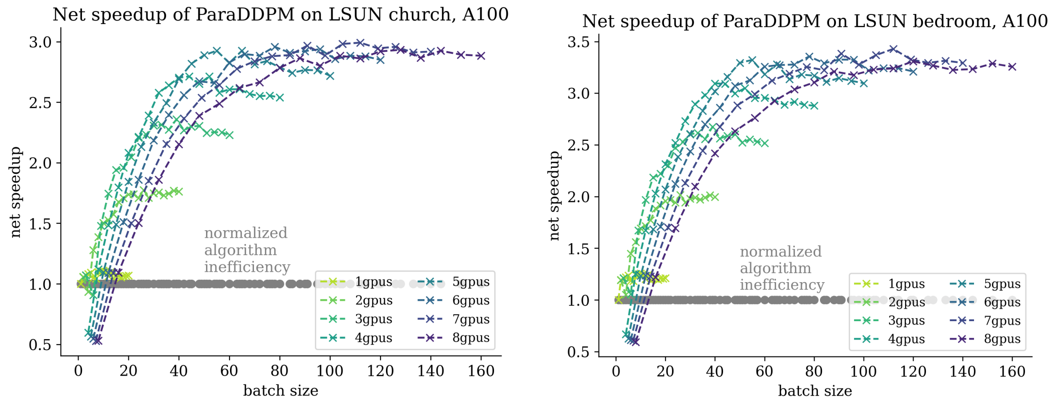
$$\|\boldsymbol{x}_t^K - \boldsymbol{x}_t^\star\|^2 \leq \lim_{n\to\infty}\sum_{j=K+1}^{n}\|\boldsymbol{x}_t^j - \boldsymbol{x}_t^{j-1}\|^2 \leq \lim_{n\to\infty}\sum_{j=K+1}^{n}\frac{1}{2^{j-K}}\|\boldsymbol{x}_t^K - \boldsymbol{x}_t^{K-1}\|^2 \leq \|\boldsymbol{x}_t^K - \boldsymbol{x}_t^{K-1}\|^2$$

Then, for each timestep t, since the inference model samples from a Gaussian with variance σ_t^2, we can bound the total variation distance.

$$\begin{aligned}
D_{\mathrm{TV}}(\mathcal{N}(\boldsymbol{x}_t^K, \sigma_t^2\boldsymbol{I}) \,||\, \mathcal{N}(\boldsymbol{x}_t^\star, \sigma_t^2\boldsymbol{I})) &\leq \sqrt{\frac{1}{2}D_{\mathrm{KL}}(\mathcal{N}(\boldsymbol{x}_t^K, \sigma_t^2\boldsymbol{I}) \,||\, \mathcal{N}(\boldsymbol{x}_t^\star, \sigma_t^2\boldsymbol{I}))} \\
&= \sqrt{\frac{\|\boldsymbol{x}_t^K - \boldsymbol{x}_t^\star\|^2}{4\sigma_t^2}} \leq \sqrt{\frac{\|\boldsymbol{x}_t^K - \boldsymbol{x}_t^{K-1}\|^2}{4\sigma_t^2}} \leq \frac{\epsilon}{T}
\end{aligned}$$

Finally, we make use of the data processing inequality, that $D_{\mathrm{TV}}(f(P) \,||\, f(Q)) \leq D_{\mathrm{TV}}(P \,||\, Q)$, so the total variation distance d_t between the sample and model distribution after t timesteps does not increase when transformed by p_θ. Then by the triangle inequality we get that $d_t \leq d_{t-1} + \epsilon/T$. giving a total variation distance d_T of at most $T\epsilon/T = \epsilon$ for the final timestep T. □

C Additional Experiments



(a) LSUN Church, close to 3x net wall clock speedup with 1000-step ParaDDPM

(b) LSUN Bedroom, over 3x net wall clock speedup with 1000-step ParaDDPM

Figure 5: Unconditional generation of 256x256 images on diffusion models prtrained on the LSUN Church and Bedroom dataset, running ParaDDPM for 1000 steps on A100 GPUs. We plot the net speedup after dividing the hardware efficiency by the algorithm inefficiency as the batch size increases.