SCORE MATCHING IN SHORT

DATA REALLY COMES FROM $g(x)$.
I MODEL THIS WITH $f(x,\theta)$.

<u>HEURISTICS</u> $\underset{\theta}{\mathrm{ARGMIN}}\ \mathbb{E}_g\left[\left\|\nabla_x \log g(x) - \nabla_x \log f(x,\theta)\right\|^2\right]$

$$= \underset{\theta}{\mathrm{ARGMIN}}\ \mathbb{E}_g\left[(\nabla_x \log f(x,\theta))^2\right] - 2\,\mathbb{E}_g\left[\nabla_x \log f(x,\theta)\,\nabla_x \log g(x)\right]$$

$$= \mathrm{ARGMIN}\ \mathbb{E}_g\left[(\nabla_x \log f(x,\theta))^2\right] - 2\int_{\mathcal{X}} \nabla_x \log f(x,\theta)\,\nabla_x g(x)\,dx$$

$$= \mathrm{ARGMIN}\ \mathbb{E}_g\left[(\nabla_x \log f(x,\theta))^2\right] - 2\left[\nabla_x \log f(x,\theta)\, g(x)\right] + 2\,\mathbb{E}\left[\nabla_x^2 \log f(x,\theta)\right]$$

$$\approx \underset{\theta}{\mathrm{ARGMIN}}\ \boxed{\mathbb{E}_D\left[(\nabla_x \log f(x,\theta))^2 + 2\,\nabla_x^2 \log f(x,\theta)\right]}$$

$$\mathbb{E}_D\left[(\nabla_x V_\theta)^2 - 2\,\Delta_x V_\theta\right]$$

HEURISTICS IS NOT ENOUGH MOTIVATED. I PROPOSE A DIFFERENT IDEA: LOOK AT <u>WGF</u> OF $D_{KL}(\nu_g \| \nu_\theta)$, $d\nu_\theta = k f\, d\lambda = k \cdot e^{-V(\cdot)} d\lambda$

WASSERSTEIN GRADIENT FLOW.

ν_g MOVES MONOTONICALLY ALONG (STRICTLY) CONVEX WASSERSTEIN GEODESICS TOWARDS ν_θ.

IF $\nu_\theta \simeq \nu_g$ THEN THE FLOW IS ALMOST STATIONARY.
SO THE WASSERSTEIN GRADIENT IS SMALL.

<u>FACT</u> $\nabla_W D_{KL}(\nu_g \| \nu_\theta)(x) = \nabla_x V(x) + \nabla_x \log g(x)$.

<u>FACT</u> IF $X \sim \nu_g$, AND I EVOLVE X VIA $\dot{X}_t = -\nabla_W D_{KL}(X_t)$
THEN LAW(X_t) FOLLOWS THE WGF.

<u>IDEA:</u> I WANT $\mathbb{E}_D\left[(-\nabla_x V_\theta - \nabla_x \log g)^2\right]$ SMALL.

$$= \mathbb{E}_D\left[\left\|\nabla_x \log f(x,\theta) - \nabla_x \log g(x)\right\|^2\right]$$