Wasserstein Diffusion
on
Multidimensional Spaces

Theo Sturm

Hausdorff Center for Mathematics
Institute for Applied Mathematics

Universität Bonn

Setting, Outline

Setting: M closed Riemannian manifold, $\mathfrak{m}$ normalized volume measure.

Wasserstein space $(\mathcal{P}(\mathrm{M}), W_2)$ can be regarded as ∞-dimensional Riemannian manifold

Heat equation on M is gradient flow for relative entropy $\mathrm{Ent}(\,.\mid \mathfrak{m})$ on $(\mathcal{P}(\mathrm{M}), W_2)$

- Is there a canonical measure on $(\mathcal{P}(\mathrm{M}), W_2)$?
- Is there a Laplacian on $(\mathcal{P}(\mathrm{M}), W_2)$?
- Is there a stochastic perturbation of the gradient flow for $\mathrm{Ent}(\,.\mid \mathfrak{m})$ (Brownian motion with drift)?

Setting, Outline

Goal: Construct a reversible diffusion process on the space $\mathcal{P}(\mathrm{M})$ of probability measures on M that is

- reversible w.r.t. the entropic measure $\mathbb{P}^\beta$ on $\mathcal{P}(\mathrm{M})$, heuristically given as

$$d\mathbb{P}^\beta(\mu) = \frac{1}{Z} e^{-\beta \,\mathrm{Ent}(\mu|\mathfrak{m})}\, d\mathbb{P}^*(\mu)$$

- associated with a regular Dirichlet form, derived from the pre-Dirichlet form

$$\mathcal{E}(f) = \frac{1}{2}\int_{\mathcal{P}(\mathrm{M})} \left\|\nabla f\right\|^2(\mu)\, d\mathbb{P}^\beta(\mu);$$

  in terms of the Wasserstein gradient in the sense of Otto calculus

- non-degenerate, at least in the case of the n-sphere and the n-torus.

Entropic Measure — Heuristics

Formal ansatz

$$d\mathbb{P}^\beta(\mu) = \frac{1}{Z} e^{-\beta\,\mathrm{Ent}(\mu|\mathfrak{m})}\, d\mathbb{P}^*(\mu) \tag{1}$$

with

- $\mathrm{Ent}(\cdot \mid \mathfrak{m})$ = *relative entropy* w.r.t. normalized volume measure $\mathfrak{m}$
- $\beta > 0$ a constant ('inverse temperature')
- $\mathbb{P}^*$ the (non-existing) 'uniform distribution' on $\mathcal{P}(\mathsf{M})$
- Z a normalizing constant.

Rigorous construction: instead of distribution of μ consider distribution of conjugate measures $\mu^{\mathfrak{c}}$.

The Conjugation Map

Put $\mathfrak{c}(x,y) = \frac{1}{2}\mathrm{d}^2(x,y)$.

- The $\mathfrak{c}$-conjugate of $\varphi : \mathrm{M} \to \mathbb{R}$ is the function

$$\varphi^{\mathfrak{c}}(x) = -\inf_{y\in M}\left[\frac{1}{2}\mathrm{d}^2(x,y) + \varphi(y)\right]$$

- φ is $\mathfrak{c}$-convex if $\varphi = (\varphi^{\mathfrak{c}})^{\mathfrak{c}}$, briefly $\varphi \in \mathcal{K}$
- $\forall \mu \in \mathcal{P}(\mathrm{M}):\ \exists! \varphi \in \tilde{\mathcal{K}} := \mathcal{K}/\mathrm{const.}$ s.t. $\mu = \exp(\nabla\varphi)_* \mathfrak{m}$
- The conjugation map

$$\mathfrak{C} : \tilde{\mathcal{K}} \to \tilde{\mathcal{K}},\ \varphi \mapsto \varphi^{\mathfrak{c}}$$

  is continuous w.r.t. topology of $H^1(\mathrm{M})/\mathrm{const.}$

Theorem

The conjugation map

$$\mathfrak{C}_{\mathcal{P}} : \quad \begin{array}{ccc} \mathcal{P}(\mathrm{M}) & \to & \mathcal{P}(\mathrm{M}) \\ \exp(\nabla\varphi)_*\mathfrak{m} & \mapsto & \exp(\nabla\varphi^{\mathfrak{c}})_*\mathfrak{m} \end{array}$$

is a continuous involutive bijection.

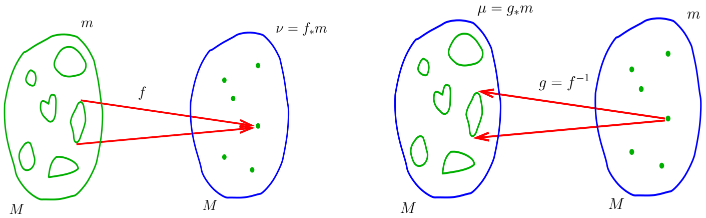
The Conjugation Map

Theorem

The conjugation map

$$\mathfrak{C}_{\mathcal{P}} : \begin{array}{ccc} \mathcal{P}(\mathsf{M}) & \to & \mathcal{P}(\mathsf{M}) \\ \exp(\nabla\varphi)_*\mathfrak{m} & \mapsto & \exp(\nabla\varphi^{\mathfrak{c}})_*\mathfrak{m} \end{array}$$

is a continuous involutive bijection.



Theorem

$$\mathrm{Ent}(\mu^{\mathfrak{c}} \mid \mathfrak{m}) = \mathrm{Ent}(\mathfrak{m} \mid \mu).$$

Entropic Measure — Heuristics

Formal ansatz $d\mathbb{P}^\beta(\mu) = \frac{1}{Z} e^{-\beta\,\mathrm{Ent}(\mu|\mathfrak{m})}\; d\mathbb{P}^*(\mu)$ is reminiscent of Feynman's heuristic picture of the Wiener measure:

$$d\mathbf{P}^\beta(dg) = \frac{1}{Z_\beta}\, e^{-\beta\cdot H(g)}\, d\mathbf{P}^*(g) \tag{2}$$

with energy $H(g) = \frac{1}{2}\int_0^1 g'(t)^2 dt$.
Given any finite partition $\{0 = t_0 < t_1 < \cdots < t_N = 1\}$ of $[0,1]$, replace $H(g)$ by

$$H_N(g) = \inf\left\{H(\tilde g):\ \tilde g \in \mathcal{C}_0,\ \tilde g(t_i) = g(t_i)\ \forall i\right\} = \sum_{i=1}^N \frac{|g(t_i) - g(t_{i-1})|^2}{2(t_i - t_{i-1})}.$$

Then (2) leads to explicit representation for the finite dimensional distributions

$$\mathbf{P}^\beta\left(g_{t_1} \in dx_1, \ldots, g_{t_N} \in dx_N\right) = \frac{1}{Z_{\beta,N}} \exp\left(-\frac{\beta}{2}\sum_{i=1}^N \frac{|x_i - x_{i-1}|^2}{t_i - t_{i-1}}\right) p_N(dx_1, \ldots, x_N). \tag{3}$$

Here $p_N(dx_1, \ldots, x_N) = \mathbf{P}^*\left(g_{t_1} \in dx_1, \ldots, g_{t_N} \in dx_N\right) =$ 'uniform distribution' on $\mathbb{R}^N$. Choosing p_N to be the N-dimensional Lebesgue measure makes the RHS of (3) a projective family of probability measures.
Kolmogorov's extension theorem: $\exists!$ projective limit, the Wiener measure $\mathbf{P}^\beta$.

Entropic Measure — Heuristics

Probability measures $\mathbf{P}(d\mu)$ on $\mathcal{P}(\mathsf{M})$ are uniquely determined by the distributions $\mathbf{P}_{M_1,\ldots,M_N}$ of $(\mu(M_1),\ldots,\mu(M_N))$ for all partitions of M into disjoint measurable subsets M_i. Conversely, if a consistent family $\mathbf{P}_{M_1,\ldots,M_N}$ of probability measures on $[0,1]^N$ (for all partitions $\dot\cup_{i=1}^N M_i = \mathsf{M}$) is given then there exists a random probability measure $\mathbf{P}$ such that

$$\mathbf{P}_{M_1,\ldots,M_N}(A) = \mathbf{P}((\mu(M_1),\ldots,\mu(M_N)) \in A)$$

for all measurable $A \subset [0,1]^N$ and all partitions $\dot\cup_{i=1}^N M_i = \mathsf{M}$.

We don't have a formula for the finite-dimensional distributions of

$$d\mathbb{P}^\beta(\mu) = \frac{1}{Z} e^{-\beta\,\mathrm{Ent}(\mu|\mathfrak{m})}\, d\mathbb{P}^*(\mu)$$

but we have a formula for the finite-dimensional distributions of

$$d\mathbb{Q}^\beta(\nu) = \frac{1}{Z} e^{-\beta\,\mathrm{Ent}(\mathfrak{m}|\nu)}\, d\mathbb{Q}^*(\nu). \qquad (4)$$

Thus new ansatz, taking into account that $\mathrm{Ent}(\mathfrak{m} \mid \nu) = \mathrm{Ent}(\nu^{\mathfrak{c}} \mid \mathfrak{m})$: Define $\mathbb{Q}^\beta$ by means of the formula for its finite-dimensional distributions and put

$$\mathbb{P}^\beta := (\mathfrak{C}_{\mathcal{P}})_* \mathbb{Q}^\beta.$$

Dirichlet-Ferguson Measure — Heuristics

Given measurable partition $\dot{\cup}_{i=1}^{N} M_i = \mathrm{M}$, heuristic ansatz

$$d\mathbb{Q}^{\beta}(\nu) = \frac{1}{Z} e^{-\beta \,\mathrm{Ent}(\mathfrak{m}|\nu)}\, d\mathbb{Q}^{*}(\nu)$$

yields formula for the finite dimensional distribution on $[0,1]^N$:

$$\mathbb{Q}^{\beta}\Big(\nu : \ \big(\nu(M_1), \ldots, \nu(M_N)\big) \in dx\Big) = \frac{1}{Z_N} e^{-\beta S_{M_1,\ldots,M_N}(x)} q_N(dx)$$

where $S_{M_1,\ldots,M_N}(x) =$ minimum of $\nu \mapsto \mathrm{Ent}(\mathfrak{m} \mid \nu)$ under the constraint $\nu(M_1) = x_1, \ldots, \nu(M_N) = x_N$, that is,

$$S_{M_1,\ldots,M_N}(x) = -\sum_{i=1}^{N} \log \frac{x_i}{\mathfrak{m}(M_i)} \cdot \mathfrak{m}(M_i)$$

and $q_N =$ 'uniform distribution' in the simplex $\big\{x \in [0,1]^N : \sum_{i=1}^{N} x_i = 1\big\}$. Requiring symmetry and invariance under merging/subdividing leads to

$$q_N(dx) = C^N \cdot \frac{dx_1 \ldots dx_{N-1}}{x_1 \cdot x_2 \cdot \ldots \cdot x_{N-1} \cdot x_N} \cdot \delta_{(1-\sum_{i=1}^{N-1} x_i)}(dx_N)$$

for some constant $C \in \mathbb{R}_+$.

Dirichlet-Ferguson Measure – Rigorous

The Dirichlet-Ferguson measure $\mathbb{Q}^\beta$ is the probability measure on $\mathcal{P}(\mathrm{M})$ with

$$\mathbb{Q}^\beta\Big(\nu : \big(\nu(M_1),\dots,\nu(M_N)\big) \in dx\Big) = c \cdot e^{-\beta S_{M_1,\dots,M_N}(x)} q_N(dx)$$

$$= \frac{\Gamma(\beta)}{\prod_{i=1}^N \Gamma(\beta\mathfrak{m}(M_i))} \cdot \prod_{i=1}^N x_i^{\beta\cdot\mathfrak{m}(M_i)-1}\, \delta_{(1-\sum_{i=1}^{N-1} x_i)}(dx_N)\, dx_{N-1}\dots dx_1.$$

The latter, indeed, defines a projective family.

Dirichlet-Ferguson Measure – Rigorous

The Dirichlet-Ferguson measure $\mathbb{Q}^\beta$ is the probability measure on $\mathcal{P}(\mathrm{M})$ with

$$\mathbb{Q}^\beta\Big(\nu : \big(\nu(M_1),\ldots,\nu(M_N)\big) \in dx\Big) = c \cdot e^{-\beta S_{M_1,\ldots,M_N}(x)} q_N(dx)$$

$$= \frac{\Gamma(\beta)}{\prod_{i=1}^N \Gamma(\beta\mathfrak{m}(M_i))} \cdot \prod_{i=1}^N x_i^{\beta\cdot\mathfrak{m}(M_i)-1}\, \delta_{(1-\sum_{i=1}^{N-1} x_i)}(dx_N)\, dx_{N-1}\ldots dx_1.$$

The latter, indeed, defines a projective family.
Alternative, more direct construction, in terms of

- iid sequence $(x_i)_{i\in\mathbb{N}}$ of points in M, distributed according to $\mathfrak{m}$
- iid sequence $(t_i)_{i\in\mathbb{N}}$ of numbers in $[0,1]$, distributed according to Beta$(1,\beta)$-distribution, i.e. $\mathrm{Prob}(t_i \in ds) = \beta(1-s)^{\beta-1}\cdot 1_{[0,1]}(s)ds$.

Define stick breaking process $\lambda_k = t_k \cdot \prod_{i=1}^{k-1}(1-t_i)$ and put

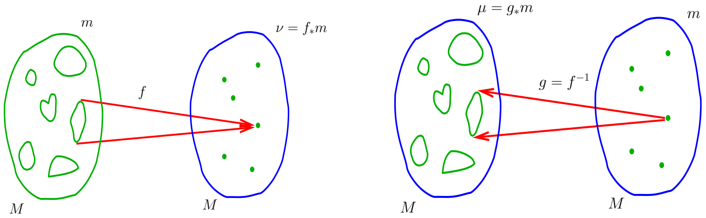
$$\nu = \sum_{k=1}^\infty \lambda_k \cdot \delta_{x_k}.$$

Then $\nu \in \mathcal{P}(\mathrm{M})$ is distributed according to $\mathbb{Q}^\beta$.

Entropic Measure – Rigorous

Definition

Entropic measure
$$\mathbb{P}^{\beta} := (\mathfrak{C}_{\mathcal{P}})_* \mathbb{Q}^{\beta}$$
is push forward of Dirichlet-Ferguson measure $\mathbb{Q}^{\beta}$ (with reference measure $\beta\mathfrak{m}$) under conjugation map $\mathfrak{C}_{\mathcal{P}} : \mathcal{P}(\mathsf{M}) \to \mathcal{P}(\mathsf{M})$.



Theorem

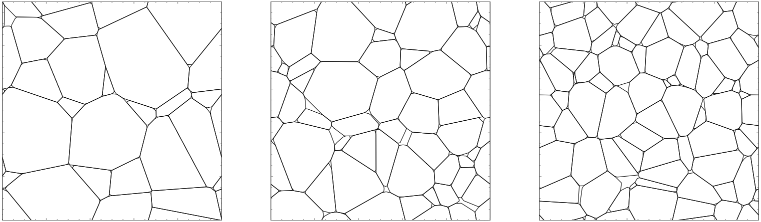
$\mathbb{P}^{\beta}$*-a.e.* μ *has no atoms and no absolutely continuous part.*

Entropic Measure – Large Deviations & Asymptotics

Theorem

$\mathbb{P}^\beta$ *satisfies a Large Deviation Principle*

$$-\inf_{\mu\in A^\circ}\mathrm{Ent}(\mu\mid\mathfrak{m})\le\liminf_{\beta\to\infty}\left[\limsup_{\beta\to\infty}\right]\frac{1}{\beta}\log\mathbb{P}^\beta(A)\le-\inf_{\mu\in\bar{A}}\mathrm{Ent}(\mu\mid\mathfrak{m}).$$



Theorem

- $\lim_{\beta\to 0}\mathbb{P}^\beta=\mathbb{P}^0:=\delta_*\mathfrak{m}$
  push forward of $\mathfrak{m}$ *under the map* $\delta:\mathrm{M}\to\mathcal{P}(\mathrm{M}), x\mapsto\delta_x$
- $\lim_{\beta\to\infty}\mathbb{P}^\beta=\mathbb{P}^\infty:=\delta_{\mathfrak{m}}.$

Wasserstein Dirichlet Form

- $f : \mathcal{P}(\mathrm{M}) \to \mathbb{R}$ is called *cylinder function*, briefly $f \in \mathrm{Cyl}(\mathcal{P}(\mathrm{M}))$, if

$$f(\mu) = F\left(\int_{\mathrm{M}} \vec{V}\, d\mu\right)$$

for suitable $k \in \mathbb{N}$, $F \in \mathcal{C}^1(\mathbb{R}^k)$ and $\vec{V} = (V_1, \dots, V_k) \in \mathcal{C}^1(\mathrm{M}, \mathbb{R}^k)$

- Squared norm of the *Otto-Wasserstein gradient* of f at $\mu \in \mathcal{P}(\mathrm{M})$

$$\left\|\nabla_{\mathrm{W}} f\right\|^2(\mu) := \sum_{i,j=1}^{k} \left(\partial_i F \cdot \partial_j F\right)\left(\int_{\mathrm{M}} \vec{V}\, d\mu\right) \int_{\mathrm{M}} \langle \nabla V_i, \nabla V_j \rangle\, d\mu$$

- Pre-Dirichlet form, defined on cylinder functions

$$\mathcal{E}^0_{\mathrm{W}}(f) := \frac{1}{2} \int_{\mathcal{P}(\mathrm{M})} \left\|\nabla_{\mathrm{W}} f\right\|^2(\mu)\, d\mathbb{P}^{\beta}(\mu).$$

Denote its relaxation in $L^2(\mathcal{P}(\mathrm{M}), \mathbb{P}^{\beta})$ by $\mathcal{E}_{\mathrm{W}}$, that is,

$$\mathcal{E}_{\mathrm{W}}(f) = \liminf_{h \to f \text{ in } L^2} \mathcal{E}^0_{\mathrm{W}}(h).$$

and the domain of the latter by $\mathcal{F}_{\mathrm{W}}$.

Wasserstein Dirichlet Form

Wasserstein Dirichlet Form

$(\mathcal{E}_{\mathsf{W}}, \mathcal{F}_{\mathsf{W}})$ is a strongly local regular Dirichlet form.

It coincides with the Cheeger energy for the metric measure space $(\mathcal{P}(\mathsf{M}), \mathsf{W}_2, \mathbb{P}^\beta)$ (which this way is proven to be infinitesimally Hilbertian).

Wasserstein Diffusion

There exists a strong Markov process $\big((\rho_t)_{t\geq 0}, (\mathbf{P}_\mu)_{\mu\in\mathcal{P}(\mathsf{M})}\big)$ with a.s. continuous trajectories ('diffusion process') properly associated with $(\mathcal{E}_{\mathsf{W}}, \mathcal{F}_{\mathsf{W}})$.

Wasserstein Dirichlet Form

Wasserstein Dirichlet Form

$(\mathcal{E}_{\mathrm{W}}, \mathcal{F}_{\mathrm{W}})$ is a strongly local regular Dirichlet form.

It coincides with the Cheeger energy for the metric measure space $(\mathcal{P}(\mathrm{M}), \mathrm{W}_2, \mathbb{P}^\beta)$ (which this way is proven to be infinitesimally Hilbertian).

Wasserstein Diffusion

There exists a strong Markov process $\big((\rho_t)_{t\geq 0}, (\mathbf{P}_\mu)_{\mu\in\mathcal{P}(\mathrm{M})}\big)$ with a.s. continuous trajectories ('diffusion process') properly associated with $(\mathcal{E}_{\mathrm{W}}, \mathcal{F}_{\mathrm{W}})$.

Open question: Is $\mathcal{E}_{\mathrm{W}} = \mathcal{E}^0_{\mathrm{W}}$ on $Dom(\mathcal{E}^0_{\mathrm{W}})$?
Or, in other words, is $(\mathcal{E}^0_{\mathrm{W}}, Dom(\mathcal{E}^0_{\mathrm{W}}))$ closable?
Affirmative answer in $n = 1$: von Renesse/St. (2009).

Challenge in $n > 1$: Is $\mathcal{E}_{\mathrm{W}} \not\equiv 0$ on its domain?
Or, equivalently, does the Wasserstein diffusion $\big((\rho_t)_{t\geq 0}, (\mathbf{P}_\mu)_{\mu\in\mathcal{P}(\mathrm{M})}\big)$ satisfy

$$\mathbb{P}^\beta\Big\{\mu : \ \mathbf{P}_\mu\big\{\exists t > 0 : \rho_t \neq \rho_0\big\} > 0\Big\} > 0 \ ?$$

Being in Motion

- Relaxed energy

$$\mathcal{E}_{\mathrm{W}}(f) = \liminf_{h \to f \text{ in } L^2} \frac{1}{2} \int_{\mathcal{P}(\mathrm{M})} \left\| \nabla_{\mathrm{W}} h \right\|^2 d\mathbb{P}^\beta = \inf_g \frac{1}{2} \int g^2 \, d\mathbb{P}^\beta$$

where the $\inf_g$ is taken over all weak upper gradients for f.

- $g : \mathcal{P}(\mathrm{M}) \to \mathbb{R}$ is a weak upper gradient for f if

$$\left| f(\mu_1) - f(\mu_0) \right| \leq \int_0^1 g(\mu_t) \cdot \left| \dot{\mu}_t \right| dt$$

for a.e. curve $(\mu_t)_{[0,1]}$ in $\big(\mathcal{P}(\mathrm{M}), \mathrm{W}_2\big)$ where "a.e." means $\mathbf{P}$-a.e. w.r.t. every

$$\mathbf{P} \in \mathcal{P}\Big(\mathcal{A}C_2\Big([0,1]; \big(\mathcal{P}(\mathrm{M}), \mathrm{W}_2\big)\Big)\Big) \quad \text{with} \quad \sup_{\mathrm{t}} \, (\mathrm{e_t})_* \mathbf{P} \leq \mathrm{C} \cdot \mathbb{P}^\beta$$

- Choose $\mathbf{P} := \hat{\Phi}_* \mathbb{P}^\beta$ for suitable Lipschitz family of isometries $\Phi_t : \mathrm{M} \to \mathrm{M}$ and

$$\hat{\Phi} : \mathcal{P}(\mathrm{M}) \to \mathcal{P}\Big(\mathcal{A}C_2\Big([0,1]; \big(\mathcal{P}(\mathrm{M}), \mathrm{W}_2\big)\Big)\Big), \quad \mu \mapsto \big((\Phi_{\mathrm{t}})_* \mu\big)_{[0,1]}$$

Being in Motion

Assume $\mathrm{M} = \mathbb{S}^n$ or $\mathbb{T}^n$.

For all $x_0, x_1 \in \mathrm{M}$: $\exists$ *Lipschitz family of isometries* $(\Phi_t)_{t\in[0,1]}$

- $\Phi_0 = \mathrm{Id}$
- $\Phi_t : \mathrm{M} \to \mathrm{M}$ is an isometry for every $t \in [0,1]$
- $\mathrm{d}(\Phi_s(y), \Phi_t(y)) \leq L \cdot |t - s|$ for all $s, t \in [0,1]$, all $y \in \mathrm{M}$, and $L = \mathrm{d}(x_0, x_1)$
- $\Phi_1(x_0) = x_1$.

Lemma

Let $(\Phi_t)_{t\in[0,1]}$ *be a Lipschitz family of isometries. Then for every* f *on* $\mathcal{P}(\mathrm{M})$

$$\mathcal{E}_{\mathrm{W}}(f) \geq \frac{1}{2L^2} \int_{\mathcal{P}(\mathrm{M})} \left| f\left((\Phi_1)_* \mu\right) - f\left(\mu\right) \right|^2 d\mathbb{P}^\beta(\mu).$$

Being in Motion

f is called antisymmetric if $f\big((\Phi_1)_*\mu\big) = -f\big(\mu\big)$ for all μ.

Theorem

For every antisymmetric Borel function f on $\mathcal{P}(\mathsf{M})$

$$\mathcal{E}_{\mathsf{W}}(f) \geq \frac{2}{L^2} \int_{\mathcal{P}(\mathsf{M})} f^2\big(\mu\big)\, d\mathbb{P}^\beta(\mu).$$

In particular, $\quad \mathcal{E}_{\mathsf{W}}(f) = 0 \quad \Longleftrightarrow \quad f \equiv 0.$

Theorem

For every non-constant Lipschitz function $V : \mathsf{M} \to \mathbb{R}$, the function $f : \mathcal{P}(\mathsf{M}) \to \mathbb{R}$, $\mu \mapsto \int_{\mathsf{M}} V\, d\mu$ has nonvanishing Wasserstein energy $\mathcal{E}_{\mathsf{W}}(f)$.

More explicitly, for every $x_0, x_1 \in \mathsf{M}$ and $\epsilon > 0$ with $|V(x_1) - V(x_0)| \geq \operatorname{Lip} V \cdot \big[\frac{1}{2} d(x_0, x_1) + 2\epsilon\big]$ and with $\eta := \mathbb{P}^\beta\big(\mathbb{B}_\epsilon(\delta_x)\big) > 0$,

$$\mathcal{E}_{\mathsf{W}}(f) \geq \frac{\eta}{8}(\operatorname{Lip} V)^2.$$