A. Background

A.1. The continuity equation

Let $T > 0$. Consider a weakly continuous family of probability measures on $\mathbb{R}^d$, $\mu : (0,T) \to \mathcal{P}_2(\mathbb{R}^d), t \mapsto \mu_t$. It satisfies a continuity equation (Ambrosio et al., 2008, Section 8.1) if there exists $(v_t)_{t\in(0,T)}$ such that $v_t \in L^2(\mu_t)$ and :

$$\frac{\partial \mu_t}{\partial t} + \mathrm{div}(\mu_t v_t) = 0 \qquad \text{in } \mathbb{R}^d \times (0,T) \tag{19}$$

holds in the sense of distributions, i.e. for any $\phi \in C_c^\infty(\mathbb{R}^d)$, the identity

$$\frac{\mathrm{d}}{\mathrm{dt}} \int_{\mathbb{R}^d} \phi(x) d\mu_t(x) = \int_{\mathbb{R}^d} \langle \nabla \phi(x), v_t(x) \rangle d\mu_t(x) \tag{20}$$

holds for any $t \in (0,T)$. By an integration by parts, the r.h.s. of the previous identity can be written:

$$\int_{\mathbb{R}^d} \langle \nabla \phi(x), v_t(x) \rangle d\mu_t(x) = -\int_{\mathbb{R}^d} \phi(x)\,\mathrm{div}(\mu_t(x) v_t(x))\mathrm{d}x. \tag{21}$$

Hence, the identity (20) can be rewritten as

$$\int_{\mathbb{R}^d} \phi(x) \frac{\partial \mu_t(x)}{\partial t} \mathrm{d}x + \int_{\mathbb{R}^d} \phi(x)\,\mathrm{div}(\mu_t(x) v_t(x))\mathrm{d}x = 0. \tag{22}$$

This equation expresses the law of conservation of mass for any volume of a moving fluid. Assume $d = 3$ and denote by (v_t^x, v_t^y, v_t^z) the projections of the vector field on the axes $(\overrightarrow{x}, \overrightarrow{y}, \overrightarrow{z})$. Then, the continuity equation has the form

$$\frac{\partial \mu_t}{\partial t} + \mathrm{div}(\mu_t v_t) = \frac{\partial \mu_t}{\partial t} + \frac{\partial(\mu_t v_t^x)}{\partial x} + \frac{\partial(\mu_t v_t^y)}{\partial y} + \frac{\partial(\mu_t v_t^z)}{\partial z} = 0. \tag{23}$$

A.2. Differentiability and convexity on the Wasserstein space

Let $(\mu, \nu) \in \mathcal{P}_2(\mathbb{R}^d)$. The Wasserstein 2 distance is defined as :

$$W_2^2(\mu,\nu) = \inf_{q \in \mathcal{Q}(\mu,\nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\|^2 dq(x,y) \tag{24}$$

where $\mathcal{Q}(\mu,\nu)$ is the set of couplings between μ and ν, *i.e.* the set of nonnegative measures q over $\mathbb{R}^d \times \mathbb{R}^d$ such that $P_{1\#}q = \mu$ (resp. $P_{2\#}q = \nu$) where $P_1 : (x,y) \mapsto x$ (resp. $P_2 : (x,y) \mapsto y$).

The Wasserstein space $(\mathcal{P}_2(\mathbb{R}^d), W_2)$ is not a Riemannian manifold but it can be equipped with a Riemannian structure and interpretation (Otto, 2001). In this geometric interpretation, the tangent space to $\mathcal{P}_2(\mathbb{R}^d)$ at μ is included in $L_2(\mu)$, and is equipped with the scalar product, defined for $f, g \in L_2(\mu)$ by:

$$\langle f, g \rangle_{L_2(\mu)} = \int_{\mathbb{R}^d} f(x) g(x) \mathrm{d}\mu(x). \tag{25}$$

Let $\mathcal{F} : \mathcal{P}_2(\mathbb{R}^d) \to \mathbb{R}$ be a functional on the Wasserstein space. We clarify in this section the notions of differentiability of $\mathcal{F}$ that we consider in this setting. The notion of Fréchet subdifferentiability and its properties have been extended to the Wasserstein framework in (Ambrosio et al., 2008, Chapter 10). We first recall that if it exists, the *first variation of $\mathcal{F}$ evaluated at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$* is the unique function $\frac{\partial \mathcal{F}(\mu)}{\partial \mu} : \mathbb{R}^d \to \mathbb{R}$ s.t.

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon}(\mathcal{F}(\mu + \epsilon \xi) - \mathcal{F}(\mu)) = \int_{\mathbb{R}^d} \frac{\partial \mathcal{F}(\mu)}{\partial \mu}(x) d\xi(x) \tag{26}$$

for all $\xi = \nu - \mu$, where $\nu \in \mathcal{P}_2(\mathbb{R}^d)$. Under mild regularity assumptions, the W_2 gradient of $\mathcal{F}$ corresponds to the gradient of the first variation of $\mathcal{F}$, as stated below.

Definition 4. (Ambrosio et al., 2008, Lemma 10.4.1). Let $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, absolutely continuous with respect to the Lebesgue measure, with density in $C^1(\mathbb{R}^d)$ and such that $\mathcal{F}(\mu) < \infty$. The *subdifferential* of $\mathcal{F}$ at μ is the map $\nabla_{W_2}\mathcal{F}(\mu)$ defined by:

$$\nabla_{W_2}\mathcal{F}(\mu)(x) = \nabla \frac{\partial \mathcal{F}(\mu)}{\partial \mu}(x) \text{ for } \mu\text{-a.e. } x \in \mathbb{R}^d, \tag{27}$$

and for every vector field $\xi \in C_c^\infty(\mathbb{R}^d; \mathbb{R}^d)$,

$$\int_{\mathbb{R}^d} \langle \nabla_{W_2}\mathcal{F}(\mu)(x), \xi(x)\rangle \mathrm{d}\mu(x) = -\int_{\mathbb{R}^d} \frac{\partial \mathcal{F}(\mu)}{\partial \mu}(x) \operatorname{div}(\mu(x)\xi(x))\mathrm{d}x. \tag{28}$$

Moreover, $\nabla_{W_2}\mathcal{F}(\mu)$ belongs to the tangent space of $\mathcal{P}_2(\mathbb{R}^d)$ at μ, which is included in $L^2(\mu)$.

A.3. Cauchy-Lipschitz assumptions for the existence and uniqueness of the Wasserstein gradient flow

Let $T > 0$, and denote by $\mathcal{L}^1$ the standard one-dimensional Lebesgue measure on $[0, T]$ and L^1 the space of measurable and integrable functions w.r.t. Lebesgue measure. The Cauchy-Lipschitz assumptions below for existence and uniqueness of the flow on $[0, T]$ are adapted to our flows from their general differential inclusion version by Bonnet & Frankowska (2021). They hold for an initial distribution $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$, the space of probability measures with compact support. If v only depends on time and not on μ_t, they then write as

- **(B$_1$)** (Carathéodory vector fields) $v : [0, T] \times \mathbb{R}^d \to \mathbb{R}^d$ is such that $t \mapsto v(t, x)$ is $\mathcal{L}^1$-measurable for all $x \in \mathbb{R}^d$ and $x \mapsto v(t, x)$ is continuous for $\mathcal{L}^1$-almost every $t \in [0, T]$.
- **(B$_2$)** (Sublinear growth) There exists a map $m(\cdot) \in L^1([0, T]; \mathbb{R}_+)$ such that $|v(t, x)| \le m(t)(1 + |x|)$ for $\mathcal{L}^1$-almost every $t \in [0, T]$ and all $x \in \mathbb{R}^d$.
- **(B$_3$)** (Lipschitz vector field) For any compact set $\mathcal{K} \subset \mathbb{R}^d$, there exists a map $l_\mathcal{K}(\cdot) \in L^1([0, T], \mathbb{R}_+)$ such that $\mathrm{Lip}(v(t, \cdot); \mathcal{K}) \le l_\mathcal{K}(t)$ for $\mathcal{L}^1$ -almost every $t \in [0, T]$.

For $R > 0$, we denote by $\mathcal{K} := B(0, R)$ the closed ball of radius R in $\mathbb{R}^d$; and for any function f, we denote by $\|f(\cdot)\|_{\infty,\mathcal{K}} := \sup_{x\in\mathcal{K}} |f(x)|$ its supremum over $\mathcal{K}$ and by $\mathrm{Lip}(f(\cdot); \mathcal{K})$ the Lipschitz constant of the restriction of f on $\mathcal{K}$. If v only depends on μ_t, then the following assumptions should hold for every $R > 0$:

- **(C$_1$)** For any $\mu \in \mathcal{P}_c\left(\mathbb{R}^d\right)$, $v_{|\mathcal{K}}(\mu)(\cdot) \in C^0\left(\mathcal{K}; \mathbb{R}^d\right)$.
- **(C$_2$)** There exists $m > 0$ such that for any $\mu \in \mathcal{P}_c\left(\mathbb{R}^d\right)$, for all $y \in \mathbb{R}^d$, we have $\|v(\mu)(y)\| \le m\left(1 + \|y\| + \int \|x\| \mathrm{d}\mu(x)\right)$
- **(C$_3$)** There exists $l_\mathcal{K} > 0$ such that for any $\mu \in \mathcal{P}_c(\mathcal{K})$, we have $\mathrm{Lip}(v(\mu)(\cdot); \mathcal{K}) \le l_\mathcal{K}$,
- **(C$_4$)** There exists $L_\mathcal{K} > 0$ such that for any $\mu, \nu \in \mathcal{P}_c(\mathcal{K})$, we have $\|v_{|\mathcal{K}}(\mu)(\cdot) - v_{|\mathcal{K}}(\nu)(\cdot)\|_{\infty,\mathcal{K}} \le L_\mathcal{K} W_2(\mu, v)$.

Relation with Assumption (A$_1$). Consider $\mu \in \mathcal{P}_c\left(\mathbb{R}^d\right)$ and take $R > 0$ such that $\mathrm{supp}(\mu) \subset \mathcal{K} := B(0, R)$. Our Assumption (A$_1$) implies that $y \mapsto v_\mu(y) = \int \nabla_2 k_\pi(x, y)\mathrm{d}\mu(x)$ is Lipschitz with constant $l_\mathcal{K} = \sup_{y\in\mathcal{K}} L(y)$, so Assumptions (C$_1$) and (C$_3$) hold. Assumption (C$_2$) corresponds to Assumption (A$_2$). Finally, for $\nu \in \mathcal{P}_c(\mathcal{K})$, since $\nabla_2 k_\pi(\cdot, y)$ is $l_\mathcal{K}$-Lipschitz for $y \in \mathcal{K}$, we have that $x \mapsto \|\nabla_2 k_\pi(x, y)\|$ is also $l_\mathcal{K}$-Lipschitz, hence

$$\|v_\mu(y) - v_\nu(y)\| \le \sup\left\{\int_M f(x)\mathrm{d}(\mu - \nu)(x) \mid \text{Lipschitz } f : \mathcal{K} \to \mathbb{R}, \mathrm{Lip}(f) \le l_\mathcal{K}\right\} = l_\mathcal{K} W_1(\mu, \nu) \le l_\mathcal{K} W_2(\mu, \nu),$$

where the last inequality between 1-Wasserstein distance W_1 and W_2 is a consequence of Jensen's inequality, so Assumption (C$_4$) is satisfied.

For other kernel-based updates as presented in Appendix A.4, the kernel is at most multiplied once by the score s, so for uniformly Lipschitz s and kernels with bounded derivatives, the Lipschitz constant is uniform for $v(\mu)(\cdot)$. This was for instance assumed in Arbel et al. (2019). However for updates involving the Stein kernel such as in KSD Descent, the analysis is more intricate. As a matter of fact, when π is a standard Gaussian distribution, we have $s(x) = x$. Hence the Stein kernel verifies, for a smooth translation-invariant kernel with $k(x, x) = 1$, that $k_\pi(x, x) = C + \|x\|^2$ for $x \in \mathbb{R}^d$, with

C a constant determined by the last term in k_π. So, for a Gaussian π, the Lipschitz constant of $v(\delta_y)(\cdot)$ is larger than $2\|y\|$, explaining the reason why we only required for a y-dependent Lipschitz constant in Assumption $(\mathbf{A}_1)$ and did not assume it to be uniform. This is related to the fact that $k_\pi(x,x)$ is in general unbounded for the Stein kernel, which precludes most of the classical assumptions made in the kernel literature.

Comment on Assumption $(\mathbf{C}_2)$. The main role of Assumption $(\mathbf{C}_2)$ is to impede the well-known phenomenon of finite-time explosion of a trajectory as for the system $z'(t) = z(t)^2$ in $\mathbb{R}$. It could in principle be replaced by any assumption preventing particles from escaping to infinity in finite time. One such lighter assumption would write as follows:

" For any $r > 0, T > 0$ and $\mu_0 \in \mathcal{P}_c\left(\mathbb{R}^d\right)$ with $\text{supp}(\mu_0) \subset rB(0,1)$ there exists $R(r,T) < \infty$ such that $\text{supp}(\mu_t) \subset R(r,T)B(0,1)$ for $(\mu_t)_{t\in[0,T]}$ solving (7)."

A.4. Descent updates of the algorithms presented in Section 3.3

From a computational viewpoint, both the SVGD, LAWGD and MMD flow algorithms, presented in Section 3.3, only propose gradient descent schemes with the following respective updates $x_n^i \leftarrow x_n^i - \gamma\mathcal{D}^i$ with

$$\mathcal{D}^i_{SVGD} = \frac{1}{N}\sum_{j=1}^N \left[k(x_n^j, x_n^i)s(x_n^j) + \nabla_1 k(x_n^j, x_n^i)\right],$$

$$\mathcal{D}^i_{LAWGD} = \frac{1}{N}\sum_{j=1}^N \nabla_2 k_{\mathcal{L}_\pi}(x_n^j, x_n^i),$$

$$\mathcal{D}^i_{MMD-GD} = \frac{1}{N}\sum_{j=1}^N \left[\nabla_2 k(x_n^j, x_n^i) - \nabla_2 k(y^j, x_n^i)\right].$$

where the MMD-GD update requires extra samples $(y^j)_{j=1}^N \sim \pi$, since it is sample-based rather than score-based. Notice that all these updates have the same iteration complexity of order $\mathcal{O}(N^2)$. Intriguingly LAWGD has the same update rule as ours (12) but for their kernel $k_{\mathcal{L}_\pi}$ which, as discussed in Section 3.3, does not incorporate an off-the-shelf kernel k, unlike the Stein kernel k_π (3).

A.5. Background on diffusion operator $\mathcal{L}_\pi$

In this section, for the convenience of the reader who is not familiar with the spectral theory of diffusion operators, we formulate Lemma 12 which provides a formal construction of the diffusion operator $\mathcal{L}_\pi = -\langle\nabla\log\pi, \nabla\rangle - \Delta$ on the space $L_2(\pi)$ and gathers a number of technical facts about it. Those facts form a background for the proofs of results related to lack of exponential convergence near equilibrium of KSD flow, which are presented in Appendix B. We provide the proof of lemma 12 in Appendix B.11.

Lemma 12. Let $\pi \propto e^{-V}$ be a probability measure on $\mathbb{R}^d$ and assume that $V : \mathbb{R}^d \to \mathbb{R}$ is in $C^1(\mathbb{R}^d)$. Let $\hat{\mathcal{L}}_\pi = \langle\nabla V, \nabla\rangle - \Delta$ on $C_c^\infty(\mathbb{R}^d)$. This operator can be extended to a positive self-adjoint operator on $L^2(\pi)$ with dense domain $\mathcal{D}(\mathcal{L}_\pi) \subset L^2(\pi)$ which we denote by $\mathcal{L}_\pi$. Moreover, $C_c^\infty(\mathbb{R}^d)$ is dense in $\mathcal{D}(\mathcal{L}_\pi)$, for the norm:

$$\|\phi\|_{\mathcal{L}_\pi} = \left(\langle\phi, \mathcal{L}_\pi\phi\rangle + \|\phi\|^2_{L_2(\pi)}\right)^{1/2}. \tag{29}$$

From that it follows, that $\mathcal{D}(\mathcal{L}_\pi)$ is the subset of the weighted Sobolev space $W_0^{1,2}(\pi)$[4] (that is, the closure of $C_c^\infty(\mathbb{R}^d)$ in $W^{1,2}(\pi)$) and for all $f \in \mathcal{D}(\mathcal{L}_\pi)$ we have

$$\|\nabla f\|^2_{L^2(\pi)} = \langle f, \mathcal{L}_\pi f\rangle_{L^2(\pi)}, \tag{30}$$

where ∇f is the weak derivative of f. This implies that the kernel of $\mathcal{L}_\pi$ consists of π-almost everywhere constant functions.

Furthermore, for any $f \in \mathcal{D}(\mathcal{L}_\pi)$ we can find a sequence $\phi_n \in C_c^\infty(\mathbb{R}^d)$, such that $\lim_{n\to\infty}\|\phi_n - f\|_{L_2(\pi)} = 0$ and $\lim_{n\to\infty}\|\mathcal{L}_\pi\phi_n - \mathcal{L}_\pi f\|_{L_2(\pi)} = 0$.

[4]Note that the meaning of 0 in the notation $W_0^{1,2}(\pi)$ differs from the meaning of 0 in $L_0^2(\pi) = \{\phi \in L^2(\pi), \int\phi d\pi = 0\}$.

A.6. A Descent lemma for KSD Descent

A descent lemma for (10) is a proposition stating that $\mathcal{F}$ decreases at each iteration of the time-discretized flow. It should hold for both continuous or discrete initializations. Arbel et al. (2019) proved a similar result for $\mathcal{F} = \frac{1}{2}\operatorname{MMD}^2$ which we recall below for completeness.

Proposition 13. (Arbel et al., 2019, Proposition 4) Let μ_n be defined by (10) for $\mathcal{F} = \frac{1}{2}\operatorname{MMD}^2$ and assume that $k \in C^{1,1}(\mathbb{R}^d \times \mathbb{R}^d)$ with l-Lipschitz gradient: $\|\nabla k(x, x') - \nabla k(y, y')\| \leq l\left(\|x - y\| + \|x' - y'\|\right)$ for all $x, x', y, y' \in \mathbb{R}^d$. Then, for $\gamma \leq \frac{2}{l}$, the sequence $(\mathcal{F}(\mu_n))_{n\geq 0}$ is decreasing and for any $n \geq 0$:

$$\mathcal{F}(\mu_{n+1}) - \mathcal{F}(\mu_n) \leq -\gamma\left(1 - \frac{3\gamma l}{2}\right)\|\nabla_{W_2}\mathcal{F}(\mu_n)\|^2_{L^2(\mu_n)}. \tag{31}$$

Although KSD Descent is a special case of MMD Descent, assuming a uniform Lipschitz constant for k_π is excessive, even for targets as simple as a single Gaussian distribution, as discussed in Appendix A.3. In contrast, we consider the weaker Assumption $(\mathbf{A}_1)$ on the Stein kernel: there exists a map $L(\cdot) \in C^0(\mathbb{R}^d; \mathbb{R}_+)$ such that, for any $y \in \mathbb{R}^d$, the maps $x \mapsto \nabla_1 k_\pi(x, y)$ and $x \mapsto \nabla_1 k_\pi(y, x)$ are $L(y)$-Lipschitz.

We shall assume below the convexity of $L(\cdot)$ and the boundedness of its L^2-norm along the trajectory, i.e. $(\|L\|_{L^2(\mu_n)})_{n\geq 0}$ is bounded. An example of more explicit, though tighter, assumptions are that L satisfies a subpolynomial growth at infinity, i.e. there exists $R > 0$, $c > 0$ and $m \in \mathbb{N}$ such that $L(x) \leq \tilde{L}(x) = c\|x\|^m$ for $x \in \mathbb{R}^d$ with $\|x\| \geq R$, combined with boundedness of the corresponding moment along the flow, i.e. there exists M_m such that $\mathbb{E}_{x\sim\mu_n}\|x\|^{2m} \leq M_m$. In this case, instead of the continuous L, one can consider the convex and continuous function defined by $\tilde{L}(x) = c\|x\|^m + \sup_{y\in B(0,R)} L(y)$.

Proposition 14. Suppose Assumption $(\mathbf{A}_1)$ holds and that $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$. Assume additionally that $L(\cdot)$ is convex, belongs to $L^2(\mu_n)$, and that $\|L(\cdot)\|_{L^2(\mu_n)} \leq M$ for any $n \geq 0$, where μ_n is defined by (10). Then, for any $\gamma \leq \frac{1}{M}$:

$$\mathcal{F}(\mu_{n+1}) - \mathcal{F}(\mu_n) \leq -\gamma\left(1 - \gamma M\right)\|\nabla_{W_2}\mathcal{F}(\mu_n)\|^2_{L^2(\mu_n)} \leq 0. \tag{32}$$

See the proof in Appendix B.12.

B. Proofs

We shall often use that, for smooth and symmetric k, $\nabla_2 k(x, y) = \nabla_1 k(y, x)$, for any $x, y \in \mathbb{R}^d$. Moreover if k is translation-invariant, $\nabla_1 k(x, y) = -\nabla_2 k(x, y)$.

Remark 2. For any $k \in C^2(\mathbb{R}^d \times \mathbb{R}^d, \mathbb{R})$ and π such that $s \in C^1(\mathbb{R}^d)$, k_π and its gradient are written

$$k_\pi(x, y) = s(x)^\top s(y)k(x, y) + s(x)^\top\nabla_2 k(x, y) + s(y)^\top\nabla_1 k(x, y) + \boldsymbol{\nabla}\cdot_1 \nabla_2 k(x, y). \tag{33}$$

$$\begin{aligned}\nabla_2 k_\pi(x, y) = {} & s(x)^\top s(y)\nabla_2 k(x, y) + Js(y)^\top s(x)k(x, y) + H_2 k(x, y)s(x) \\ & + Js(y)^\top\nabla_1 k(x, y) + \nabla_2(\boldsymbol{\nabla}\cdot_1 \nabla_2 k(x, y)).\end{aligned} \tag{34}$$

If k is translation-invariant, since $\nabla_1 k(x, y) = -\nabla_2 k(x, y)$, we have:

$$k_\pi(x, y) = s(x)^\top s(y)k(x, y) + (s(x) - s(y))^\top\nabla_2 k(x, y) + \boldsymbol{\nabla}\cdot_1 \nabla_2 k(x, y), \tag{35}$$

$$\begin{aligned}\nabla_2 k_\pi(x, y) = {} & s(x)^\top s(y)\nabla_2 k(x, y) + Js(y)^\top s(x)k(x, y) + H_1 k(x, y)s(x) \\ & - Js(y)^\top\nabla_2 k(x, y) - \nabla_2(\boldsymbol{\nabla}\cdot_2 \nabla_2 k(x, y)).\end{aligned} \tag{36}$$

Notice that, in this case, odd derivatives of k vanish for $x = y$.

B.1. Proof of Lemma 1

To prove that Assumption $(\mathbf{A}_1)$ holds, we want to show that $x \mapsto \nabla_1 k_\pi(x, y)$ and $x \mapsto \nabla_2 k_\pi(x, y)$ are $L(y)$-Lipschitz for some $L(\cdot) \in C^0(\mathbb{R}^d, \mathbb{R}_+)$ that is μ-integrable for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, so L should have a quadratic growth at most. We will leverage the regularity of k and s to write this problem as that of upper-bounding over $\mathbb{R}^d$ the gradients of these functions (which are related to the Hessian of k_π). To prove that Assumption ($\mathbf{A}_4$) holds, we want on the other hand to identify an integrable function that upper bounds $\|H_1 k_\pi(x,y)\|_{op}$.

We denote by $Js(x)$ the Jacobian of s at x, which is also the Hessian of $\log(\pi)$, and by $Hs(x)$ the Hessian of s at x, which is also the tensor of third derivatives of $\log(\pi)$, i.e. $Hs(x)_{ijk} = \frac{\partial^3 \log(\pi)}{\partial x_i \partial x_j \partial x_k}(x)$.

We compute the gradient of the first term in (33):

$$\nabla_x \left(s(x)^\top s(y) k(x,y) \right) = Js(x)^\top s(y) k(x,y) + s(x)^\top s(y) \nabla_1 k(x,y). \tag{37}$$

Differentiating a second time, we get that

$$H_1 \left(s(x)^\top s(y) k(x,y) \right) = Hs(x)s(y)k(x,y) + \nabla_1 k(x,y) s(y)^\top Js(x) + \left(Js(x)^\top s(y) \right) \nabla_1 k(x,y)^\top + s(x)^\top s(y) H_1 k(x,y),$$

where $Hs(x)s(y)$ is a $d \times d$ matrix of entries $[Hs(x)s(y)]_{jk} = \sum_{i=1}^n Hs(x)_{ijk} s(y)_i$.

Then, the gradient of the second term in (33) is

$$\nabla_x \left(s(x)^\top \nabla_2 k(x,y) \right) = Js(x)^\top \nabla_2 k(x,y) + \nabla_{1,2} k(x,y) s(x), \tag{38}$$

where $\nabla_{1,2} k(x,y) = [\partial_{x_i} \partial_{y_j} k(x,y)]_{i,j}$, and its Hessian is given by:

$$H_1 \left(s(x)^\top \nabla_2 k(x,y) \right) = Hs(x) \nabla_2 k(x,y) + Js(x)^\top \nabla_{1,2} k(x,y) + \nabla_{1,2} k(x,y) Js(x) + \nabla_{1,1,2} k(x,y) s(x), \tag{39}$$

where $\nabla_{1,1,2} k(x,y) = [\partial_{x_i, x_j} \partial_{y_l} k(x,y)]_{i,j,l}$ is a tensor of third derivatives of k.

The Hessian of the last two terms in (33) is straightforward to compute. Hence, collecting all the terms together, we derive that

$$\begin{aligned} H_1 k_\pi(x,y) = & Hs(x)s(y)k(x,y) + \nabla_1 k(x,y) s(y)^\top Js(x) + \left(Js(x)^\top s(y) \right) \nabla_1 k(x,y)^\top + s(x)^\top s(y) H_1 k(x,y) \\ & Hs(x) \nabla_2 k(x,y) + Js(x)^\top \nabla_{1,2} k(x,y) + \nabla_{1,2} k(x,y) Js(x) + \nabla_{1,1,2} k(x,y) s(x) \\ & + \nabla_{1,1,1} k(x,y) s(y) + H_1 \operatorname{Tr}(\boldsymbol{\nabla} \cdot_2 \nabla_1 k(x,y)). \end{aligned} \tag{40}$$

In this expression, the only problematic terms to upper bound $H_1 k_\pi$ uniformly w.r.t. x are the ones where $s(x)$ appears (since $Js(x)$ and $Hs(x)$ are bounded by assumption). However, since s is C_1-Lipschitz, we have that $\|s(x)\| \leq \|s(y)\| + C_1 \|x - y\|$. We then use that the kernel, its derivatives, and the derivatives up to order 3 multiplied by $\|x - y\|$ are bounded for all x and y by a given $B \geq 0$. Hence, for instance,

$$\|s(x)^\top s(y) H_1 k(x,y)\| \leq \|s(y)\| \cdot (\|s(y)\| + C_1 \|x - y\|) \cdot \|H_1 k(x,y)\| \leq B \|s(y)\|^2 + C_1 B \|s(y)\|.$$

As $\|Js(x)\|_{op} \leq C_1$ and $\|Hs(x)\|_{op} \leq C_2$ for all $x \in \mathbb{R}^d$, we have that $\|s(x)\| \leq C_1 \|x\| + \|s(0)\|$ and a similar computation gives that

$$\begin{aligned} \|H_1 k_\pi(x,y)\|_{op} \leq & C_2 B \|s(y)\| + 2C_1 B \|s(y)\| + B \|s(y)\|^2 + C_1 B \|s(y)\| \\ & + C_2 B + 2C_1 B + B \|s(y)\| + C_1 B + B \|s(y)\| + B \end{aligned} \tag{41}$$

$$\leq \tilde{C}_2 \|y\|^2 + \tilde{C}_0, \tag{42}$$

where $\tilde{C}_2 \geq 0$ and $\tilde{C}_0 \geq 0$ only depend on B, C_2, C_1 and $\|s(0)\|$. Consequently, the function $(x,y) \mapsto \|H_1 k_\pi(x,y)\|_{op}$ is integrable for any $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, so Assumption ($\mathbf{A}_3$) holds.[5]

Similarly, one can show based on (34) that

$$\begin{aligned} \nabla_{1,2} k_\pi(x,y) = & (Js(x)^\top s(y)) \nabla_2 k(x,y)^\top + s(x)^\top s(y) \nabla_{1,2} k(x,y) + Js(y)^\top Js(x) k(x,y) + \nabla_1 k(x,y) s(x)^\top Js(y) \\ & + H_2 k(x,y) Js(x) + \nabla_{1,2,2} k(x,y) s(x) + Js(y)^\top H_1 k(x,y) + \nabla_{1,2} \operatorname{Tr}(\boldsymbol{\nabla} \cdot_2 \nabla_1 k(x,y)). \end{aligned}$$

[5] We could do without the assumption of Lipschitzianity of s to prove that Assumption ($\mathbf{A}_3$) holds. We could also merely require that the derivatives of k are bounded, without considering the multiplication by $\|x - y\|$. As a matter of fact, assuming that $\|Hs(x)\|_{op} \leq C_2$ for all $x \in \mathbb{R}^d$ implies that $\|Js(x)\|_{op} \leq C_2 \|x\| + C_1$ and $\|s(x)\| \leq C_2 \|x\|^2 + C_0$ for some $C_0 \geq 0$ and $C_1 \geq 0$. Hence (40) yields an upper bound that is quadratic in $\|x\|$ and $\|y\|$, so $\mu \otimes \nu$-integrable as in Assumption ($\mathbf{A}_3$).

Using the same inequalities that led to (42), we can find two constants $\hat{C}_2 \geq 0$ and $\hat{C}_0 \geq 0$ such that

$$\|\nabla_{1,2} k_\pi(x,y)\|_{op} \leq \hat{C}_2 \|y\|^2 + \hat{C}_0.$$

Setting $L(y) = \max(\tilde{C}_2, \hat{C}_2)\|y\|^2 + \max(\tilde{C}_0, \hat{C}_0)$, the functions $x \mapsto \nabla_1 k_\pi(x,y)$ and $x \mapsto \nabla_2 k_\pi(x,y)$ are $L(y)$-Lipschitz since $L(y)$ is an upper bound of the norm of their Jacobians. Furthermore, $L(\cdot) \in C^0(\mathbb{R}^d, \mathbb{R}_+)$, and is μ-integrable for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$, so Assumption **($\mathbf{A}_1$)** holds.

To prove that Assumption **($\mathbf{A}_4$)** holds, notice that, based on (33),

$$\begin{aligned}|k_\pi(x,x)|_{op} =& \left| \|s(x)\|^2 k(x,x) + s(x)^\top \nabla_2 k(x,x) + s(x)^\top \nabla_1 k(x,x) + \boldsymbol{\nabla} \cdot_1 \nabla_2 k(x,x) \right| \\ \leq& B\left(\|s(x)\|^2 + 2\|s(x)\| + 1\right).\end{aligned}$$

Since, for all $x \in \mathbb{R}^d$, $\|s(x)\| \leq C_1\|x\| + \|s(0)\|$, the function $x \mapsto k_\pi(x,x)$ is bounded by a quadratic function and is thus μ-integrable for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$. This shows that Assumption **($\mathbf{A}_4$)** holds.

We now prove that Assumption **($\mathbf{A}_2$)** is also satisfied if there exists $M > 0$ and M_0 such that, for all $x \in \mathbb{R}^d$, $\|s(x)\| \leq M\sqrt{\|x\|} + M_0$. Then, based on (34) and as $\sqrt{\|x\|} \leq \|x\| + 1$,

$$\begin{aligned}\|\nabla_2 k_\pi(x,y)\| &\leq \|s(x)\| \cdot \|s(y)\| \cdot \|\nabla_2 k(x,y)\| + BC_1\|s(x)\| + B\|s(x)\| + BC_1 + B \\ &\leq (\|s(y)\| + C_1\|x-y\|) \cdot \|s(y)\| \cdot \|\nabla_2 k(x,y)\| + (BC_1 + B) \cdot (M\sqrt{\|x\|} + M_0 + 1) \\ &\leq B\|s(y)\|^2 + BC_1\|s(y)\| + (BC_1 + B) \cdot (M\|x\| + M + M_0 + 1) \\ &\leq B(M\sqrt{\|y\|} + M_0)^2 + BC_1(M\|y\| + M + M_0) + (BC_1 + B) \cdot (M\|x\| + M + M_0 + 1) \\ &\leq m(1 + \|y\| + \|x\|),\end{aligned}$$

for some $m > 0$ depending on B, M, C_1 and M_0. Integrating over μ for any $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ gives that $\| \int \nabla_2 k_\pi(x,y) \mathrm{d}\mu(x)\| \leq m\left(1 + \|y\| + \int \|x\|\, d\mu(x)\right)$, so Assumption **($\mathbf{A}_2$)** holds, which concludes the proof.

B.2. Proof of Proposition 2

Lemma 15. Suppose Assumption **($\mathbf{A}_1$)** holds. The first variation of $\mathcal{F}$ evaluated at $\mu \in \mathcal{P}_2(\mathbb{R}^d)$ is then the function defined for any $y \in \mathbb{R}^d$ by:

$$\frac{\partial \mathcal{F}(\mu)}{\partial \mu}(y) = \int_{\mathbb{R}^d} k_\pi(x,y) \mathrm{d}\mu(x).$$

Proof. Let $\mu, \nu \in \mathcal{P}_2(\mathbb{R}^d)$, and $\xi = \mu - \nu$. Using the symmetry of k_π, we get

$$\begin{aligned}\frac{1}{\epsilon}\left[\mathcal{F}(\mu + \epsilon\xi) - \mathcal{F}(\mu)\right] &= \frac{1}{2\epsilon}\left[\iint_X k_\pi(x,y) d(\mu+\epsilon\xi)(x) d(\mu+\epsilon\xi)(y) - \iint_X k_\pi(x,y) d\mu(x) d\mu(y)\right] \\ &= \iint_{\mathbb{R}^d} k_\pi(x,y) d\mu(x) d\xi(y) + \frac{\epsilon}{2}\iint_X k_\pi(x,y) d\xi(x) d\xi(y)\end{aligned}$$

Hence,

$$\lim_{\epsilon \to 0} \frac{1}{\epsilon}(\mathcal{F}(\mu + \epsilon\xi) - \mathcal{F}(\mu)) = \iint_{\mathbb{R}^d} k_\pi(x,y) d\mu(x) d\xi(y).$$

□

Assume μ satisfies the assumptions of Definition 4. To obtain the expression for the W_2 gradient of $\mathcal{F}$, we first need to exchange the integration and the gradient with respect to y. Since $y \mapsto \nabla_2 k_\pi(x,y)$ is $L(x)$-Lipschitz, we have for all $v \in \mathbb{R}^d, 0 < \epsilon \leq 1$:

$$\left| \frac{k_\pi(x, y+\epsilon v) - k_\pi(x,y)}{\epsilon} - \langle \nabla_2 k_\pi(x,y), v\rangle \right| \leq \frac{L(x)\epsilon\|v\|_2^2}{2} \leq \frac{L(x)\|v\|_2^2}{2}$$

where the right hand side is integrable. Therefore by the Lebesgue dominated convergence theorem, we have the following interchange of the gradient and the integral when computing the W_2-gradient:

$$\nabla_{W_2}\mathcal{F}(\mu)(y) = \nabla \frac{\partial \mathcal{F}(\mu)}{\partial \mu}(y) = \int \nabla_2 k_\pi(x,y) \mathrm{d}\mu(x). \tag{43}$$

Then, using the continuity equation (7) and an integration by parts,[6] the dissipation of $\mathcal{F}$ along its W_2 gradient flow is obtained as follows:

$$\frac{\mathrm{d}\mathcal{F}(\mu_t)}{\mathrm{d}t} = \int \frac{\partial \mathcal{F}(\mu_t)}{\partial \mu_t}\frac{\partial \mu_t}{\partial t} = \int \frac{\partial \mathcal{F}(\mu_t)}{\partial \mu} \operatorname{div}\left(\mu_t \nabla \frac{\partial \mathcal{F}(\mu_t)}{\partial \mu}\right) = -\int \left\|\nabla \frac{\partial \mathcal{F}(\mu_t)}{\partial \mu}\right\|^2 \mathrm{d}\mu_t.$$

Plugging (43) in the r.h.s. of this equality leads to the final formula.

B.3. Proof of Proposition 3

We compute the second time derivative $\ddot{\mathcal{F}}(\rho_t)$ where ρ_t is a path from μ to $(I + \nabla\psi)_{\#}\mu$ given by: $\rho_t = (I + t\nabla\psi)_{\#}\mu$, for all $t \in [0, 1]$. By Lemma 22, $\ddot{\mathcal{F}}(\rho_t)$ is well-defined and $\ddot{\mathcal{F}}(\mu) = \ddot{\mathcal{F}}(\rho_t)|_{t=0}$ is given by:

$$\ddot{\mathcal{F}}(\mu) = \int \left[\nabla\psi^\top(z)\nabla_1\nabla_2 k_\pi(z, z')\nabla\psi(z')\right] d\mu(z')d\mu(z) + \int \left[\nabla\psi^\top(z) H_1 k_\pi(z, z')\nabla\psi(z)\right] d\mu(z')d\mu(z). \tag{44}$$

B.4. Proof of Corollary 4

Under Assumption ($\mathbf{A}_3$), by the dominated convergence theorem, we can exchange the order of the integral and derivative in the second term of (44):

$$\iint \left[\nabla\psi^\top(z) H_1 k_\pi(z, z')\nabla\psi(z)\right] d\rho_t(z')d\rho_t(z) = \int \left[\nabla\psi^\top(z) H_1 \left(\int k_\pi(z, z')d\rho_t(z')\right)\nabla\psi(z)\right] d\rho_t(z). \tag{45}$$

The latter (45) vanishes when $\rho_t = \pi$, thanks to the property of the Stein kernel : $\int k_\pi(z, z')\mathrm{d}\pi(z') = 0$.

Hence, by considering $\psi \in C_c^\infty(\mathbb{R}^d)$ and a path ρ_t from π to $(I + \nabla\psi)_{\#}\pi$ given by: $\rho_t = (I + t\nabla\psi)_{\#}\pi$, for all $t \in [0, 1]$.,

$$\ddot{\mathcal{F}}(\pi) = \mathrm{Hess}_\pi(\psi, \psi),$$

where

$$\mathrm{Hess}_\pi(\psi, \psi) = \iint \left[\nabla\psi^\top(z)\nabla_1\nabla_2 k_\pi(z, z')\nabla\psi(z')\right] \mathrm{d}\pi(z')\mathrm{d}\pi(z).$$

By an integration by parts[7] w.r.t. x, we have, for $\mathcal{L}_\pi : f \mapsto -\Delta f + \langle -\nabla \log \pi, \nabla f\rangle$,

$$\begin{aligned}
\mathrm{Hess}_\pi(\psi, \psi) &= \sum_{i,j=1}^d \iint \frac{\partial\psi(x)}{\partial x_i} . \frac{\partial^2 k_\pi(x, y)}{\partial x_i \partial y_j} . \frac{\partial\psi(y)}{\partial y_j} \mathrm{d}\pi(x)\mathrm{d}\pi(y) \\
&= -\sum_{i,j=1}^d \iint \frac{\partial^2\psi(x)}{\partial x_i^2} . \frac{\partial k_\pi(x, y)}{\partial y_j} . \frac{\partial\psi(y)}{\partial y_j} \mathrm{d}\pi(x)\mathrm{d}\pi(y) - \sum_{i,j=1}^d \iint \frac{\partial\psi(x)}{\partial x_i} . \frac{\partial \log\pi(x)}{\partial x_i} . \frac{\partial k_\pi(x, y)}{\partial y_j} . \frac{\partial\psi(y)}{\partial y_j} \mathrm{d}\pi(x)\mathrm{d}\pi(y) \\
&= \sum_{j=1}^d \iint \mathcal{L}_\pi\psi(x) . \frac{\partial k_\pi(x, y)}{\partial y_j} . \frac{\partial\psi(y)}{\partial y_j} \mathrm{d}\pi(x)\mathrm{d}\pi(y).
\end{aligned}$$

We then repeat the same steps, performing an integration by parts w.r.t. y_j and using Lemma 21 in the last equality,

$$\mathrm{Hess}_\pi(\psi, \psi) = \int \mathcal{L}_\pi\psi(x) k_\pi(x, y) \mathcal{L}_\pi\psi(y) \mathrm{d}\pi(x)\mathrm{d}\pi(y) = \|S_{\pi,k_\pi}(\mathcal{L}_\pi\psi)\|^2_{\mathcal{H}_{k_\pi}}.$$

Notice that under Assumption ($\mathbf{A}_4$), we have $\mathcal{H}_{k_\pi} \subset L^2(\pi)$, so S_{π,k_π} and its adjoint are well-defined (see Section 2.3).

[6] By the regularity of μ, from the assumptions of Definition 4, there are no boundary terms in the integration.

[7] First, differentiate the product $\frac{\partial\psi(x)}{\partial x_i}\pi(x)$, then integrate $\frac{\partial^2 k_\pi(x,y)}{\partial x_i \partial y_j}$ w.r.t. x_i. Since ψ is compactly supported, it vanishes at infinity. Since π has a C^1-density w.r.t. the Lebesgue measure, any integral over $\mathrm{d}\pi$ bearing on the boundary of the support of π vanishes. We thus do not have any extra integral on the boundary when performing the integration by parts.

B.5. Proof of Proposition 5

Denote by $L^2_0(\pi)$ the closed subspace of $L^2(\pi)$ consisting of functions ϕ such that $\int \phi(x)d\pi(x) = 0$. For any $f \in \mathcal{D}(\mathcal{L}_\pi)$, $\int \mathcal{L}_\pi f(x)d\pi(x) = 0$; hence $\mathrm{Im}(\mathcal{L}_\pi)$ is a subset of $L^2_0(\pi)$, which is itself a subset of $L^2(\pi)$.

Let $T_{\pi,k_\pi} = S^*_{\pi,k_\pi} \circ S_{\pi,k_\pi}$, and assume that $V = -\log(\pi)$ is a $C^1(\mathbb{R}^d)$ function. We want to show that exponential decay near equilibrium (16) holds, if and only if $\mathcal{L}_\pi^{-1} : L^2_0(\pi) \to L^2_0(\pi)$, the inverse of $\mathcal{L}_\pi|_{L^2_0(\pi)}$, is a well-defined linear operator, and

$$\langle \phi, T_{\pi,k_\pi}\phi\rangle_{L_2(\pi)} \geq \lambda \langle \phi, \mathcal{L}_\pi^{-1}\phi\rangle_{L_2(\pi)} \tag{46}$$

holds for all $\phi \in L^2_0(\pi)$.

Proof. First, assume that the exponential convergence near equilibrium holds. We will show that $\mathcal{L}_\pi^{-1}$ is a well-defined bounded linear operator on $L^2_0(\pi)$, and then that the inequality 46 holds. Let $\psi \in C^\infty_c(\mathbb{R}^d)$. By Corollary 4, the Hessian of $\mathcal{F}$ at π can be written as

$$\mathrm{Hess}_\pi(\psi,\psi) = \langle S_{\pi,k_\pi}(\mathcal{L}_\pi\psi), S_{\pi,k_\pi}\mathcal{L}_\pi\psi\rangle_{\mathcal{H}_{k_\pi}} = \langle T_{\pi,k_\pi}(\mathcal{L}_\pi\psi), \mathcal{L}_\pi\psi\rangle_{L^2(\pi)},$$

where $T_{\pi,k_\pi} = S^*_{\pi,k_\pi} \circ S_{\pi,k_\pi}$. By Lemma 12, we have that $\|\nabla\psi\|^2_{L^2(\pi)} = \langle \psi, \mathcal{L}\psi\rangle_{L^2(\pi)}$, and exponential convergence near equilibrium can be written:

$$\langle T_{\pi,k_\pi}(\mathcal{L}_\pi\psi), \mathcal{L}_\pi\psi\rangle_{L^2(\pi)} \geq \lambda\langle \psi, \mathcal{L}_\pi\psi\rangle_{L^2(\pi)}. \tag{47}$$

Note, that by Assumption ($\mathbf{A}_4$), $T_{k_\pi,\pi}$ is a bounded operator (Steinwart & Christmann, 2008, Theorem 4.27), and it follows by an application of the Cauchy-Schwartz inequality that (47) implies:

$$\|\mathcal{L}_\pi\psi\|^2_{L^2(\pi)} \geq \frac{\lambda}{\|T_{k_\pi,\pi}\|_{op}}\langle \psi, \mathcal{L}_\pi\psi\rangle_{L^2(\pi)}, \tag{48}$$

where $\|\cdot\|_{op}$ denotes the operator norm. Now, let ψ be an arbitrary element of $\mathcal{D}(\mathcal{L}_\pi)$, the domain of $\mathcal{L}_\pi$. By Lemma 12, there exists a sequence $(\psi_n)_{n=1}^\infty \subset C^\infty_c(\mathbb{R}^d)$ converging strongly to ψ, such that $\mathcal{L}_\pi\psi_n$ converges strongly to $\mathcal{L}_\pi\psi$ as well. Hence, (48) holds for all $\psi \in \mathcal{D}(\mathcal{L}_\pi)$.

We will now show, that the spectrum of $\mathcal{L}_\pi$, $\sigma(\mathcal{L}_\pi)$ is contained in $\{0\} \cup [\frac{\lambda}{\|T_{k_\pi,\pi}\|_{op}}, \infty)$. Suppose that there exists a $\sigma \in (0, \frac{\lambda}{\|T_{k_\pi,\pi}\|_{op}}) \cap \sigma(\mathcal{L}_\pi)$. If σ is in the point spectrum of $\mathcal{L}_\pi$, then by definition there would exist a vector $v \in \mathcal{D}(\mathcal{L}_\pi)$ such that $\mathcal{L}_\pi v = \sigma v$, which would contradict the inequality (48). On the other hand, if σ is not in the point spectrum of $\mathcal{L}_\pi$, then by Weyl's criterion (Pankrashkin, 2014, Theorem 7.22) we can find an orthonormal sequence $(u_n)_{n=1}^\infty \in \mathcal{D}(\mathcal{L}_\pi)$ such that $(\mathcal{L}_\pi - \sigma)u_n$ converges to 0 in $L_2(\pi)$. An obvious calculation shows that this would contradict (48). Hence, $\sigma(\mathcal{L}_\pi) \subset \{0\} \cup [\frac{\lambda}{\|T_{k_\pi,\pi}\|}, \infty)$.

We note that $L^2_0(\pi)$ is itself a Hilbert space with the inner product inherited from $L^2(\pi)$. The image of $\mathcal{L}_\pi$ is contained in $L^2_0(\pi)$ (recall that for any $f \in \mathcal{D}(\mathcal{L}_\pi)$, $\int \mathcal{L}_\pi f(x)d\pi(x) = 0$), and it is dense in $L^2_0(\pi)$ since $\overline{\mathrm{Im}(T)} = (\mathrm{Ker}(T))^\perp$ for self-adjoint operators T (Brezis, 2010, Corollary 2.18). Furthermore, since $\sigma(\mathcal{L}_\pi) \subset \{0\} \cup [\frac{\lambda}{\|T_{k_\pi,\pi}\|_{op}}, \infty)$ and the kernel of $\mathcal{L}_\pi$ consists of constants (Appendix A.5), the operator $\tilde{\mathcal{L}}_\pi = \mathcal{L}_\pi|_{L^2_0(\pi)}$ (the restriction of $\mathcal{L}_\pi$ to $L^2_0(\pi)$) is strictly positive and self-adjoint on $L^2_0(\pi)$. Since consequently $\sigma(\mathcal{L}_\pi^{-1}) \in (0, \frac{\|T_{k_\pi,\pi}\|_{op}}{\lambda}]$, $\mathcal{L}_\pi^{-1} := \tilde{\mathcal{L}}_\pi^{-1}$ is a well-defined, bounded, self-adjoint and positive operator from $L^2_0(\pi)$ to $L^2_0(\pi)$. We now take an arbitrary $\psi \in \mathcal{D}(\mathcal{L}_\pi)$ and denote $\phi = \mathcal{L}_\pi\psi$. We can write $\psi = \psi' + C$, where $C = \int \psi(x)d\pi(x)$ is a constant and $\psi' \in L^2_0(\pi) \cap \mathcal{D}(\mathcal{L}_\pi)$. We then have $\mathcal{L}_\pi\psi = \mathcal{L}_\pi\psi'$ and $\mathcal{L}_\pi^{-1}\mathcal{L}_\pi\psi = \psi' = \psi - C$. We also note, that since $\phi \in \mathrm{Im}(\mathcal{L}_\pi) \subset L^2_0(\pi)$ we have $\langle C, \phi\rangle_{L^2(\pi)} = 0$. Now by a direct substitution of $\phi = \mathcal{L}_\pi\psi$ in (47) we obtain:

$$\langle T_{\pi,k_\pi}\phi, \phi\rangle_{L^2(\pi)} \geq \lambda\langle \psi' + C, \phi\rangle_{L^2(\pi)} = \lambda\langle \mathcal{L}_\pi^{-1}\phi, \phi\rangle_{L^2(\pi)}. \tag{49}$$

for all ϕ in the image of $\mathcal{L}_\pi$. Given that this image is dense in $L^2_0(\pi)$, and that $T_{k_\pi,\pi}$ and $\mathcal{L}_\pi$ are continuous, this is equivalent to (49) holding for all $\phi \in L^2_0(\pi)$.

Now we prove the reverse implication, that is that if $\mathcal{L}_\pi^{-1}$ is well-defined, bounded, and (49) holds for all $\phi \in L^2_0(\pi)$, then (47) holds for all $\psi \in C^\infty_c(\mathbb{R}^d)$. This follows trivially from the fact, that for $\psi \in C^\infty_c(\mathbb{R}^d)$ we have $\mathcal{L}_\pi\psi \in L^2_0(\pi)$ and $\mathcal{L}_\pi^{-1}\mathcal{L}_\pi\psi = \psi - C$ where $C = \int \psi(x)d\pi(x)$ as previously. Again, using the fact that $\langle C, \mathcal{L}_\pi\psi\rangle_{L^2(\pi)} = 0$ and a direct substitution $\phi = \mathcal{L}_\pi\psi$ into (49), we obtain (47) for all $\psi \in C^\infty_c(\mathbb{R}^d)$. □

B.6. Proof of Corollary 6

We begin this section by an additional result, that shows that the assumption that the spectrum of $\mathcal{L}_\pi^{-1}$ is necessarily discrete if exponential decay near equilibrium (16) holds and the RKHS of k_π is infinite dimensional.

Lemma 16. If exponential decay near equilibrium (16) holds, and the RKHS for k_π is infinite dimensional, then $\mathcal{L}_\pi^{-1}$ has a discrete spectrum.

Proof. By Proposition 5, exponential convergence near equilibrium implies that $\mathcal{L}_\pi^{-1}$ is a well-defined bounded linear operator on $L_0^2(\pi)$ and the inequality (18) holds for all $\phi \in L_2^0(\pi)$. Let $(l_n)_{n\in\mathbb{N}}$ be the eigenvalues of T_{k,π_k} in descending order. Using the max-min variational formula for the eigenvalues of a compact operator, and applying Proposition 5, we get:

$$l_n = \sup_{\substack{E\subset L_0^2(\pi)\\ dim(E)=n}} \inf_{\substack{\phi\in E\\ \phi\neq 0}} \frac{\langle \phi, T_{k,\pi_k}\phi\rangle_{L^2(\pi)}}{\langle \phi,\phi\rangle_{L^2(\pi)}} \geq \lambda \sup_{\substack{E\subset L_0^2(\pi)\\ dim(E)=n}} \inf_{\substack{\phi\in E\\ \phi\neq 0}} \frac{\langle \phi, \mathcal{L}_\pi^{-1}\phi\rangle_{L^2(\pi)}}{\langle \phi,\phi\rangle_{L^2(\pi)}}. \tag{50}$$

We will now show, that this implies that the spectrum of $\mathcal{L}_\pi^{-1}$ is discrete. Let $\mathcal{L}_\pi^{1/2}$ be the square root of $\mathcal{L}_\pi|_{L_0^2(\pi)}$, the restriction of $\mathcal{L}_\pi$ to $L_0^2(\pi)$. As a consequence of the analysis in Appendix B.5, the operator $\mathcal{L}_\pi^{1/2}$ is well-defined, strictly positive and self-adjoint from $\mathcal{D}(\mathcal{L}_\pi)\cap L_0^2(\pi)$ to $L_0^2(\pi)$. By retricting the supremum to ϕ of the form $\phi = \mathcal{L}_\pi^{1/2}\psi$, where $\psi \in \mathcal{D}(\mathcal{L}_\pi)\cap L_0^2(\pi)$, we obtain the following lower bound:

$$\begin{aligned}
\sup_{\substack{E\subset L_0^2(\pi)\\ dim(E)=n}} \inf_{\substack{\phi\in E\\ \phi\neq 0}} \frac{\langle \phi, \mathcal{L}_\pi^{-1}\phi\rangle_{L^2(\pi)}}{\langle \phi,\phi\rangle_{L^2(\pi)}} &\geq \sup_{\substack{E\subset \mathcal{L}_\pi^{1/2}(L_0^2(\pi)\cap\mathcal{D}(\mathcal{L}_\pi))\\ dim(E)=n}} \inf_{\substack{\phi\in E\\ \phi\neq 0}} \frac{\langle \phi, \mathcal{L}_\pi^{-1}\phi\rangle_{L^2(\pi)}}{\langle \phi,\phi\rangle_{L^2(\pi)}} \\
&= \sup_{\substack{E\subset L_0^2(\pi)\cap\mathcal{D}(\mathcal{L}_\pi)\\ dim(E)=n}} \inf_{\substack{\psi\in E\\ \psi\neq 0}} \frac{\langle \psi,\psi\rangle_{L^2(\pi)}}{\langle \psi, \mathcal{L}_\pi\psi\rangle_{L^2(\pi)}} = \left(\inf_{\substack{E\subset L_0^2(\pi)\cap\mathcal{D}(\mathcal{L}_\pi)\\ dim(E)=n}} \sup_{\substack{\psi\in E\\ \psi\neq 0}} \frac{\langle \psi, \mathcal{L}_\pi\psi\rangle_{L^2(\pi)}}{\langle \psi,\psi\rangle_{L^2(\pi)}} \right)^{-1}
\end{aligned} \tag{51}$$

It now follows that $\mathcal{L}_\pi|_{L_0^2(\pi)}$ has discrete spectrum. Indeed, if it was not the case, by (Pankrashkin, 2014, Theorem 8.1) the quotients within the parenthesis on the right hand side of (51), would be upper bounded by $\inf \sigma_{\text{ess}}(\mathcal{L}_\pi|_{L_0^2(\pi)}) > 0$, where σ_{ess} is the essential spectrum, and thus the right hand side of (51) would be lower bounded by a constant equal to $\left(\inf \sigma_{\text{ess}}(\mathcal{L}_\pi|_{L_0^2(\pi)})\right)^{-1} > 0$. This is not possible since the eigenvalues $(l_n)_{n\in\mathbb{N}}$ converge to zero. Since $\mathcal{L}_\pi$ is positive and unbounded on $L_0^2(\pi)$ and its spectrum is purely discrete, it follows that $\mathcal{L}_\pi^{-1}$ is compact and, by extension, $(\mathcal{L}_\pi + I)^{-1} : L^2(\pi) \to L^2(\pi)$ is compact, which means that $\mathcal{L}_\pi$ has compact resolvent. □

The proof of Corollary 6 now follows readily:

Proof of Corollary 6. We denote by $(l_n)_{n\in\mathbb{N}}$ the eigenvalues of T_{k,π_k} in descending order and by $(\lambda_n)_{n\in\mathbb{N}}$ the eigenvalues of compact operator $\mathcal{L}_\pi^{-1} : L_0^2(\pi) \to L_0^2(\pi)$ also in descending order. By max-min variational characterization of eigenvalues for compact operators, it follows from (50) that $l_n \geq \lambda.\lambda_n$, hence $\lambda_n = \mathcal{O}(l_n)$. □

B.7. Bound on eigenvalue decay and proof of Theorem 7

Lemma 17. Let $\gamma_d \sim \mathcal{N}(0, I_d)$ be the standard d-dimensional Gaussian measure, and let $\mathcal{L}_{\gamma_d} = -\Delta + \langle x, \nabla\rangle$ be the Ornstein-Uhlenbeck operator on $L^2(\gamma_d)$. If we denote by ρ_n the n-th smallest eigenvalue of $\mathcal{L}_{\gamma_d}$, then we have:

$$\rho_n = \mathcal{O}(n^{1/d}).$$

Proof. For a multi-index $\alpha = (k_1, \ldots, k_d)$, the multivariate Hermite polynomial is for any $x = (x_1, \ldots, x_d) \in \mathbb{R}^d$:

$$H_\alpha(x) = \prod_{i=1}^{d} H_{k_i}(x_i)$$

where for $k_i \in \mathbb{N}$, H_{k_i} denotes the usual one-dimensional k_ith-order Hermite polynomial. It is well known that multivariate Hermite polynomials form an orthogonal basis of $L^2(\gamma_d)$ and that we have

$$\mathcal{L}_{\gamma_d} H_\alpha = |\alpha| H_\alpha$$

where $|\alpha| = \sum_{i=1}^d k_i$ (Bakry et al., 2013, Section 2.7.1). Therefore any $k \in \mathbb{N}$ is an eigenvalue of $\mathcal{L}_{\gamma_d}$, with multiplicity equal to

$$S_k = \binom{k+d-1}{d-1}$$

which is the number of solutions of the equation $k = \sum_{i=1}^d k_i$, where $\{k_i\}_{i=1}^d$ takes its values in the set of nonnegative integers. This means, that

$$\rho_n = k \iff \sum_{i=1}^{k-1} S_i < n \leq \sum_{i=1}^{k} S_i.$$

Since S_i is a polynomial in i of degree $d-1$, then $\sum_{i=1}^k S_i$ is a polynomial in k of degree d. Therefore we have $\rho_n = \mathcal{O}(n^{1/d})$. □

Corollary 18. The conclusion of Lemma 17 holds for any Schrödinger operator on $L_2(\mathbb{R}^d)$, defined for $L > 0$ as follows,

$$\mathcal{H}_{\nu_L} := -\Delta + \frac{1}{4} L^2 \|x\|^2 - \frac{1}{2} dL. \tag{52}$$

Proof. Let $\nu_L \sim \mathcal{N}(0, \frac{1}{L} I_d)$ be a normal measure, and let $\mathcal{L}_{\nu_L} = -\Delta + L\langle x, \nabla \rangle$ be the associated Ornstein-Uhlenbeck operator on $L_2(\nu_L)$. It is easy to see that the map $R_L : L_2(\gamma_d) \to L_2(\nu_L)$ given by:

$$(R_L \phi)(x) = \phi(\sqrt{L} x)$$

is unitary, and that $\mathcal{L}_{\gamma_d} = R_L^* \mathcal{L}_{\nu_L} R_L$. Furthermore, it is standard that $\mathcal{L}_{\nu_L}$ is unitarily equivalent to $\mathcal{H}_{\nu_L}$, see (Pavliotis, 2014, Proposition 4.7). Since unitary equivalence preserves the spectrum, the thesis follows. □

Lemma 19. Suppose that $\pi \propto e^{-V}$ where V is a $C^2(\mathbb{R}^d)$ potential such that ∇V is L-Lipschitz and assume that $\mathcal{L}_\pi$ has discrete spectrum. If we denote by $\tilde{\lambda}_n$ the n-th smallest eigenvalue of the operator $\mathcal{L}_\pi$ (counting the multiplicities), then

$$\tilde{\lambda}_n \leq \mathcal{O}(n^{1/d})$$

Proof. It is easy to show, that if ∇V is Lipschitz and $\int e^{-V(x)} \mathrm{d}x$ is finite, then $\lim_{|x| \to \infty} V(x) = \infty$. This is a consequence of the fact that $\lim_{R \to \infty} \int_{|x| > R} e^{-V(x)} \mathrm{d}x = 0$ and that on the set $\{y : ||x - y|| \leq 1, \langle \nabla V(x), y - x \rangle \leq 0\}$ we have $-V(y) \geq -V(x) - \frac{L}{2} \|x - y\|_2^2$. Therefore $V(x)$ has to diverge to ∞ as $|x| \to \infty$. It follows, that V attains a minimum on $\mathbb{R}^d$. Assume, without loss of generality, that V attains its (not necessarily unique) minimum at 0. The operator $\mathcal{L}_\pi$ is unitarily equivalent to the Schrödinger operator (Equation 4.130, Pavliotis (2014))

$$\mathcal{H}_\pi = -\Delta + \frac{1}{4} \|\nabla V\|^2 - \frac{1}{2} \Delta V$$

on $L_2(\mathbb{R}^d)$. Since unitary equivalence preserves the spectrum, this operator has by assumption a discrete spectrum. We will show that the eigenvalues of $\mathcal{H}_{\nu_L} + dLI$ dominate those of $\mathcal{H}_\pi$ where $\mathcal{H}_{\nu_L}$ is defined in (52). For any $\phi \in D(\mathcal{H}_L)$, we have

$$\begin{aligned} +\infty > \langle \phi, \mathcal{H}_{\nu_L} + dLI\phi \rangle_{L_2(\mathbb{R}^d)} &= \langle \phi, \mathcal{H}_\pi \phi \rangle_{L_2(\mathbb{R}^d)} + \frac{1}{4} \int \left(L^2 \|x\|^2 - \|\nabla V(x)\|^2 + 2dL + 2\Delta V(x) \right) \phi(x)^2 \mathrm{d}x \\ &\geq \langle \phi, \mathcal{H}_\pi \phi \rangle_{L_2(\mathbb{R}^d)}, \end{aligned}$$

where the last inequality follows from the L-Lipschitz regularity of ∇V. It follows from the above calculation that the domain of the quadratic form $\langle \phi, \mathcal{H}_{\nu_L} + dLI \rangle_{L_2(\mathbb{R}^d)}$, denoted $Q(\mathcal{H}_{\nu_L} + dLI)$ is contained in the domain of the quadratic form $\langle \phi, \mathcal{H}_\pi \phi \rangle_{L_2(\mathbb{R}^d)}$, denoted $Q(\mathcal{H}_\pi)$. It is now a simple consequence of the the Rayleigh-Ritz variational formula that the eigenvalues of $\mathcal{H}_{\nu_L} + dLI$ dominate those of $\mathcal{H}_\pi$ (Pankrashkin, 2014, Corollary 8.6), that is, $\tilde{\lambda}_n \leq \rho_n + dL$ where ρ_n is the n-th smallest eigenvalue of the operator $\mathcal{H}_{\nu_L}$. By Corollary 18, we have that $\rho_n = \mathcal{O}(n^{1/d})$, which concludes the proof. □

Proof of Theorem 7. Let $(l_n)_{n=1}^{\infty}$ be the eigenvalues of $T_{k_\pi,\pi}$ and $(\lambda_n)_{n=1}^{\infty}$ be the eigenvalues of $\mathcal{L}_\pi^{-1}$ both in descending order. Since $T_{k_\pi,\pi}$ is a trace class operator (Steinwart & Christmann, 2008, Theorem 4.27), we know that $\sum_{n=1}^{\infty} l_n < +\infty$. By Corollary 18 exponential convergence near equilibrium would imply that $\sum_{n=1}^{\infty} \lambda_n < +\infty$ as well. On the other hand we have that $\lambda_n = \tilde{\lambda_n}^{-1}$, where $\tilde{\lambda}_n$ are the positive eigenvalues of $\mathcal{L}_\pi$ in a nondecreasing order. By Lemma 19 we have $\tilde{\lambda}_n = \mathcal{O}(n^{1/d})$, which implies that there exists an $N \in \mathbb{N}$ and a constant $C > 0$, such that for all $n \geq N$ we have $\tilde{\lambda}_n \leq Cn^{1/d}$ and thus $\lambda_n \geq Cn^{-1/d}$. This contradicts the summability of λ_n, and concludes the proof. □

B.8. Proof of Lemma 8

Let $\mathcal{H}_0 = \left\{\sum_{i=1}^m \alpha_i k_\pi(\cdot, x_i); m \in \mathbb{N}; \alpha_1, \ldots, \alpha_m \in \mathbb{R}; x_1, \ldots, x_m \in \mathbb{R}^d\right\}$. Recall that $\mathcal{H}_{k_\pi}$ is the set of functions on $\mathbb{R}^d$ for which there exists an $\mathcal{H}_0$-Cauchy sequence $(f_n)_n \in \mathcal{H}_0^{\mathbb{N}}$ converging pointwise to f. Let $c \in \mathbb{R}$. Let $f_0 = \sum_{i=1}^m \alpha_i k_\pi(\cdot, x_i) \in \mathcal{H}_0$ and assume that $f_0 = c$. Integrating f_0 w.r.t. π yields

$$c = \int f_0(x) d\pi(x) = \sum_{i=1}^m \alpha_i \int k_\pi(x_i, x) d\pi(x) = 0.$$

Hence $c = 0$. Similarly, let $f \in \mathcal{H}_{k_\pi}$ such that $f = c$, fix $(f_n)_n \in \mathcal{H}_0^{\mathbb{N}}$ such that $\|f_n - f\|_{\mathcal{H}_{k_\pi}} \to 0$.

Under Assumption (**A**$_4$), the injection $\iota : \mathcal{H}_{k_\pi} \to L^2(\pi)$ is continuous, linear and bounded. Indeed, for any $g \in \mathcal{H}_{k_\pi}$

$$\|\iota g\|_{L^2(\pi)}^2 = \int g(x)^2 d\pi(x) = \int \langle g, k_\pi(x, \cdot)\rangle_{\mathcal{H}_{k_\pi}}^2 d\pi(x) \leq \|g\|_{\mathcal{H}_{k_\pi}}^2 \int k_\pi(x,x) d\pi(x) =: c_\pi^2 \|g\|_{\mathcal{H}_{k_\pi}}^2,$$

where $c_\pi < +\infty$ exists by Assumption (**A**$_4$). So $\|f_n - f\|_{\mathcal{H}_{k_\pi}} \to 0$ implies that $\|f_n - f\|_{L^2(\pi)} \to 0$. Since $\int f_n(x) d\pi(x) = 0$, by the reproducing property and Cauchy-Schwartz inequality, we have:

$$|c| = \left|\int f(x) d\pi(x)\right| = \left|\int (f_n - f)(x) d\pi(x)\right| \leq \int |\langle f_n - f, k_\pi(x, \cdot)\rangle_{\mathcal{H}_{k_\pi}}| d\pi(x) \leq \|f_n - f\|_{\mathcal{H}_{k_\pi}} c_\pi^2.$$

which shows that $c = 0$ since $\|f_n - f\|_{\mathcal{H}_{k_\pi}} \to 0$.

B.9. Proof of Proposition 10

Rather than providing a proof only for the more restrictive smooth submanifolds considered in Proposition 10, we express the result below for general closed nonempty sets. This formulation involves contingent cones which are crucial quantities for studying the invariance of non-smooth sets. They can be informally described as the collection of directions at x that point either inward or that are tangent to the set $\mathcal{M}$. More formally, for $d_{\mathcal{M}}(y)$ the distance of $y \in \mathbb{R}^d$ to $\mathcal{M}$, $T_{\mathcal{M}}(x) := \{v | \liminf_{h \to 0^+} d_{\mathcal{M}}(x + hv)/h = 0\}$. Non-smooth sets were experimentally met whenever we considered Gaussian mixtures with more than three components for which there was no simple axis of symmetry (as on Figure 8). These are cases more intricate than the ones considered in Lemma 11.

Proposition 20. Let $\mathcal{M} \subset \mathbb{R}^d$ be a closed nonempty set and $\mu_0 \in P_c(\mathbb{R}^d)$ with $\text{supp}(\mu_0) \subset \mathcal{M}$. Assume that, for a deterministic $(v_{\mu_t})_{t \geq 0}$ satisfying the Caratheodory-Lipschitz Assumptions (**C**$_1$)-(**C**$_4$), we have $v_{\mu_t}(x) \in T_{\mathcal{M}}(x)$ where $T_{\mathcal{M}}(x)$ is the contingent cone of $\mathcal{M}$ at $x \in \mathcal{M}$. Then $\mathcal{M}$ is flow-invariant for (4).

Proof. Consider any $x_0 \in \text{supp}(\mu_0)$. By Assumption (**C**$_2$) and Gronwall's lemma, $x'(t) = v_{\mu_t}(x(t))$ can only generate trajectories that do not explode in finite time, i.e. there is no $T < \infty$ such that $\limsup_{t \to T^-} \|x(t)\| = +\infty$. Since $v(\mu)(\cdot)$ is continuous by Assumption (**C**$_1$), and $v_{\mu_t}(x) \in T_{\mathcal{M}}(x)$, we can apply an invariance result (Aubin & Frankowska, 1990, Theorem 10.4.1) stating that all the generated trajectories $x(\cdot)$ stay within $\mathcal{M}$ at all times. This can be informally understood as using directions that are always tangent or pointing within $\mathcal{M}$ cannot push $x(\cdot)$ outside of $\mathcal{M}$.

The superposition principle (Ambrosio & Crippa, 2014, Theorem 3.4) states that $\text{supp}(\mu_t)$ is contained in the set of positions $x(t)$ reached by all the trajectories satisfying $x'(t) = v_{\mu_t}(x(t))$ for some $x_0 = x(0) \in \text{supp}(\mu_0)$. Consequently $\text{supp}(\mu_t) \subset \mathcal{M}$, namely $\mathcal{M}$ is flow-invariant for (4). □

B.10. Proof of Lemma 11

Assume that $d\pi(x) \propto e^{-V(x)} d\lambda(x)$, then $s(x) = -\nabla V(x)$. Translating and reindexing, w.l.o.g. we can take $\mathcal{M} = \text{span}(e_I, \ldots, e_d)$. Thus by symmetry of π w.r.t. $\mathcal{M}$, we have that $s_i(x) = (\nabla \log \pi(x))_i = 0$ for all $i < I$ and $x \in \mathcal{M}$, so $s(x) \in \mathcal{M}$. Hence, for every $x \in \mathcal{M}$, $\partial_j s_i(x) = 0$, for $i < I$ and $I \leq j \leq d$. Therefore, for any $u \in \mathcal{M}$ $[J(s)(x)]u = [J(s)(x)]_{|\mathcal{M}}u$, i.e. $\mathcal{M}$ is left invariant under the action of any Jacobian matrix $[J(s)(x)]$ for $x \in \mathcal{M}$.

Consider a radial kernel $k(x,y) = \phi(\|x-y\|^2/2)$ with $\phi \in C^3(\mathbb{R})$. Then $\nabla_2 k(x,y) = (x-y)\phi'(\|x-y\|^2/2)$ and $H_1 k(x,y) = \phi'(\|x-y\|^2/2)I_d + (x-y)\otimes(x-y)\phi''(\|x-y\|^2/2)$. As, for all $x, y \in \mathcal{M}$, $x-y \in \mathcal{M}$, $\nabla_2 k(x,y) \in \mathcal{M}$ and $\mathcal{M}$ is left invariant under the action of $H_1 k(x,y)$. Since $\boldsymbol{\nabla}\boldsymbol{\cdot}_1 \nabla_2 k(x,y)$ is radial as well, $\nabla_2(\boldsymbol{\nabla}\boldsymbol{\cdot}_1 \nabla_2 k(x,y)) \in \operatorname{span}(x-y)$. We have thus shown that all the terms in (36) belong to $\mathcal{M}$ if $x, y \in \mathcal{M}$, so $\nabla_2 k_\pi(x,y) \in \mathcal{M}$.

Consequently $\nabla_{W_2}\mathcal{F}(\mu)(y) = \mathbb{E}_{x\sim\mu}[\nabla_2 k_\pi(x,y)] \in \mathcal{M}$ for any μ such that $\operatorname{supp}(\mu) \subset \mathcal{M}$. Since the contingent cone of a subspace at any point is equal to the subspace itself, we conclude by applying Proposition 20 to $\mathcal{M}$.

B.11. Proof of Lemma 12

We start by noting, that $\tilde{\mathcal{L}}_\pi(\phi) = \nabla V \cdot \nabla\phi$ is well-defined on $C_c^\infty(\mathbb{R}^d)$. Denote $Z = \int_{\mathbb{R}^d} e^{-V(x)}\mathrm{d}x$. Using integration by parts, we have for $\phi, \psi \in C_c^\infty(\mathbb{R}^d)$:

$$\begin{aligned}
\langle \tilde{\mathcal{L}}_\pi \phi, \psi\rangle_{L_2(\pi)} &= \frac{1}{Z}\int_{\mathbb{R}^d} (\nabla V \cdot \nabla\phi(x) - \Delta\phi(x))\psi(x)e^{-V(x)}\mathrm{d}x \\
&= \frac{1}{Z}\int_{\mathbb{R}^d} (\nabla V \cdot \nabla\phi(x))\psi(x)e^{-V(x)} + \nabla\phi(x)\cdot(\nabla\psi(x)e^{-V(x)})\mathrm{d}x \\
&= \frac{1}{Z}\int \langle \nabla\phi(x), \nabla\psi(x)\rangle e^{-V(x)}\mathrm{d}x = \langle \nabla\phi, \nabla\psi\rangle_{L_2(\pi)}.
\end{aligned}$$

It follows that $\tilde{\mathcal{L}}_\pi$ is symmetric, and since $\langle \tilde{\mathcal{L}}_\pi f, f\rangle = \|\nabla f\|^2_{L_2(\pi)}$, it is positive as well. We can now define $\mathcal{L}_\pi$ as the Friedriechs extension of $\tilde{\mathcal{L}}_\pi$ over $L_2(\pi)$ (Pankrashkin, 2014, Definition 2.17). This means that, when we consider the bilinear form $F(\phi,\psi) = \langle \tilde{\mathcal{L}}_\pi\phi, \psi\rangle$, there exists the smallest closed bilinear form $\overline{F}$ on $L_2(\pi)$ which extends it (Pankrashkin, 2014, Proposition 2.16). The operator associated with $\overline{F}$ is a self-adjoint extension of $\tilde{\mathcal{L}}_\pi$ on $L_2(\pi)$, which is also positive. We denote this extension by $\mathcal{L}_\pi$. Furthermore, the domain of $\mathcal{L}_\pi$ is by definition contained in the domain of $\overline{F}$, which is the closure of C_c^∞ for the norm (29) (Pankrashkin, 2014, Proposition 2.8). The claim of density of $C_c^\infty(\mathbb{R}^d)$ in $\mathcal{D}(\mathcal{L}_\pi)$ for the norm (29) now follows.

Recall, that we have shown that for $\phi \in C_c^\infty(\mathbb{R}^d)$ we have $\langle \phi, \mathcal{L}_\pi\phi\rangle = \|\nabla\phi\|^2_{L^2(\pi)}$. Therefore

$$\|\phi\|^2_{\mathcal{L}_\pi} = \left(\|\nabla\phi\|^2_{L^2(\pi)} + \|\phi\|^2_{L^2(\pi)}\right)$$

which is the $W^{1,2}(\pi)$ Sobolev norm. This means, that the domain of the closure $\overline{F}$ is equal to $W_0^{1,2}(\pi)$, and hence $\mathcal{D}(\mathcal{L}_\pi) \subset W_0^{1,2}(\pi)$. In fact, one can establish that $\mathcal{D}(\mathcal{L}_\pi) = W_0^{1,2}(\pi) = W^{1,2}(\pi)$, though we will not need this. It now follows that for any $f \in \mathcal{D}(\mathcal{L}_\pi)$ there exists a weak derivative $\nabla f \in L_2(\pi)$, and it is easy to establish the equality $\|\nabla f\|^2_{L_2(\pi)} = \langle f, \mathcal{L}_\pi f\rangle_{L_2(\pi)}$ by approximating f in the norm $\|\cdot\|_{W^{1,2}(\pi)}$.

It is now easy to show, that the kernel of $\mathcal{L}_\pi$ consists of constant functions. If $\mathcal{L}_\pi f = 0$, we have $\|\nabla f\|^2_{L^2(\pi)} = \langle f, \mathcal{L}_\pi f\rangle_{L^2(\pi)} = 0$. One can deduce, that if the weak derivative is zero then the function is constant by a standard argument using mollifiers that we sketch below:

Define the convolution in $L^2(\pi)$ by $f \star g(x) = \int f(x-y)g(y)d\pi(y)$. By standard properties of the convolution, for any $\phi \in C_c^\infty$, we have $f \star \phi \in C^\infty \cap L^2(\pi)$ and $\nabla(f\star\phi) = f \star (\nabla\phi)$, and if f has a weak derivative then by definition $f\star(\nabla\phi) = (\nabla f)\star\phi$. Let ϕ_n be now a mollifier then again, by standard arguments, $f \star \phi_n$ converges to f in $L_2(\pi)$, and also $f\star\phi_n$ is smooth with $\nabla(f\star\phi_n) = (\nabla f)\star\phi_n = 0$, hence $f\star\phi_n$ is constant. It follows that f is constant.

We are left to show that the set $\{(\phi, \mathcal{L}_\pi\phi) \in L^2(\pi)\times L^2(\pi) : \phi \in C_c^\infty(\mathbb{R}^d)\}$ is dense in the graph of $L^2(\pi)$ for the topology inherited from the normed space $L^2(\pi)\times L^2(\pi)$. The domain of the closed form $\overline{F}$ with the norm (29) is a Hilbert space (Pankrashkin, 2014, Definition 2.5), with an inner product denoted by $\langle\cdot,\cdot\rangle_{\mathcal{L}_\pi}$. We have that for any $f \in \mathcal{D}(\mathcal{L}_\pi)$ there exists a sequence $(\phi_n)_{n=1}^\infty$ such that $\lim_{n\to\infty}\|\phi_n - f\|_{\mathcal{L}_\pi} = 0$. By the Cauchy-Schwarz inequality, we have that for any $\psi \in C_c^\infty$:

$$|\langle \psi, \mathcal{L}_\pi(f-\phi_n)\rangle_{L^2(\pi)} + \langle \psi, f-\phi_n\rangle_{L^2(\pi)}| = |\langle \psi, f-\phi_n\rangle_{\mathcal{L}_\pi}| \leq \|\psi\|_{\mathcal{L}_\pi}\|f-\phi_n\|_{\mathcal{L}_\pi},$$

and since ϕ_n converges to f in $L^2(\pi)$ strongly, we obtain $\lim_{n\to\infty}\langle\psi, \mathcal{L}_\pi(f-\phi_n)\rangle = 0$ for all $\psi \in C_c^\infty(\mathbb{R}^d)$. Since $C_c^\infty(\mathbb{R}^d)$ is dense in $L^2(\pi)$, it follows that $\mathcal{L}_\pi\phi_n$ converges weakly to $\mathcal{L}_\pi f$. By a version of Mazur's lemma on weakly and strongly closed convex sets (Renardy & Rogers, 2006, Lemma 10.19), it follows that there exists a sequence of finite convex combinations of $\mathcal{L}_\pi \phi_n$ converging strongly to $\mathcal{L}_\pi f$. More precisely, there exists a function $N : \mathbb{N} \to \mathbb{N}$ and a sequence of sets of real numbers $\{\alpha_{k,m}\}_{m=k}^{N(k)}$ such that $\alpha_{k,m} \geq 0$, $\sum_{m=k}^{N(k)} \alpha_{k,m} = 1$, and

$$\lim_{k\to\infty} \sum_{n=k}^{N(k)} \alpha_{k,n} \mathcal{L}_\pi \phi_n = \mathcal{L}_\pi f$$

in the strong topology on $L^2(\pi)$. It follows by linearity of $\mathcal{L}_\pi$, that the functions $\psi_k = \sum_{n=k}^{N(k)} \alpha_{k,n}\phi_n$ are C_c^∞ functions that converge to f in $L^2(\pi)$ strongly, and $\mathcal{L}_\pi \psi_k$ converges to $\mathcal{L}_\pi f$ strongly in $L^2(\pi)$ as well.

B.12. Proof of Proposition 14

To perform the computations related to Lemma 22 for the induction formula (10), we need a compactly-supported push-forward, however $\nabla_{W_2}\mathcal{F}(\mu_n)(\cdot)$ is not compactly supported in general. We consequently leverage the compactness of the measure it is applied on to perform our analysis. We thus first show by induction that, owing to (10), $\mu_n \in \mathcal{P}_c(\mathbb{R}^d)$ for all $n \in \mathbb{N}$, since $\mu_0 \in \mathcal{P}_c(\mathbb{R}^d)$. Assume that $\mu_n \in \mathcal{P}_c(\mathbb{R}^d)$, then, by definition of the push-forward operation, since $\nabla_{W_2}\mathcal{F}(\mu_n) \in C^1(\mathbb{R}^d; \mathbb{R}^d)$,

$$\operatorname{supp}(\mu_{n+1}) \subset (I - \gamma\nabla_{W_2}\mathcal{F}(\mu_n))\,(\operatorname{supp}(\mu_n)) \subset \operatorname{supp}(\mu_n) + \gamma\left(\max_{x\in\operatorname{supp}(\mu_n)} \|\nabla_{W_2}\mathcal{F}(\mu_n)(x)\|_2\right) B(0,1) =: S_n. \quad (53)$$

Hence $\mu_{n+1} \in \mathcal{P}_c(\mathbb{R}^d)$ as claimed.

Consider a path between μ_n and μ_{n+1} of the form $\rho_t = (I - \gamma t \nabla_{W_2}\mathcal{F}(\mu_n))_\#\, \mu_n$. Set $\phi(x) = -\gamma\nabla_{W_2}\mathcal{F}(\mu_n)(x)$, and $s_t(x) = x + t\phi(x)$ which is distributed according to ρ_t for x distributed as μ_n. In general the function $(I - \gamma\nabla_{W_2}\mathcal{F}(\mu_n))$ is not compactly supported so we cannot apply Lemma 22 outright. But since the push-forward is only applied to the compactly supported μ_n in the definition of ρ_t, we can find, through a mollifier, a function $f_{n,t} \in C^1(\mathbb{R}^d; \mathbb{R}^d)$ such that it coincides with $(I - \gamma t\nabla_{W_2}\mathcal{F}(\mu_n))$ on S_n and is supported on $S_n + B(0,1)$. So $\rho_t = (f_{n,t})_\#\mu_n$ and we can apply Lemma 22. Hence $t \mapsto \mathcal{F}(\rho_t)$ is differentiable and absolutely continuous, and consequently

$$\mathcal{F}(\rho_1) - \mathcal{F}(\rho_0) = \dot{\mathcal{F}}(\rho_0) + \int_0^1 \left[\dot{\mathcal{F}}(\rho_t) - \dot{\mathcal{F}}(\rho_0)\right] dt, \quad (54)$$

where

$$\dot{\mathcal{F}}(\rho_t) = \mathbb{E}_{\substack{x\sim\mu_n\\ x'\sim\mu_n}}\left[\nabla_1 k_\pi\left(s_t(x), s_t(x')\right)^\top (\phi(x))\right].$$

The first term in the r.h.s. of (54) is

$$\dot{\mathcal{F}}(\rho_0) = -\gamma\mathbb{E}_{x\sim\mu_n}\left[\|\nabla_{W_2}\mathcal{F}(\mu_n)(x)\|^2\right].$$

Since $s_t(x) - s_{t'}(x) = (t - t')\,\phi(x)$, by Assumption $(\mathbf{A}_1)$, we derive through Jensen's and Cauchy-Schwarz inequalities that

$$\begin{aligned}
&\left|\dot{\mathcal{F}}(\rho_t) - \dot{\mathcal{F}}(\rho_{t'})\right| = \left|\mathbb{E}_{\substack{x\sim\mu_n\\ x'\sim\mu_n}}\left[\left(\nabla_1 k_\pi\left(s_t(x), s_t(x')\right) - \nabla_1 k_\pi\left(s_{t'}(x), s_{t'}(x')\right)\right)^\top \phi(x)\right]\right| \\
&\leq \mathbb{E}_{\substack{x\sim\mu_n\\ x'\sim\mu_n}}\left[\left(\|\nabla_1 k_\pi\left(s_t(x), s_t(x')\right) - \nabla_1 k_\pi\left(s_{t'}(x), s_t(x')\right)\| + \|\nabla_1 k_\pi\left(s_{t'}(x), s_t(x')\right) - \nabla_1 k_\pi\left(s_{t'}(x), s_{t'}(x')\right)\|\right)\|\phi(x)\|\right] \\
&\leq |t-t'|\mathbb{E}_{\substack{x\sim\mu_n\\ x'\sim\mu_n}}\left[\left(L(s_t(x'))\|\phi(x)\| + L(s_{t'}(x))\,\|\phi(x')\|\right)\|\phi(x)\|\right] \\
&\leq |t-t'|\left(\mathbb{E}_{x'\sim\mu_n}\left[L(s_t(x'))\right]\mathbb{E}_{x\sim\mu_n}\left[\|\phi(x)\|^2\right] + \mathbb{E}_{x\sim\mu_n}\left[L(s_{t'}(x))\|\phi(x)\|\right]\mathbb{E}_{x'\sim\mu_n}\left[\|\phi(x')\|\right]\right) \\
&\leq |t-t'|\left(\mathbb{E}_{x\sim\mu_n}\left[L(s_t(x))\right]\mathbb{E}_{x\sim\mu_n}\left[\|\phi(x)\|^2\right] + \mathbb{E}_{x\sim\mu_n}\left[L(s_{t'}(x))^2\right]^{\frac{1}{2}}\mathbb{E}_{x\sim\mu_n}\left[\|\phi(x)\|^2\right]^{\frac{1}{2}}\mathbb{E}_{x'\sim\mu_n}\left[\|\phi(x')\|^2\right]^{\frac{1}{2}}\right) \\
&\leq \gamma^2|t-t'|\left(\|L\|_{L^1(\rho_t)} + \|L\|_{L^2(\rho_{t'})}\right)\mathbb{E}_{x\sim\mu_n}\left[\|\nabla_{W_2}\mathcal{F}(\mu_n)(x)\|^2\right] \text{ since } \phi(x) = -\gamma\nabla_{W_2}\mathcal{F}(\mu_n)(x).
\end{aligned}$$

Hence, for $t' = 0$, we have

$$\left|\dot{\mathcal{F}}(\rho_t) - \dot{\mathcal{F}}(\rho_0)\right| \leq t\gamma^2\left(\|L\|_{L^1(\rho_t)} + \|L\|_{L^2(\mu_n)}\right)\mathbb{E}_{x\sim\mu_n}\left[\|\nabla_{W_2}\mathcal{F}(\mu_n)(x)\|^2\right].$$

Then, since $\rho_t = ((1-t)I + t(I+\phi))_{\#}\mu_n$, by convexity of L, we can write :

$$\|L\|_{L^1(\rho_t)} = \int |L((1-t)x + t(x+\phi(x))|\mathrm{d}\mu_n(x)$$
$$\leq (1-t)\|L\|_{L^1(\mu_n)} + t\|L\|_{L^1(\mu_{n+1})} \leq (1-t)\|L\|_{L^2(\mu_n)} + t\|L\|_{L^2(\mu_{n+1})} \leq M.$$

We can thus upper bound the second term in the r.h.s. of (54),

$$\int_0^1 |\dot{\mathcal{F}}(\rho_t) - \dot{\mathcal{F}}(\rho_0)|dt \leq \int_0^1 \left(t2M\gamma^2\mathbb{E}_{x\sim\mu_n}\left[\|\nabla_{W_2}\mathcal{F}(\mu_n)(x)\|^2\right]\right)\mathrm{d}t = \gamma^2 M\mathbb{E}_{x\sim\mu_n}\left[\|\nabla_{W_2}\mathcal{F}(\mu_n)(x)\|^2\right].$$

Finally, since $\gamma M \leq 1$, (54) leads to

$$\mathcal{F}(\mu_{n+1}) - \mathcal{F}(\mu_n) \leq -\gamma(1-\gamma M)\mathbb{E}_{x\sim\mu_n}\left[\|\nabla_{W_2}\mathcal{F}(\mu_n)(x)\|^2\right] \leq 0.$$

C. Additional results

C.1. On the kernel integral operator

Lemma 21. For any $f, g \in L^2(\pi)$, we have that

$$\langle S_{\pi,k_\pi} f, S_{\pi,k_\pi} g\rangle_{\mathcal{H}_{k_\pi}} = \iint f(x)^\top g(y) k_\pi(x,y)\mathrm{d}\pi(x)\mathrm{d}\pi(y).$$

Proof. By using the reproducing property, we deduce that

$$\langle S_{\pi,k_\pi} f, S_{\pi,k_\pi} g\rangle_{\mathcal{H}^d_{k_\pi}} = \langle \int k_\pi(x,\cdot)f(x)d\pi(x), \int k_\pi(y,\cdot)g(y)d\pi(y)\rangle_{\mathcal{H}^d_{k_\pi}}$$
$$= \sum_{i=1}^d \langle \int k_\pi(x,\cdot)f_i(x)d\pi(x), \int k_\pi(y,\cdot)g_i(y)d\pi(y)\rangle_{\mathcal{H}_{k_\pi}}$$
$$= \sum_{i=1}^d \iint f_i(x)k_\pi(x,y)g_i(y)\mathrm{d}\pi(x)\mathrm{d}\pi(y) = \iint f(x)^\top g(y)k_\pi(x,y)\mathrm{d}\pi(x)\mathrm{d}\pi(y).$$

□

C.2. On the differentiation of the squared KSD

For Lemma 22 below, our computations are similar to the ones in Arbel et al. (2019, Lemma 22 and 23) with some terms getting simpler owing to the Stein's property of the Stein kernel, but under a weaker assumption than a uniform Lipschitz constant for the kernel (see the discussion in Appendix A.6).

Lemma 22. Let $q \in \mathcal{P}_2(\mathbb{R}^d)$ and $\phi \in C_c^1(\mathbb{R}^d)$. Consider the path ρ_t from q to $(I+\nabla\phi)_{\#}q$ given by: $\rho_t = (I + t\nabla\phi)_{\#}q$, for all $t \in [0,1]$. Suppose Assumptions ($\mathbf{A}_1$) and ($\mathbf{A}_3$) hold. Then, $\mathcal{F}(\rho_t)$ is twice differentiable in t with

$$\dot{\mathcal{F}}(\rho_t) = \mathbb{E}_{\substack{(x)\sim q\\(x')\sim q}}\left[\nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))^\top\nabla\phi(x)\right],$$
$$\ddot{\mathcal{F}}(\rho_t) = \mathbb{E}_{\substack{(x)\sim q\\(x')\sim q}}\left[\nabla\phi(x')^\top\nabla_1\nabla_2 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))\nabla\phi(x)\right]$$
$$+ \mathbb{E}_{\substack{(x)\sim q^*\\(x')\sim q^*}}\left[\nabla\phi(x)^\top H_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))\nabla\phi(x)\right].$$

Proof. The function $f : t \mapsto k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))$ is differentiable for all $x, x' \in \mathbb{R}^d$, and its time derivative is :

$$\dot{f} = \nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))^\top\nabla\phi(x) + \nabla_2 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))^\top\nabla\phi(x')$$
$$= \nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))^\top\nabla\phi(x) + \nabla_1 k_\pi(x'+t\nabla\phi(x'), x+t\nabla\phi(x))^\top\nabla\phi(x') \quad (55)$$

using the symmetry of k_π. The two terms on the r.h.s. of the former equation are symmetric w.r.t. x and x', so we will focus on the first one. By the Cauchy-Schwartz inequality and Assumption ($\mathbf{A}_1$),

$$\begin{aligned}|\nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))^\top \nabla\phi(x)| &\leq \|\nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))\|\|\nabla\phi(x)\| \\ &\leq (L(x+t\nabla\phi(x))\|x'+t\nabla\phi(x')\| + \|\nabla_1 k_\pi(x+\nabla\phi(x),0)\|)\,\|\nabla\phi(x)\|\end{aligned}$$

The r.h.s. of the above inequality is integrable in x because $\nabla\phi$ is compactly supported, L is continuous, and because $x' \mapsto \|x'+t\nabla\phi(x')\|$ is integrable since $q \in \mathcal{P}_2(\mathbb{R}^d)$.Therefore, by the differentiation lemma (Klenke, 2013, Theorem 6.28), $\mathcal{F}(\rho_t)$ is differentiable and $\dot{\mathcal{F}}(\rho_t) = \mathbb{E}_{(x)\sim q,(x')\sim q}[\dot{f}]$, i.e.

$$\dot{\mathcal{F}}(\rho_t) = \mathbb{E}_{\substack{(x)\sim q \\ (x')\sim q}}\left[\nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))^\top \nabla\phi(x)\right].$$

Now define the function $g : t \mapsto \nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))^\top \nabla\phi(x)$. Its time derivative writes as

$$\dot{g} = \nabla\phi(x)^\top \nabla_2\nabla_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))\nabla\phi(x') + \nabla\phi(x)^\top H_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))\nabla\phi(x).$$

The first term on the r.h.s. is integrable in (x, x') because it is compactly supported and continuous. We now deal with the second term. Under Assumption ($\mathbf{A}_3$), $H_1 k_\pi(x, y)$ is dominated by a q-integrable function. Then,

$$|\nabla\phi(x)^\top H_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))\nabla\phi(x)| \leq \|H_1 k_\pi(x+t\nabla\phi(x), x'+t\nabla\phi(x'))\|_{op}\|\nabla\phi(x)\|^2$$

The r.h.s. of the above inequality is integrable because $(x, y) \mapsto H_1 k_\pi(x, y)$ is integrable and $\nabla\phi$ is bounded. □

D. Additional experiments

Implementation details. The code for KSD Descent is written in Python using Pytorch (Paszke et al., 2019). We use Matplotlib (Hunter, 2007) for figures, Scipy (Virtanen et al., 2020) for the L-BFGS implementation, as well as Numpy (Harris et al., 2020). It is available at `https://github.com/pierreablin/ksddescent`.

Different initializations and variance for Gaussian mixtures. To discuss in greater detail the results of our second toy example presented in Section 5.1 and illustrated on Figure 3, we performed some more experiments for several choices of initialization and variance, for a mixture of two Gaussians with equal variance (see Figure 7). We also investigated the support of the stationary points of KSD Descent for a mixture of three Gaussians with different variances (see Figure 8). As discussed in Appendix B.9, the support is not necessarily a smooth submanifold or an axis of symmetry.

A comparison between KSD Descent and Stein points. We finally compare KSD Descent and Stein points (Chen et al., 2018). We choose a low dimensional problem, since the approach of Stein points cannot scale to large dimensions. We use the classical 2-D "banana" density. Figure 9 shows the behavior of the two algorithms. Interestingly, KSD Descent succeeds and does not fall into spurious local minima, even tough the density is not log-concave. We posit that this happens because here the potential $\log(\pi)$ does not have saddle points.

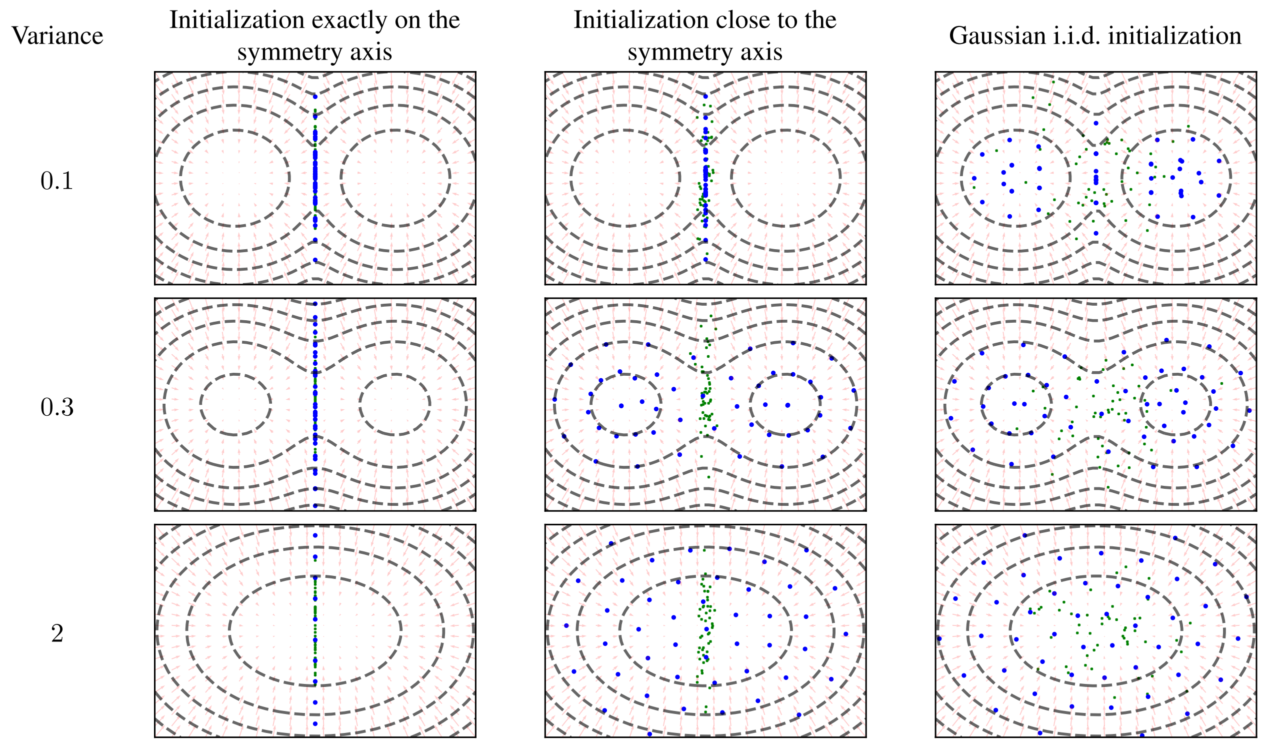

Figure 7. Results of KSD Descent for a mixture of two gaussians, depending on their variance and on the initialization of the algorithm. The green crosses indicate the initial particle positions, while the blue ones are the final positions.

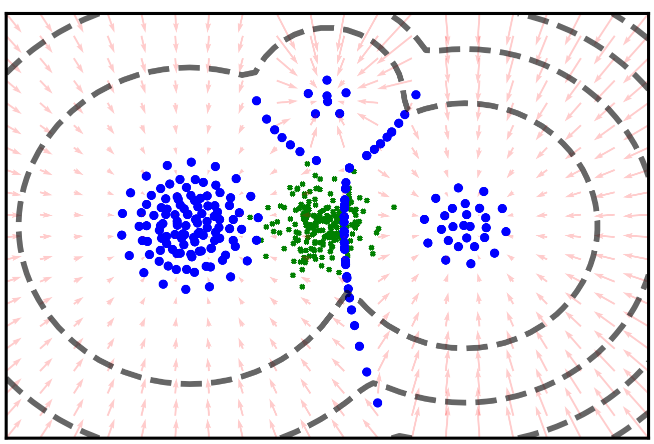

Figure 8. Using KSD Descent to sample from an unbalanced mixture of Gaussians. Some particles get stuck in spurious zones, which are not a straight line nor a manifold. The green crosses indicate the initial particle positions, while the blue ones are the final positions.

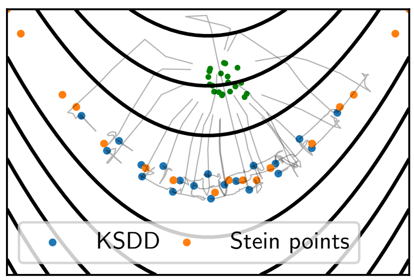

Figure 9. Comparison of KSD Descent and Stein points on the “banana” dataset. Green points are the initial points for KSD Descent. Both methods work successfully here, even though it is not a log-concave distribution. We posit that KSD Descent succeeds because there is no saddle point in the potential.