Journal of Machine Learning Research 23 (2022) 1-68 Submitted 11/21; Revised 4/22; Published 8/22

On quantitative Laplace-type convergence results for some exponential probability measures, with two applications

Valentin De Bortoli VALENTIN.DEBORTOLI@GMAIL.COM
Department of Statistics
University of Oxford, UK

Agnès Desolneux AGNES.DESOLNEUX@ENS-PARIS-SACLAY.FR
Centre Borelli,
CNRS and ENS Paris-Saclay, France

Editor: Sebastien Bubeck

Abstract

Laplace-type results characterize the limit of sequence of measures $(\pi_\varepsilon)_{\varepsilon>0}$ with density w.r.t. the Lebesgue measure $(\mathrm{d}\pi_\varepsilon/\mathrm{d}\lambda)(x) \propto \exp[-U(x)/\varepsilon]$ when the temperature $\varepsilon > 0$ converges to 0. If a limiting distribution π_0 exists, it concentrates on the minimizers of the potential U. Classical results require the invertibility of the Hessian of U in order to establish such asymptotics. In this work, we study the particular case of norm-like potentials U and establish quantitative bounds between π_ε and π_0 w.r.t. the Wasserstein distance of order 1 under an invertibility condition of a generalized Jacobian. One key element of our proof is the use of geometric measure theory tools such as the coarea formula. We apply our results to the study of maximum entropy models (microcanonical/macrocanonical distributions) and to the convergence of the iterates of the Stochastic Gradient Langevin Dynamics (SGLD) algorithm at low temperatures for non-convex minimization.

Keywords: Laplace methods, Geometric measure theory, Stochastic Approximation, Entropy models

1. Introduction

Asymptotic expansions of integrals are ubiquitous in probability theory with applications in simulated annealing (Gelfand and Mitter, 1993; Pelletier, 1998), Bayesian inference (Haughton, 1988; Tierney and Kadane, 1986; Rue et al., 2009), statistical physics (Ellis and Rosen, 1982) or chaos expansion (Korshunov et al., 2015). In this paper, we are interested in families of Gibbs probability measures π_ε given by the density w.r.t. the Lebesgue measure $(\mathrm{d}\pi_\varepsilon/\mathrm{d}\lambda)(x) \propto \exp[-\|F(x)\|^k/\varepsilon]$, for any $x \in \mathbb{R}^d$, where $k \in \mathbb{N}$ and $F : \mathbb{R}^d \to \mathbb{R}^p$ is some function with $F(0) = 0$. In particular, we derive non asymptotic bounds between $(\pi_\varepsilon)_{\varepsilon>0}$ and π_0 when $\varepsilon \to 0$.

Taking a step back, one way to deal with integrals of the form $\int_{\mathbb{R}^d} \exp[-U(x)/\varepsilon]\mathrm{d}x$ where $U : \mathbb{R}^d \to \mathbb{R}_+$ is to rely on Laplace-type techniques, see Bleistein and Handelsman (1975); Olver (1997); De Bruijn (1981); Evgrafov (2020); Erdélyi (1956); Fedoryuk (1989); Bender and Orszag (2013); Wong (2001) for instance. Assuming that U admits a unique minimizer $U(0) = 0$ and under additional regularity conditions, it can be shown that if the Hessian of U evaluated at 0 is invertible then $\int_{\mathbb{R}^d} \exp[-U(x)/\varepsilon]\mathrm{d}x \sim C_U \varepsilon^{d/2}$, with explicit constant $C_U >$

0 (series expansions are also available under similar conditions). This result on asymptotic integrals can be immediately used to prove the convergence of probability measures with density w.r.t. the Lebesgue measure $(\mathrm{d}\pi_\varepsilon/\mathrm{d}\lambda)(x) \propto \exp[-U(x)/\varepsilon]$, see Hwang (1980) for instance. However the invertibility of the Hessian is a restricting condition which is not satisfied in the setting of this paper where $U(x) = \|F(x)\|^k$ in the case $p \leq d$.

When the Hessian $\nabla^2 U(0)$ is not invertible, the asymptotic integral expansion becomes *degenerate* and is more difficult to analyze. However, several approaches have been proposed in order to tackle this issue. For instance in Rytova and Yarovaya (2016) a multidimensional version of the Watson lemma is established under restrictive conditions on the potential U near its singularities. Another approach consists in integrating over the manifold of minimizers and apply local change of variables Korshunov et al. (2015), see also Barbe (2003); Breitung (2006). In particular, Korshunov et al. (2015) derives an asymptotic integral expansion similar to the one obtained in the non-degenerate case, under some invertibility condition of the minors of the Hessian. Similarly, Hwang (1980) uses the tubular neighborhood theorem to obtain that π_ε converges to a limiting measure under the assumption that the minimizers of U can be partitioned into a collection of manifolds. Closer to the method introduced in the present paper, Combet (2006); Arnold et al. (2012) propose to use the so-called *Gelfand-Leray forms* to tackle the non-degeneracy problem. By integrating over the level sets it can be shown that

$$\int_{\mathsf{V}} \varphi(x) \exp[-U(x)/\varepsilon] \mathrm{d}x = \int_0^{+\infty} \exp[-t/\varepsilon] (\int_{U^{-1}(t) \cap \mathsf{V}} \varphi(x) \omega_U(x)) \mathrm{d}t \ , \tag{1}$$

where V is an open set, $\varphi \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R})$ a test function and ω_U the so-called Gelfand-Leray form associated with U. Then by establishing regularity properties for the functional $t \mapsto \int_{U^{-1}(t)} \varphi(x)\omega_U(x)$ one can recover asymptotic integral expansion in the case where U is analytic, see Combet (2006); Arnold et al. (2012). This theory does not rely on any invertibility condition on the Hessian of U but the exponents appearing in the asymptotic expansion are usually not available in closed form in general since their derivation relies on resolution of singularities (Hironaka, 1964).

In this paper, we consider a different approach which also does not rely on invertibility conditions on the Hessian of U at singularity points but allows us to derive quantitative expansions. To do so, we restrict the set of functions $U : \mathbb{R}^d \to \mathbb{R}_+$ to the set of norm-like functions, *i.e.* U is norm-like if there exists $F : \mathbb{R}^d \to \mathbb{R}^p$ and $k \in \mathbb{N}$ such that for any $x \in \mathbb{R}^d$, $U(x) = \|F(x)\|^k$ (however our results can be extended to general potentials U under classical invertibility conditions on the Hessian of U). Instead of invertibility conditions of the Hessian of U we require invertibility conditions on the generalized Jacobian of F to be defined below. The setting of norm-like potentials is of particular interest in machine learning as it arises in many applications such as Variational AutoEncoders (VAEs) or macrocanonical/microcanonical distributions, as we will discuss. Our approach relies on tools from the geometric measure theory and in particular the coarea formula, see Ambrosio et al. (2000) for instance. Indeed, using this formula we are able to establish a similar result as (1) where the Gelfand-Leray form is replaced by a twisted Hausdorff measure ν. By establishing the Lipschitz regularity of the mapping $t \mapsto \int_{U^{-1}(t)\cap\mathsf{V}} \varphi(x)\mathrm{d}\nu(x)$ we are able to provide quantitative bounds for the integral expansion. This expansion is the key to our main result which establishes quantitative bounds w.r.t. the Wasserstein distance of order 1 between π_ε and a limit probability measure π_0 for $\varepsilon > 0$ small enough. Concurrently to this work, Bras and Pagès (2021) establish similar quantitative results in the case where the Hessian of U is invertible.

One of our main motivation for this study is the application of Laplace-type results to maximum entropy distributions. In particular, using our quantitative bounds, we are able to provide a link between two possible maximum entropy distributions commonly used in statistical physics and image processing (Bruna and Mallat, 2018; De Bortoli et al., 2021). Given a reference measure μ and constraints $F : \mathbb{R}^d \to \mathbb{R}^p$ it is possible to define the *microcanonical distribution*, which corresponds to the probability measure π_{mic} with the minimum Kullback-Leibler divergence w.r.t. μ and such that π_{mic} is supported on the set $\{x \in \mathbb{R}^d : F(x) = 0\}$. Similarly, one can define the *macrocanonical distribution* with level $\varepsilon > 0$ denoted $\pi_{\mathrm{mac}}^{\varepsilon}$ such that $\pi_{\mathrm{mac}}^{\varepsilon}$ minimizes the Kullback-Leibler divergence w.r.t. μ and satisfies the integrability condition $\int_{\mathbb{R}^d} \|F(x)\|^2 \mathrm{d}\pi_{\mathrm{mac}}^{\varepsilon}(x) = \varepsilon$. Using results from information geometry (Csiszár, 1975) and under mild regularity conditions, $\pi_{\mathrm{mac}}^{\varepsilon}$ can be written as a Gibbs measure and its limiting behavior when $\varepsilon \to 0$ can be investigated in the context of asymptotic expansions of integrals. In this work, we show that the limit of the macrocanonical distributions when $\varepsilon \to 0$ is a twisted microcanonical distribution and propose an algorithm to asymptotically recover the original microcanonical distribution. Another motivation for the study of the limiting behavior of distributions with norm-like potentials comes from the study of the posterior of the latent variables in VAE. In particular, we show that under assumptions on the decoding neural network, the posterior distribution of a VAE concentrates with explicit rates.

Finally, we also consider the non-asymptotic study of Stochastic Gradient Langevin Dynamics (SGLD) (Welling and Teh, 2011), a popular algorithm used to approximate the minimizers of non-convex functions in a machine learning setting. Given a potential (or population risk) $U : \mathbb{R}^d \to \mathbb{R}_+$ which can be written for any $x \in \mathbb{R}^d$ as $U(x) = \int_{\mathsf{Z}} u(x, z) \mathrm{d}\mu(z)$, we assume that we have access to an empirical version of this risk for any $n \in \mathbb{N}$ and $z^{1:n} = \{z_i\}_{i=1}^n \in \mathsf{Z}^n$ given by $U_n(x, z^{1:n}) = (1/n) \sum_{i=1}^n u(x, z_i)$ as well as an unbiased estimator of its gradient denoted g. Conditionally to the samples $z^{1:n}$, SGLD corresponds to the recursion associated with an unadjusted Langevin dynamics (Roberts and Tweedie, 1996; Durmus and Moulines, 2017; Dalalyan, 2017) with target $U_n(\cdot, z^{1:n})/\varepsilon$ for a small value of $\varepsilon > 0$ and where the gradient of $U_n(\cdot, z^{1:n})$ is replaced by its unbiased estimator g. Since the invariant measure of the underlying Langevin diffusion is given by π_ε with density w.r.t. the Lebesgue measure $(\mathrm{d}\pi_\varepsilon/\mathrm{d}\lambda)(x) \propto \exp[-U_n(x, z^{1:n})/\varepsilon]$ we obtain that for small values of the parameter $\varepsilon > 0$, the samples $(X_k)_{k \in \mathbb{N}}$ are approximately concentrated around the minimizers of $U_n(\cdot, z^{1:n})$. In existing works on SGLD (Raginsky et al., 2017; Chau and Rasonyi, 2019; Gao et al., 2018; Erdogdu et al., 2018; Nguyen et al., 2019; Xu et al., 2017) all quantitative theoretical guarantees are obtained w.r.t. the sequence of $(\mathbb{E}[U(X_k)])_{k \in \mathbb{N}}$. On the contrary, we characterize the limiting distribution of SGLD and establish non-asymptotic quantitative bounds between SGLD and this limit. Finally, we show that the support of the limiting distribution is included into (but not necessarily equal to) the set of minimizers of U when n is large. This last result relies on the notion of *thermodynamic barrier* which we introduce in the context of integral expansion. To our knowledge this is the first time that the importance of such a barrier to derive parametric Laplace-type results is highlighted (as opposed to the *kinetic barrier* which is a well-known quantity in simulated annealing (Hajek, 1988)).

To summarize, our main contributions are three-fold:

(i) We establish quantitative bounds between π_ε and π_0 under the assumption that the potential U is norm-like and additional regularity conditions. In particular we provide an upper-bound between π_ε and π_0 w.r.t. the Wasserstein distance of order 1. We emphasize that our results also hold under the more classical invertibility condition on the Hessian of U. We show a first application of our results with a study of the concentration of posteriors in VAEs.

(ii) We apply our results to show that under mild conditions the limit of a natural class of macrocanonical distributions is *not* the original microcanonical distribution. However, we prove that a twisted sequence of macrocanonical distributions converges to the microcanonical distribution. This observation allows us to construct a Langevin-based algorithm in order to sample from this distribution. We illustrate our method in low-dimensional settings.

(iii) Finally, we apply our theory to study the behavior of the iterates of SGLD in the context of non-convex optimization. In particular, we characterize the limiting distribution of SGLD at low temperature and show that it concentrates on the minimizers of the population loss for a large number of samples using the notion of *thermodynamic barrier*. Note that in this study, we no longer assume that U is norm-like but instead rely on invertibility conditions on the Hessian of U.

The rest of the paper is organized as follows. In Section 2 we present our main results, *i.e.* quantitative bounds between π_ε and the limiting distribution π_0. In Section 3, we present our two main applications: the links between the macrocanonical and microcanonical maximum entropy distributions in Section 3.1, and a study of the convergence of SGLD for non-convex minimization in Section 3.2. The proofs of our results are gathered in Section 4.

Notation

Let $d \in \mathbb{N}^\star$. We denote $\{e_i\}_{i=1}^d$ the canonical basis of $\mathbb{R}^d$. Let $\langle \cdot, \cdot \rangle$ be the Euclidean scalar product over $\mathbb{R}^d$, and $\|\cdot\|$ be the corresponding norm. We denote $\mathrm{B}(x, r)$ the ball with center $x \in \mathbb{R}^d$ and radius $r > 0$ w.r.t. the norm $\|\cdot\|$. Similarly, we denote $\mathrm{B}_\infty(x, r)$ the ball with center $x \in \mathbb{R}^d$ and radius $r > 0$ w.r.t. the norm $\|\cdot\|_\infty$ given for any $x = (x_1, \dots, x_d) \in \mathbb{R}^d$ by $\|x\|_\infty = \sup_{i \in \{1,\dots,d\}} |x_i|$.

Let $\mathsf{A} \subset \mathbb{R}^d$, we denote $\mathrm{diam}(\mathsf{A}) = \sup_{x,y \in \mathsf{A}} \|x - y\|$ its diameter and A^c its complementary set. Let $\mathcal{B}(\mathbb{R}^d)$ denote the Borel σ-field of $\mathbb{R}^d$. Let U be an open set of $\mathbb{R}^d$, $n \in \mathbb{N}^\star$ and let $\mathrm{C}^n(\mathsf{U}, \mathbb{R}^p)$ be the set of the n-differentiable $\mathbb{R}^p$-valued functions defined over U. If $p = 1$ we simply denote $\mathrm{C}^n(\mathsf{U})$. Let $f \in \mathrm{C}^1(\mathsf{U})$, we denote by ∇f its gradient. Furthermore, if $f \in \mathrm{C}^2(\mathsf{U})$ we denote $\nabla^2 f$ its Hessian and Δf its Laplacian. By convention we denote $\nabla^0 f = f$. We also denote $\mathrm{C}(\mathsf{U}, \mathbb{R}^p)$ the set of continuous functions defined over U. Let $f : \mathsf{A} \to \mathbb{R}^p$ with $p \in \mathbb{N}^\star$ and $\mathsf{A} \subset \mathbb{R}^d$. The function f is said to be $\mathtt{L}$-Lipschitz with $\mathtt{L} \geq 0$ if for any $x, y \in \mathsf{A}$, $\|f(x) - f(y)\| \leq \mathtt{L}\,\|x - y\|$. Let $F : \mathbb{R}^d \to \mathbb{R}^p$ such that F is differentiable at $x \in \mathbb{R}^d$ and denote $\mathrm{D}F(x) = (\partial_j F_i(x))_{i \in \{1,\dots,p\}, j \in \{1,\dots,d\}}$ the Jacobian matrix of F which is a $p \times d$ matrix. We define the generalized Jacobian $\mathrm{J}F : \mathbb{R}^d \to \mathbb{R}_+$ given for any $x \in \mathbb{R}^d$ by

$$\mathrm{J}F(x) = \begin{cases} \det(\mathrm{D}F(x)\mathrm{D}F(x)^\top)^{1/2} \ , & \text{if } d \geq p, \\ \det(\mathrm{D}F(x)^\top \mathrm{D}F(x))^{1/2} \ , & \text{if } d \leq p. \end{cases}$$

Finally, if $F \in \mathrm{C}^\ell(\mathbb{R}^d, \mathbb{R}^p)$ with $\ell \in \mathbb{N}$, we define $\mathrm{D}^j F$ for any $j \in \{1, \dots, \ell\}$, recursively by $\mathrm{D}^{j+1}F = \mathrm{D}(\mathrm{D}^j F)$. Note that for any $j \in \{0, \dots, \ell\}$, $\mathrm{D}^j F$ can be represented as a $p \times d \times \cdots \times d$ tensor (where d appears j times) with symmetric last j coordinates. Hence, we can define $\mathrm{D}^{j+1}F^\top$ the tensor where the first and last dimension have been exchanged, which is a $d \times \dots d \times p$ tensor. If $p = 1$ then, we write $\nabla^j F = \mathrm{D}^j F$ for any $j \in \{1, \dots, \ell\}$.

Let $(\mathsf{X}, \mathcal{X})$ be a measurable space. We denote by $\mathbb{F}(\mathsf{X})$ the set of the $\mathcal{X}/\mathcal{B}(\mathbb{R})$-measurable real functions over X. Let $\mathbb{M}(\mathcal{X})$ be the set of finite signed measures over $\mathcal{X}$ and let $\mu \in \mathbb{M}(\mathcal{X})$. For $f \in \mathbb{F}(\mathsf{X})$ a μ-integrable function we denote

$$\mu[f] = \textstyle\int_{\mathsf{X}} f(x) \mathrm{d}\mu(x) \ ,$$

the integral of f w.r.t. to μ when it is well-defined. We also define $\mathscr{P}(\mathsf{X}, \mathcal{X})$ the set of probability measures over $\mathcal{X}$ and when there is no ambiguity on the sigma-field we simply denote it by $\mathscr{P}(\mathsf{X})$. We denote by λ the Lebesgue measure on $\mathbb{R}^d$. Let $(\mathsf{X}, \mathcal{X})$ and $(\mathsf{Y}, \mathcal{Y})$ be two measurable spaces. A Markov kernel K is a mapping $\mathrm{K} : \mathsf{X} \times \mathcal{Y} \to [0,1]$ such that for any $x \in \mathsf{X}$, $\mathrm{K}(x, \cdot) \in \mathscr{P}(\mathsf{Y}, \mathcal{Y})$ and for any $\mathsf{A} \in \mathcal{Y}$, $\mathrm{K}(\cdot, \mathsf{A})$ is measurable.

Let $\mu, \nu \in \mathscr{P}(\mathsf{X}, \mathcal{X})$. A probability measure $\zeta \in \mathscr{P}(\mathsf{X} \times \mathsf{X}, \mathcal{X} \otimes \mathcal{X})$ is said to be a transference plan between μ and ν if for any $\mathsf{A} \in \mathcal{X}$, $\zeta(\mathsf{A} \times \mathcal{X}) = \mu(\mathsf{A})$ and $\zeta(\mathcal{X} \times \mathsf{A}) = \nu(\mathsf{A})$. We denote by $\mathbf{T}(\mu, \nu)$ the set of all transference plans between μ and ν. We define the Wasserstein metric/distance of order $\ell \geq 1$ $\mathbf{W}_\ell(\mu, \nu)$ between μ and ν by

$$\mathbf{W}_\ell(\mu, \nu)^\ell = \inf_{\zeta \in \mathbf{T}(\mu, \nu)} \textstyle\int_{\mathsf{X}^2} \|x - y\|^\ell \mathrm{d}\zeta(x, y) \ .$$

We denote $\mathcal{H}^r$ the Hausdorff measure of order $r > 0$, given for any $\mathsf{A} \subset \mathbb{R}^d$ by $\mathcal{H}^r(\mathsf{A}) = \lim_{\delta \to 0} \mathcal{H}^{r,\delta}(\mathsf{A})$, where for any $\delta > 0$ we have

$$\mathcal{H}^{r,\delta}(\mathsf{A}) = \inf\{(\alpha_r/2^r) \textstyle\sum_{i \in \mathbb{N}} \mathrm{diam}(\mathsf{U}_i)^r \ : \ \mathsf{A} \subset \cup_{i \in \mathbb{N}} \mathsf{U}_i \ , \ \mathrm{diam}(\mathsf{U}_i) \leq \delta\} \ ,$$

where $\alpha_r = \pi^{r/2}/\Gamma(\frac{r}{2} + 1)$ with Γ the usual Gamma function. Basic facts on the Hausdorff measure are gathered in Appendix C. Finally, we denote by $\mathtt{Poly}(k, \mathsf{A})$ the set of polynomials of $k \in \mathbb{N}$ variables with coefficients in $\mathsf{A} \subset \mathbb{R}$.

2. Limit theorems

In this section, we state our main theorem and draw links with previous approximation results for probability integrals. Let $k \in \mathbb{N}$, $\bar{\varepsilon} > 0$, $F : \mathbb{R}^d \to \mathbb{R}^p$ with $d, p \in \mathbb{N}^\star$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$, $\int_{\mathbb{R}^d} \exp[-\|F(x)\|^k / \varepsilon] \mathrm{d}x < +\infty$. For any $\varepsilon \in (0, \bar{\varepsilon})$, we define $\pi_\varepsilon \in \mathscr{P}(\mathbb{R}^d)$ such that for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$

$$\pi_\varepsilon(\mathsf{A}) = \textstyle\int_{\mathsf{A}} \exp[-\|F(x)\|^k / \varepsilon] \mathrm{d}x \Big/ \int_{\mathbb{R}^d} \exp[-\|F(x)\|^k / \varepsilon] \mathrm{d}x \ . \tag{2}$$

Our goal is to study the behavior of π_ε when $\varepsilon \to 0$. In particular, we identify a limit $\pi_0 \in \mathscr{P}(\mathbb{R}^d)$ and derive non-asymptotic convergence bounds. For any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ we let

(when it is well-defined)

$$\pi_0(\mathsf{A}) = \int_{F^{-1}(0)\cap \mathsf{A}} \mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-\min(d,p)}(x) \Big/ \int_{F^{-1}(0)} \mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-\min(d,p)}(x) \ .$$

In what follows we consider the following assumption on F which implies that π_ε is well-defined for any $\varepsilon \geq 0$.

H1 *$F \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^p)$ and $F(0) = 0$. For any $x \in F^{-1}(0)$, $\mathrm{J}F(x) \neq 0$. There exist $\mathtt{m}, \alpha > 0$ and $R \geq 0$ such that for any $x \in \mathbb{R}^d$ with $\|x\| \geq R$, $\|F(x)\| \geq \mathtt{m}\, \|x\|^{\alpha}$.*

A few remarks are in order. First, the assumption $F(0) = 0$ is only technical and can be replaced by $F^{-1}(0) \neq \emptyset$. Similarly, the smoothness assumption $F \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^p)$ allows us to avoid some technicalities in the proofs but can be relaxed. Second, we assume that $x \mapsto \|F(x)\|$ grows at least polynomially when $\|x\| \to +\infty$. This condition can also be relaxed to handle sub-polynomial growth at infinity. However, changing the rate of growth might affect the quantitative convergence properties of $(\pi_\varepsilon)_{\varepsilon>0}$ towards π_0. A study of sub-polynomial growth is left for future work. Finally the assumption that for any $x \in F^{-1}(0)$, $\mathrm{J}F(x) \neq 0$ is necessary in order to define π_0.

Theorem 1 *Assume* **H**1*. Then for any $\varepsilon \geq 0$, π_ε is well-defined. Let $\mathsf{U} \subset \mathbb{R}^d$ be open, bounded and such that $F^{-1}(0) \subset \mathsf{U}$. Let $\varphi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R})$ and $C_\varphi \geq 0$ such that for any $x \in \mathbb{R}^d$*

$$|\varphi(x)| \leq C_\varphi \exp[C_\varphi \, \|x\|^{\alpha k}] \ , \tag{3}$$

with $\alpha > 0$ defined in **H**1 *and $k \in \mathbb{N}^\star$ in (2). Then $\lim_{\varepsilon\to 0} |\pi_\varepsilon[\varphi] - \pi_0[\varphi]| = 0$. In addition, if φ is $M_{1,\varphi}$-Lipschitz on U with $M_{1,\varphi} \geq 0$, then there exist $A \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+), \bar{\varepsilon} \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ such that for any $\varepsilon \in (0, \bar{\varepsilon}(C_\varphi))$ we have*

$$|\pi_\varepsilon[\varphi] - \pi_0[\varphi]| \leq A(C_\varphi)(1 + M_{0,\varphi} + M_{1,\varphi})\varepsilon^{1/k} \ ,$$

with $M_{0,\varphi} = \sup\{|\varphi(x)| \ : \ x \in \mathsf{U}\}$, $A, \bar{\varepsilon}$ functions that do not depend on φ, A is non-decreasing and $\bar{\varepsilon}$ is non-increasing.

Proof We provide a sketch of the proof. The whole proof is postponed to Section 4.1. We first start by showing that we can restrict our study to versions of π_ε and π_0 with truncated support. This is done by studying the decay of $\pi_\varepsilon(\mathsf{K}^\mathrm{c})$ where K is some compact set. Then our study differs depending on if $d \geq p$ or $d \leq p$. If $d \geq p$ then we use tools from geometric measure theory, and in particular the coarea formula in combination with the Lipschitz property of applications of the form $t \mapsto \int_{F^{-1}(t)} \Phi(x)\mathrm{d}\mathcal{H}^{d-p}(x)$ for regular mappings Φ. More precisely, we apply the coarea formula with the mapping F. Doing so, we have $\int_{\mathsf{U}} \exp[-\|F(x)\|^k/\varepsilon]\varphi(x)\mathrm{d}x = \int_{\mathsf{V}} \exp[-\|t\|^k/\varepsilon] \int_{F^{-1}(t)} \varphi(x)\mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-p}(x)\mathrm{d}t$, where U, V are two open sets. Hence, controlling the regularity of $t \mapsto \int_{F^{-1}(t)} \varphi(x)\mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-p}(x)$ is the key to control the convergence rate. In the case $d \leq p$ we adapt arguments from Morse theory and Laplace theory to derive quantitative bounds on $|\pi_\varepsilon[\varphi] - \pi_0[\varphi]|$. These analyses are first conducted with smooth test functions $\varphi \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$ and then we relax this hypothesis using smoothing arguments. ■

We highlight a few key points from this theorem and draw links with the existing literature.

(a) Our result is related to the Laplace-type convergence results of Wong (2001); Korshunov et al. (2015); Barbe (2003); Breitung (2006); Bleistein and Handelsman (1975); Olver (1997); De Bruijn (1981); Evgrafov (2020); Erdélyi (1956); Fedoryuk (1989); Bender and Orszag (2013). In these works, the authors study integrals of the form $\int_{\mathbb{R}^d} \varphi(x) \exp[-U(x)/\varepsilon]\mathrm{d}x$ in the limiting case where $\varepsilon \to 0$ under regularity assumptions on U and φ, and the non-degeneracy condition that for any $x \in \arg\min_{\mathbb{R}^d} U$, $\nabla^2 U(x)$ is invertible. We say that $U : \mathbb{R}^d \to \mathbb{R}_+$ is norm-like with exponent 2 if there exist $p \in \mathbb{N}^\star$ and $F : \mathbb{R}^d \to \mathbb{R}^p$ such that $U(x) = \|F(x)\|^2$. In this case, and under additional regularity assumptions, we have for any $x \in \mathbb{R}^d$

$$\nabla^2 U(x) = 2\mathrm{D}^2 F(x)^\top F(x) + 2\mathrm{D}F(x)^\top \mathrm{D}F(x) \; .$$

In particular for any $x \in F^{-1}(0)$, $\nabla^2 U(x)$ is invertible if and only if $\mathrm{J}F(x) \neq 0$ which is precisely the non-degeneracy condition imposed in **H**1 (note that this directly implies that $d \leq p$). For any $U \in \mathrm{C}^3(\mathbb{R}^d, \mathbb{R}_+)$ with Lipschitz third order derivatives, there exist $p \in \mathbb{N}^\star$ and $F \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^p)$ with Lipschitz derivatives such that for any $x \in \mathbb{R}^d$, $U(x) = \|F(x)\|^2$, see Bony et al. (2006); Fefferman and Phong (1978). Hence, our restriction to functions $U : \mathbb{R}^d \to \mathbb{R}_+$ of the form $U = \|F\|^k$ for some $F : \mathbb{R}^d \to \mathbb{R}^p$ with $p, k \in \mathbb{N}^\star$ is not too constraining under additional regularity assumptions.

(b) We emphasize that our results can be extended to a general potential $U \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}_+)$ if the Hessian of U is invertible on the set $U^{-1}(0)$, thus recovering the usual setting of Laplace-type convergence results.

(c) As highlighted previously, we aim at providing quantitative bounds for $|\pi_\varepsilon[\varphi] - \pi_0[\varphi]|$. In the case where $d \leq p$ we adapt classical arguments from Morse and Laplace theory. Our main contribution is the use of geometric measure theory to cover the case $d \geq p$ which cannot be treated with classical Laplace arguments. The case $d \geq p$ is of particular interest in machine learning applications requiring to sample from distributions of the form $(\mathrm{d}\pi/\mathrm{d}\lambda)(x) \propto \exp[-\|F(x)\|^2]$, where $F : \mathbb{R}^d \to \mathbb{R}^p$ is an encoder neural network with $d \geq p$. In Section 3.1 we investigate this situation in details in the case of generative modeling with macrocanonical and microcanonical distributions.

(d) We notice that choosing $k > 1$ hinders the convergence towards $\pi_0[\varphi]$. The intuition behind this result is that for large values of $k \in \mathbb{N}$ it is harder to distinguish global minimizers from their neighborhood (for example in the case where $F(x) = x$ and $d = p = 1$, we have that $\|F(0.1)\|^1 = 0.1$ whereas $\|F(0.1)\|^3 = 10^{-3}$).

(e) Finally, in Raginsky et al. (2017) we have that $\pi_\varepsilon[U] - \min_{\mathbb{R}^d} U \leq C\varepsilon$ under mild assumptions. In the case where $U = \|F\|^2$ our results suggest that $\pi_\varepsilon[U] - \min_{\mathbb{R}^d} U \leq C\varepsilon^{1/2}$ which appears to be suboptimal. However, our result is more general as it is valid for every locally Lipschitz function. In order to recover the rate $\mathcal{O}(\varepsilon)$ one must modify the proof of Theorem 1 to leverage the fact that $\nabla U(x) = 0$ for any $x \in \mathbb{R}^d$ which is a minimizer of U. More precisely, under the assumptions of Theorem 1 and if $\varphi \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R})$ with $\nabla^\ell \varphi(x) = 0$ for any $\ell \in \{0, \ldots, p\}$ with $p \in \mathbb{N}$ and $x \in \mathbb{R}^d$ which is a minimizer of U, we get that $|\pi_\varepsilon[\varphi] - \pi_0[\varphi]| \leq C_\varphi \varepsilon^{(1+p)/2}$.

As a by-product of Theorem 1 we obtain the following corollary which establishes quantitative bounds for the Wasserstein distance of order 1 between π_ε and π_0.

Corollary 2 *Assume* $\mathbf{H}1$*. Then for any* $\varepsilon \geq 0$*,* π_ε *is well-defined. In addition, there exist* $A_1 \geq 0$ *and* $\bar{\varepsilon} > 0$ *such that for any* $\varepsilon \in [0, \bar{\varepsilon})$ *we have*

$$\mathbf{W}_1(\pi_\varepsilon, \pi_0) \leq A_1 \varepsilon^{1/k} ,$$

where we recall that $k \in \mathbb{N}^\star$ *is defined in* (2)*.*

Proof Denote $\mathrm{Lip} = \{\varphi : \mathbb{R}^d \to \mathbb{R} : |\varphi(x) - \varphi(y)| \leq \|x - y\| \text{ , for any } x, y \in \mathbb{R}^d\}$. In addition, denote $\mathrm{Lip}_0 = \{\varphi : \mathbb{R}^d \to \mathbb{R} : |\varphi(x) - \varphi(y)| \leq \|x - y\| \text{ , for any } x, y \in \mathbb{R}^d, \ \varphi(0) = 0\}$. First, using (Villani, 2009, Theorem 5.10) we have for any $\varepsilon > 0$.

$$\mathbf{W}_1(\pi_\varepsilon, \pi_0) = \sup\{\pi_\varepsilon[\varphi] - \pi_0[\varphi] : \varphi \in \mathrm{Lip}\} = \sup\{|\pi_\varepsilon[\varphi] - \pi_0[\varphi]| : \varphi \in \mathrm{Lip}_0\} . \tag{4}$$

Let $r = \alpha k$ and $\beta_r = \lceil 1/r \rceil$. For any $\varphi \in \mathrm{Lip}_0$ and $x \in \mathbb{R}^d$ with $\|x\| \geq 1$ we have

$$\beta_r \exp[\beta_r \|x\|^{\alpha k}] \geq \beta_r^{\beta_r}/(\beta_r!) \|x\|^{r\beta_r} \geq \|x\| \geq |\varphi(x)| .$$

Similarly, if $\|x\| \leq 1$ we have that $|\varphi(x)| \leq 1 \leq \beta_r \exp[\beta_r \|x\|]$. Hence for any $x \in \mathbb{R}^d$, $|\varphi(x)| \leq \beta_r \exp[\beta_r \|x\|]$. For any $\varphi \in \mathrm{Lip}_0$ we have $M_{1,\varphi} = 1$. In addition, the set $F^{-1}(0)$ is compact since $\lim_{\|x\| \to +\infty} \|F(x)\| = +\infty$ and $F \in \mathrm{C}(\mathbb{R}^d, \mathbb{R}^p)$. Therefore there exists $R > 0$ such that $F^{-1}(0) \subset \mathrm{B}(0, R)$. In particular, note that $\mathsf{U} = \mathrm{B}(0, R)$ is open and bounded. We have that for any $\varphi \in \mathrm{Lip}_0$ and $x \in \mathsf{U}$

$$|\varphi(x)| \leq \|x\| \leq R .$$

Therefore we get that $M_{0,\varphi} = R$. Combining these results with Theorem 1 we have for any $\varphi \in \mathrm{Lip}_0$ and $\varepsilon \in [0, \bar{\varepsilon}(\lceil 1/r \rceil))$

$$|\pi_\varepsilon[\varphi] - \pi_0[\varphi]| \leq A(C_\varphi)(1 + M_{0,\varphi} + M_{1,\varphi})\varepsilon^{1/k} \leq 2A(\lceil 1/r \rceil)(1 + R)\varepsilon^{1/k} .$$

We conclude the proof upon combining this result and (4). ■

Concurrently to our work, Bras and Pagès (2021) establish similar quantitative bounds w.r.t. the Wasserstein distance of order 1 without the norm-like assumption on the potential U but assuming that the Hessian of U is invertible on $\arg\min\{U(x) : x \in \mathbb{R}^d\}$, see (Bras and Pagès, 2021, Lemma 4.6). To do so, the authors derive estimates of the form $\mathbf{W}_1(\pi_{\varepsilon_1}, \pi_{\varepsilon_2}) \leq C|\varepsilon_2^{1/2} - \varepsilon_1^{1/2}|$ with $C \geq 0$ for $\varepsilon_1, \varepsilon_2 > 0$ using a coupling lemma and results from Morse theory. The final bound is obtained by letting $\varepsilon_2 \to 0$ in the previous inequality and the fact that $\lim_{\varepsilon \to 0} \pi_\varepsilon = \pi_0$ weakly. Theorem 2 extends these bounds to the case where the Hessian of U is no longer invertible under a norm-like condition on the potential.

Also, note that Hwang (1980) establishes that $\lim_{\varepsilon \to 0} \pi_\varepsilon = \pi_0$ weakly without any norm-like assumption on U under the condition that $\mathsf{C}^\star = \arg\min\{U(x) : x \in \mathbb{R}^d\}$ can be partitioned into a collection of manifolds and that for any $x \in \mathsf{C}^\star$, the Hessian of U along the normal plan to the manifold associated with x is invertible (we denote by $\nabla^2_{\mathrm{N}} U(x)$ this quantity). In Theorem 23, we show that if $U = \|F\|^2$ then we have that $\det(\nabla^2_{\mathrm{N}} U(x)) = \mathrm{J}F(x)^2$ for any $x \in F^{-1}(0)$. Hence, the invertibility condition can be written as: for any $x \in F^{-1}(0)$, $\mathrm{J}F(x) \neq 0$ similarly to $\mathbf{H}\,1$ and the limiting measure identified by Hwang (1980) is exactly π_0.

The next result is an extension of Theorem 1 where, given $k \in \mathbb{N}^\star$, $\Psi : \mathbb{R}^d \to \mathbb{R}_+$, π_ε is replaced by π_ε^Ψ defined for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$\pi_\varepsilon^\Psi(\mathsf{A}) = \textstyle\int_{\mathsf{A}} \Psi(x) \exp[-\|F(x)\|^k/\varepsilon]\mathrm{d}x \Big/ \textstyle\int_{\mathbb{R}^d} \Psi(x) \exp[-\|F(x)\|^k/\varepsilon]\mathrm{d}x \ . \tag{5}$$

Similarly, for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ we let (when it is well-defined)

$$\pi_0^\Psi(\mathsf{A}) = \textstyle\int_{F^{-1}(0)\cap\mathsf{A}} \Psi(x)\mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-\min(d,p)}(x) \Big/ \textstyle\int_{F^{-1}(0)} \Psi(x)\mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-\min(d,p)}(x) \ . \tag{6}$$

We consider the following assumption on Ψ which ensures that π_ε^Ψ is well-defined for $\varepsilon \geq 0$ when combined with **H**1.

H2 $\Psi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R}_+)$, *there exists* C_Ψ *such that for any* $x \in \mathbb{R}^d$, $\Psi(x) \leq C_\Psi \exp[C_\Psi \|x\|^{\alpha k}]$ *and* $\Psi(x) > 0$ *for any* $x \in F^{-1}(0)$.

Note that in **H**2, $\alpha > 0$ is given in **H**1 and $k \in \mathbb{N}^\star$ is given in (2). Under this assumption, we derive the following quantitative bounds.

Theorem 3 *Assume* **H**1 *and* **H**2. *Then for any* $\varepsilon \geq 0$, π_ε^Ψ *is well-defined. Let* $\mathsf{U} \subset \mathbb{R}^d$ *be open, bounded and such that* $F^{-1}(0) \subset \mathsf{U}$. *Let* $\varphi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R})$ *and* $C_\varphi \geq 0$ *such that for any* $x \in \mathbb{R}^d$

$$|\varphi(x)| \leq C_\varphi \exp[C_\varphi \|x\|^{\alpha k}] \ , \tag{7}$$

with $\alpha > 0$ *defined in* **H**1 *and* $k \in \mathbb{N}^\star$ *in* (2). *Then* $\lim_{\varepsilon\to 0} \left|\pi_\varepsilon^\Psi[\varphi] - \pi_0^\Psi[\varphi]\right| = 0$. *In addition, if* φ *and* Ψ *are respectively* $M_{1,\varphi}$*-Lipschitz and* $M_{1,\Psi}$*-Lipschitz on* U *with* $M_{1,\varphi}, M_{1,\Psi} \geq 0$, *then there exist* $A \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$, $\bar{\varepsilon} \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+^\star)$ *such that for any* $\varepsilon \in (0, \bar{\varepsilon}(C_\varphi))$ *we have*

$$|\pi_\varepsilon^\Psi[\varphi] - \pi_0^\Psi[\varphi]| \leq A(C_\varphi)(1 + M_{0,\varphi} + M_{1,\varphi})\varepsilon^{1/k} \ ,$$

with $M_{0,\varphi} = \sup\{|\varphi(x)| \ : \ x \in \mathsf{U}\}$, $A, \bar{\varepsilon}$ *functions that do not depend on* φ, A *non-decreasing and* $\bar{\varepsilon}$ *non-increasing.*

Proof The proof is postponed to Section 4.1. ■

The same remarks formulated after Theorem 1 hold for this extension. In particular the norm-like assumption on U can be omitted and replaced by a classical invertibility condition on the Hessian at $U^{-1}(0)$. Note that Theorem 1 is a direct consequence of Theorem 3 upon letting $\Psi = 1$. Reciprocally, we do not have that Theorem 3 is a consequence of Theorem 1 upon replacing φ by $\varphi\Psi$. Indeed, it must be noted that, contrary to φ, Ψ also appears in the normalizing constant of π_ε^Ψ. An important consequence of Theorem 3 is the case where for any $x \in \mathbb{R}^d$, $\Psi(x) = \mathrm{J}F(x)$. Indeed, doing so we obtain that π_0^Ψ is the uniform distribution on $F^{-1}(0)$, *i.e.* the maximum entropy distribution with support $F^{-1}(0)$. We discuss this special case in Section 3.1. Similarly to Theorem 2 we can also establish the following corollary.

Corollary 4 *Assume* $\mathbf{H}1$ *and* $\mathbf{H}2$*. Then for any* $\varepsilon \geq 0$, π_ε^Ψ *is well-defined. In addition, assume that* Ψ *is Lipschitz continuous. Then there exist* $A_2 \geq 0$ *and* $\bar{\varepsilon} > 0$ *such that for any* $\varepsilon \in (0, \bar{\varepsilon})$ *we have*

$$\mathbf{W}_1(\pi_\varepsilon^\Psi, \pi_0^\Psi) \leq A_2 \varepsilon^{1/k} ,$$

where we recall that $k \in \mathbb{N}^\star$ *is defined in* (2).

Proof The proof is similar to the one of Theorem 2. ■

We conclude this section with a first application of our results to the study of the concentration of the posterior of a Variational AutoEncoder (VAE). Let $p, d \in \mathbb{N}$ with $d > p$. Assume we have access to a dataset $x_{1:n} = \{x_i\}_{i=1}^n \in (\mathbb{R}^d)^n$ with $n \in \mathbb{N}$ such that $\{x_i\}_{i=1}^n$ are i.i.d. samples from $\mu \in \mathscr{P}(\mathbb{R}^d)$ a target data distribution. We consider a generative model to approximately sample from μ defined as follows: let η be an easy-to-sample distribution in $\mathbb{R}^p$ with density p_η and for any $z \in \mathsf{Z}$, let $p_\theta(\cdot|z) = \mathcal{N}(m_\theta(z), \varepsilon \operatorname{Id}/2)$ with $\varepsilon > 0$ and $m_\theta : \mathbb{R}^p \to \mathbb{R}^d$ given by a neural network [1] with parameters $\theta \in \Theta$. The generative model is trained by maximizing an Evidence Lower Bound (ELBO) which requires introducing an encoding probability distribution. Assuming that the generative model is trained we obtain a set of parameters $\theta^\star \in \Theta$ and the generative model is then given for any $x \in \mathbb{R}^d$ by

$$p_{\theta^\star}(x) = \textstyle\int_{\mathbb{R}^p} p_\eta(z) p_{\theta^\star}(x|z) \mathrm{d}z .$$

For encoding purposes, we are also interested in the posterior distribution given for any $x \in \mathbb{R}^d$, $z \in \mathbb{R}^p$ by $p_{\theta^\star}(z|x) = p_\eta(z) p_{\theta^\star}(x|z)/p_{\theta^\star}(x)$. Note that for any $x \in \mathbb{R}^d$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^p)$ we have

$$\textstyle\int_{\mathsf{A}} p_{\theta^\star}(z|x) \mathrm{d}z = \int_{\mathsf{A}} p_\eta(z) \exp[-\|x - m_\theta(z)\|^2/\varepsilon] \mathrm{d}z / \int_{\mathbb{R}^p} p_\eta(z) \exp[-\|x - m_\theta(z)\|^2/\varepsilon] \mathrm{d}z .$$

In particular, we have that for any $x \in \mathbb{R}^d$ the distribution $\pi_{x,\varepsilon}$ with density $p_{\theta^\star}(\cdot|x)$ is of the form (5) with $\Psi \leftarrow p_\eta$ and $F_x(z) = x - m_\theta(z)$. Under $\mathbf{H}1$ and $\mathbf{H}2$, we define $\pi_{x,0}$ such that for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^p)$

$$\pi_{x,0}(\mathsf{A}) = \textstyle\int_{m_\theta^{-1}(x) \cap \mathsf{A}} p_\eta(z) \mathrm{J}F_x(z)^{-1} \mathrm{d}\mathcal{H}^0(z) \Big/ \int_{m_\theta^{-1}(x)} p_\eta(z) \mathrm{J}F_x(z)^{-1} \mathrm{d}\mathcal{H}^0(z) \; .$$

Assuming that p_η is Lipschitz, we can apply Theorem 4 and there exist $A_{2,x} \geq 0$ and $\bar{\varepsilon}_x > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon}_x)$ we have

$$\mathbf{W}_1(\pi_{x,\varepsilon}, \pi_{x,0}) \leq A_{2,x} \varepsilon^{1/2} .$$

This result provides quantitative convergence bounds for the posterior distribution $p_{\theta^\star}(\cdot|x)$ towards a limiting distribution $\pi_{x,0}$. Note that $\pi_{x,0}$ will assign more mass on the points where (a) the prior distribution p is high ; (b) the generalized Jacobian is small, i.e. the mapping $m_{\theta^\star}$ does not fluctuate too much around $z \in m_{\theta^\star}^{-1}(x)$.

---

[1] Note that in practice the covariance matrix is also parameterized by a neural network but for simplicity we keep it fixed.

3. Applications

In this section, we present two applications of our results. First, we draw links between two different maximum entropy distributions in Section 3.1. Then, we study the convergence of SGLD for non-convex minimization in Section 3.2.

3.1 Maximum entropy distributions

3.1.1 Maximum entropy distributions

Let $\mathsf{X} \subset \mathbb{R}^d$ be a compact space with $\mathrm{int}(\mathsf{X}) \neq \emptyset$ and $G : \mathsf{X} \to \mathbb{R}^p$ measurable. The maximum entropy distribution with constraint G is denoted π_{Ent} and is given by

$$\pi_{\mathrm{Ent}} = \arg\max\{H(\pi) \, : \, \pi \in \mathscr{P}_{\lambda}(\mathsf{X}), \ \pi[\|G\|] < +\infty, \ \pi[G] = 0\} \ , \tag{8}$$

where $\mathscr{P}_{\lambda}(\mathsf{X})$ is the set of probability measures on X which admit a density w.r.t. the Lebesgue measure, $\pi[G] = \int_{\mathsf{X}} G(x)\mathrm{d}\pi(x)$ and $H(\pi) = -\int_{\mathsf{X}} \log(\mathrm{d}\pi/\mathrm{d}\lambda)(x)\mathrm{d}\pi(x)$ for any $\pi \in \mathscr{P}_{\lambda}(\mathsf{X})$, where λ is the Lebesgue measure on $\mathbb{R}^d$. Note that $H(\pi)$ is well-defined (but can be infinite) since X is compact and in this case $\mathrm{KL}(\pi|\mu) = -H(\pi) + \log(\lambda(\mathsf{X}))$, where μ is the uniform measure on X and we recall that λ is the Lebesgue measure.

Such maximum entropy distributions naturally arise in many areas such as statistical physics (Jaynes, 1957; Bruna and Mallat, 2018), econometrics (Golan, 2008), generative modeling (Portilla and Simoncelli, 2000; Lu et al., 2016; De Bortoli et al., 2021), reinforcement learning (Ziebart et al., 2008) or image processing (Geman and Geman, 1984; Desolneux and Leclaire, 2017). In texture synthesis applications such as Lu et al. (2016); De Bortoli et al. (2021), the feature mapping G is given by a pretrained neural network as in Gatys et al. (2015). In what follows, we introduce two extensions of (8). More precisely, we use the analogy between the maximization of the entropy and the minimization of an appropriate Kullback-Leibler divergence to extend (8) on non-compact spaces, following De Bortoli et al. (2021).

Macrocanonical distributions Given $\mu \in \mathscr{P}(\mathbb{R}^d)$ and $G : \mathbb{R}^d \to \mathbb{R}^p$ measurable, we say that $\pi^\star \in \mathscr{P}(\mathbb{R}^d)$ is a macrocanonical distribution (Bruna and Mallat, 2018) with reference measure μ and constraint G if it satisfies

$$\pi^\star \in \arg\min\{\mathrm{KL}(\pi|\mu) \, : \, \pi \in \mathscr{P}(\mathbb{R}^d), \ \pi[\|G\|] < +\infty, \ \pi[G] = 0\} \ .$$

Using results from information geometry (Csiszár, 1975) any macrocanonical distribution can be written as an exponential distribution under mild assumptions.

Proposition 5 ((De Bortoli et al., 2021)) *Assume that $G \in \mathrm{C}(\mathbb{R}^d, \mathbb{R}^p)$ and that there exist $\mathsf{\alpha}, \mathsf{\beta}, \eta > 0$ with $\mathsf{\beta} > \mathsf{\alpha}$ such that*

$$\sup\{\|G(x)\|(1 + \|x\|^{\mathsf{\alpha}})^{-1} \, : \, x \in \mathbb{R}^d\} < +\infty \ , \qquad \int_{\mathbb{R}^d} \exp[\eta\|x\|^{\mathsf{\beta}}]\mathrm{d}\mu(x) < +\infty \ .$$

In addition, assume that for any $\theta \in \mathbb{R}^p$ with $\|\theta\| = 1$, we have $\mu(\{x \in \mathbb{R}^d \, : \, \langle \theta, G(x) \rangle < 0\}) > 0$. Then there exists a unique macrocanonical distribution with constraint G and reference measure μ denoted $\pi^\star$. In addition, there exists $\theta^\star \in \mathbb{R}^p$ such that for any $x \in \mathbb{R}^d$

$$(\mathrm{d}\pi^\star/\mathrm{d}\mu)(x) = \exp[-\langle \theta^\star, G(x) \rangle - L(\theta^\star)] \ , \qquad L(\theta^\star) = \log \int_{\mathbb{R}^d} \exp[-\langle \theta^\star, G(x) \rangle]\mathrm{d}\mu(x) \ .$$

We refer to De Bortoli et al. (2021) for a relaxation of these conditions and a detailed study of the macrocanonical distribution. Theorem 5 ensures that macrocanonical distributions can be written as Gibbs measures. We emphasize that these distributions satisfy a constraint in expectation. An alternative way to define maximum entropy distributions under constraints is to ensure almost sure equality. This tighter constraint yields new maximum entropy distributions called microcanonical distributions.

Microcanonical distributions A microcanonical distribution $\pi^\star \in \mathscr{P}(\mathbb{R}^d)$ with reference measure μ and constraint G can be defined as

$$\pi^\star \in \arg\min\{\mathrm{KL}(\pi|\mu) \ : \ \pi \in \mathscr{P}(\mathbb{R}^d), \ \pi[\|G\|] = 0\} \ . \tag{9}$$

We emphasize that imposing (9) is equivalent to impose that $\pi^\star$ has minimum Kullback-Leibler divergence w.r.t. μ among all distributions which satisfy $G = 0$ almost surely. Note that if there exists a microcanonical distribution $\pi^\star$ with $\mathrm{KL}(\pi^\star|\mu) < +\infty$ then $\mu(\mathsf{A}) > 0$ where $\mathsf{A} = \{x \in \mathbb{R}^d \ : \ G(x) = 0\}$ and in this case $\pi^\star = \mu(\cdot \cap \mathsf{A})/\mu(\mathsf{A})$. In what follows, we will say that $\pi^\star$ is the uniform microcanonical distribution with constraint G if $G^{-1}(0)$ is compact with $\mathcal{H}^{d-\min(d,p)}(G^{-1}(0)) < +\infty$ and $\pi^\star$ is the uniform distribution on $G^{-1}(0)$, *i.e.* for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$, we have

$$\pi^\star(\mathsf{A}) = \textstyle\int_{G^{-1}(0)\cap\mathsf{A}} \mathrm{d}\mathcal{H}^{d-\min(d,p)}(x)/\mathcal{H}^{d-\min(d,p)}(G^{-1}(0)) \ . \tag{10}$$

Note that contrary to macrocanonical distributions which admit a representation as a Gibbs measure, see Theorem 5, the microcanonical distributions are concentrated on the set $G^{-1}(0)$. In the next section, we draw links between these two maximum entropy distributions.

3.1.2 From macrocanonical to microcanonical

In order to draw links between macrocanonical and microcanonical distributions we consider a specific sequence of macrocanonical distributions associated with the constraint of the form $G_\varepsilon = \|F\|^2 - \varepsilon$ for $\varepsilon > 0$, and where $F : \ \mathbb{R}^d \to \mathbb{R}^p$. Let $\mu \in \mathscr{P}(\mathbb{R}^d)$ be a reference probability measure. Under the conditions of Theorem 5, we define a family of measures $\{\rho_\varepsilon\}_{\varepsilon>0}$ such that for any $\varepsilon > 0$, ρ_ε is the macrocanonical distribution associated with G_ε and μ. More precisely, for any $\varepsilon > 0$, there exists $\theta_\varepsilon \in \mathbb{R}$ such that for any $x \in \mathbb{R}^d$

$$(\mathrm{d}\rho_\varepsilon/\mathrm{d}\mu)(x) = \exp[-\theta_\varepsilon\|F(x)\|^2 - L_\varepsilon] \ , \quad L_\varepsilon = \log \textstyle\int_{\mathbb{R}^d} \exp[-\theta_\varepsilon\|F(x)\|^2]\mathrm{d}\mu(x) \ .$$

Our goal is to study the behavior of the family $\{\rho_\varepsilon\}_{\varepsilon>0}$ when $\varepsilon \to 0$. In particular, we show that there exists a reference measure μ such that the limit $\pi_0 = \lim_{\varepsilon\to0}\rho_\varepsilon$ exists where π_0 is the uniform microcanonical distribution (10). We start with the following proposition which ensures that $\lim_{\varepsilon\to0}\theta_\varepsilon = +\infty$ with linear rate.

Proposition 6 *Let $\mu \in \mathscr{P}(\mathbb{R}^d)$ and $F : \ \mathbb{R}^d \to \mathbb{R}^p$. Assume that the conditions of Theorem 5 with $G_\varepsilon = \|F\|^2 - \varepsilon$ for any $\varepsilon > 0$ are satisfied and let ρ_ε be the macrocanonical distribution with constraint G_ε and reference measure μ. Assume that there exists $\Psi : \ \mathbb{R}^d \to \mathbb{R}_+$ such that μ admits a density w.r.t. the Lebesgue measure given by Ψ. In addition, assume that* **H**1 *and* **H**2 *hold. Then, we have that $\theta_\varepsilon \sim_{\varepsilon\to0} p/(2\varepsilon)$.*

Proof We provide a sketch of the proof. The whole proof is postponed to Section 4.2. In the case where $d \geq p$ we use tools from geometric measure theory to derive an equivalent of $\int_{\mathbb{R}^d} \Psi(x)\|F(x)\|^2 \exp[-\|F(x)\|^2/\varepsilon]\mathrm{d}x$ when $\varepsilon \to 0$ akin to Theorem 12. In the case $d \leq p$, using Morse lemma and classical Laplace analysis we show a similar result. We conclude upon combining these results and that $\rho_\varepsilon[\|F\|^2] = \varepsilon$. ∎

Note that the equivalence in Theorem 6 does not depend on the reference measure μ. In Section 4.2, we prove an extension of Theorem 6, see Theorem 18. In particular we show that $G_\varepsilon = \|F\|^2 - \varepsilon$ can be replaced by $G_\varepsilon = \|F\|^k - \varepsilon$ with $k \in \mathbb{N}^\star$ under similar conditions. In particular, we get that for any $k \in \mathbb{N}^\star$, $\theta_\varepsilon \sim_{\varepsilon \to 0} \mathcal{C}_k/\varepsilon$ with $\mathcal{C}_k$ explicit if $k = 2$. Finally note that under the same conditions as Theorem 6, we have that $\pi_\varepsilon^\Psi[\|F\|^2] \sim_{\varepsilon \to 0} p/(2\varepsilon)$, where we recall that for any $\varepsilon > 0$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$

$$\pi_\varepsilon^\Psi(\mathsf{A}) = \int_{\mathsf{A}} \Psi(x) \exp[-\|F(x)\|^2/\varepsilon]\mathrm{d}x / \int_{\mathbb{R}^d} \Psi(x) \exp[-\|F(x)\|^2/\varepsilon]\mathrm{d}x \ .$$

This comes from the facts that $\rho_\varepsilon = \pi_{\theta_\varepsilon}^\Psi$ and $\rho_\varepsilon[\|F\|^2] = \varepsilon$.

The following proposition establishes quantitative bounds w.r.t. the Wasserstein of order 1 between $\{\rho_\varepsilon\}_{\varepsilon>0}$ and a limiting measure. We obtain this result upon combining Theorem 6 and Theorem 2.

Proposition 7 *Assume the same conditions as Theorem 6. Then there exist $\bar{\varepsilon} > 0$ and $A_3 \geq 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon}]$ we have*

$$\mathbf{W}_1(\rho_\varepsilon, \pi_0^\Psi) \leq A_3 \varepsilon^{1/2} \ ,$$

where π_0^Ψ is given in (6).

Theorem 7 establishes quantitative bounds between the family of macrocanonical distributions $\{\rho_\varepsilon\}_{\varepsilon>0}$ and a limiting distribution π_0^Ψ. Note that this limiting distribution is *not* the uniform microcanonical distribution $\pi^\star$ defined in (10) in general, even though it is supported on the set $F^{-1}(0)$. However, it is possible to sample from this $\pi^\star$ by choosing μ such that for any $x \in F^{-1}(0)$, $\Psi(x) = (\mathrm{d}\mu/\mathrm{d}\lambda)(x) = \mathrm{J}F(x)$.

3.1.3 Some simple experiments

Methodology In this experimental section we consider two simple examples of functions $F : \mathbb{R}^d \to \mathbb{R}^p$ with $d = 1$ or 2 and $p = 1$. We consider the associated distributions $\{\pi_\varepsilon\}_{\varepsilon>0}$ and $\{\pi_\varepsilon^\Psi\}_{\varepsilon>0}$ and their limit as ε goes to 0, showing in particular that for $\Psi = \mathrm{J}F$, it converges to the uniform microcanonical distribution that is the uniform distribution on $F^{-1}(0)$. We also check experimentally the scaling relation of Proposition 6.

Zeros of a polynomial In this first example, let $P : \mathbb{R} \to \mathbb{R}$ be a polynomial. We are interested in the zeros of P, and, more precisely in sampling in a uniform way on the set of the zeros of P. Using the notations of the previous sections, we set $F = P$. For $\varepsilon > 0$, let us define the two distributions, with respective densities with respect to the Lebesgue measure on $\mathbb{R}$ given for any $x \in \mathbb{R}$ by

$$(\mathrm{d}\pi_\varepsilon/\mathrm{d}\lambda)(x) = \exp[-F(x)^2/\varepsilon] / \int_{\mathbb{R}} \exp[-F(\tilde{x})^2/\varepsilon]\mathrm{d}\tilde{x} \ ,$$
$$(\mathrm{d}\pi_\varepsilon^\Psi/\mathrm{d}\lambda)(x) = \mathrm{J}F(x) \exp[-F(x)^2/\varepsilon] / \int_{\mathbb{R}} \mathrm{J}F(\tilde{x}) \exp[-F(\tilde{x})^2/\varepsilon]\mathrm{d}\tilde{x} \ ,$$

where for any $x \in \mathbb{R}$, $\Psi(x) = \mathrm{J}F(x) = |P'(x)|$.

In Figure 1, we display the polynomial P where for any $x \in \mathbb{R}$, $P(x) = x(x-0.5)(x-1.7)(x-2.5)$. We also check numerically the scaling relation of Proposition 6. Then, in Figure 2, we illustrate the different behaviors of π_ε and π_ε^Ψ when ε is small ($\varepsilon = 10^{-3}$ in our experiments). We also present the histograms obtained using of $N = 10^4$ independent samples of π_ε and of π_ε^Ψ. These samples were simply obtained by the numerical CDF inversion method. The stars indicate the target limit distributions (that is the uniform distribution on the zeros of P for the limit of π_ε^Ψ).

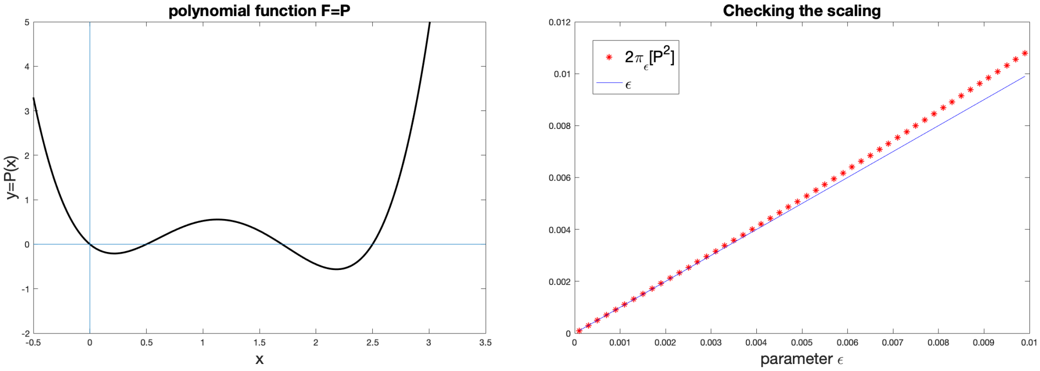

Figure 1: Left: graph of the polynomial $x \mapsto P(x)$, that has 4 zeros. Right: verifying the scaling relation of Theorem 6 by plotting $\varepsilon \mapsto 2\pi_\varepsilon[P^2]$, which is equivalent to ε as ε goes to 0.

Two-dimensional ellipse In this second example, we consider the function $F : \mathbb{R}^2 \to \mathbb{R}$ given for any $x = (x_1, x_2) \in \mathbb{R}^2$ by $F(x) = a_1 x_1^2 + a_2 x_2^2 - 1$ with $a_1, a_2 > 0$. For $\varepsilon > 0$, we define π_ε and π_ε^Ψ whose densities w.r.t. the Lebesgue measure are given for any $x \in \mathbb{R}^2$ by

$$(\mathrm{d}\pi_\varepsilon/\mathrm{d}\lambda)(x) = \exp[-\|F(x)\|^2/\varepsilon] / \textstyle\int_{\mathbb{R}^2} \exp[-\|F(\tilde{x})\|^2/\varepsilon] \mathrm{d}\tilde{x} \, ,$$
$$(\mathrm{d}\pi_\varepsilon^\Psi/\mathrm{d}\lambda)(x) = \mathrm{J}F(x) \exp[-\|F(x)\|^2/\varepsilon] / \textstyle\int_{\mathbb{R}^2} \mathrm{J}F(\tilde{x}) \exp[-\|F(\tilde{x})\|^2/\varepsilon] \mathrm{d}\tilde{x} \, .$$

To sample from π_ε and π_ε^Ψ, we use two Markov chains given by the Unadjusted Langevin Algorithm (Roberts and Tweedie, 1996; Durmus and Moulines, 2017; Dalalyan, 2017), given respectively by $X_0, Y_0 \in \mathbb{R}^2$ and the recursions

$$X_{n+1} = X_n - \gamma(\nabla\|F\|^2(X_n)/\varepsilon) + \sqrt{2\gamma} Z_{n+1} \, ,$$
$$Y_{n+1} = Y_n - \gamma(\nabla\|F\|^2(Y_n)/\varepsilon - \nabla \log \mathrm{J}F(Y_n)) + \sqrt{2\gamma} Z_{n+1} \, ,$$

where $\gamma > 0$ is a stepsize and $\{Z_n\}_{n \in \mathbb{N}}$ is a family of independent Gaussian random variables with zero mean and identity covariance matrix. For any $x \in \mathbb{R}^2$ we have

$$\mathrm{J}F(x) = 2(a_1^2 x_1^2 + a_2^2 x_2^2)^{1/2} \, , \quad \nabla\|F\|^2(x) = 2F(x)\nabla F(x) = 4(a_1 x_1^2 + a_2 x_2^2 - 1) \begin{pmatrix} a_1 x_1 \\ a_2 x_2 \end{pmatrix} \, ,$$
$$\nabla \log \mathrm{J}F(x) = (a_1^2 x_1^2 + a_2^2 x_2^2)^{-1} \begin{pmatrix} a_1^2 x_1 \\ a_2^2 x_2 \end{pmatrix} \, .$$

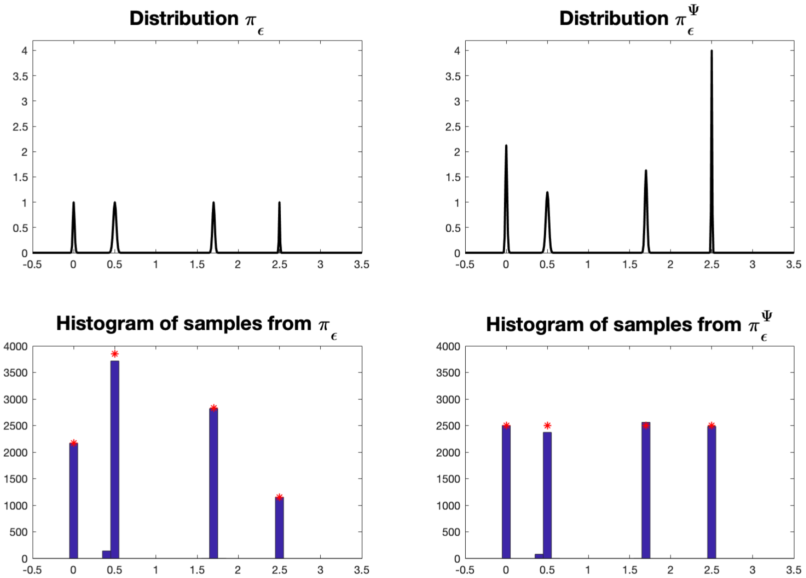

Figure 2: Distributions π_ε and π_ε^Ψ (first line), and histogram of their samples from them (second line). This experiment shows that the limit distribution of π_ε^Ψ as ε goes to 0 is the uniform microcanonical distribution given by the uniform distribution on the zeros of P.

The set $F^{-1}(0)$ is an ellipse, of cartesian equation $a_1 x_1^2 + a_2 x_2^2 = 1$ and has a polar parametrization given for any $\theta \in [0, 2\pi)$ by

$$r(\theta) = (a_2 + (a_1 - a_2)\cos^2\theta)^{1/2} \ .$$

This is not an arc-length parametrization, and the infinitesimal length of the curve element between θ and $\theta + d\theta$ is given by $\ell(\theta)d\theta$ where for any $\theta \in (0, 2\pi]$ we have

$$\ell(\theta) = (r(\theta)^2 + r'(\theta)^2)^{1/2} = (a_2^2 + (a_1^2 - a_2^2)\cos^2\theta)^{1/2}(a_2 + (a_1 - a_2)\cos^2\theta)^{-3/2} \ .$$

In Figure 3, we set $a_1 = 1$, $a_2 = 4$, $\varepsilon = 10^{-3}$, $\gamma = 10^{-5}$, $N = 9 \times 10^7$ iterations in the Markov chains, and histograms with bins of size 0.05π. In Figure 3, we compare two different histograms: the ones of the angle values (θ) for the points of the chain X_n (diamonds), of the chain Y_n (stars), and the two distributions with density $\theta \mapsto \ell(\theta)/\int_{-\pi}^{\pi}\ell(\theta)\mathrm{d}\theta$ (black curve) and $\theta \mapsto \ell(\theta)\mathrm{J}F(\theta)^{-1}/\int_{-\pi}^{\pi}\ell(\theta)\mathrm{J}F(\theta)^{-1}\mathrm{d}\theta$ (red curve). Note that $\theta \mapsto \ell(\theta)/\int_{-\pi}^{\pi}\ell(\theta)\mathrm{d}\theta$ corresponds to the Hausdorff measure on the ellipse pushed by the polar parametrisation $\theta \mapsto r(\theta)$, *i.e.* the uniform microcanonical distribution pushed by $\theta \mapsto r(\theta)$. We observe that the Markov chain corrected with the generalized Jacobian indeed achieves the uniform microcanonical distribution on the set $F^{-1}(0)$. Finally, we also check experimentally in Figure 4 the scaling relation of Proposition 6.

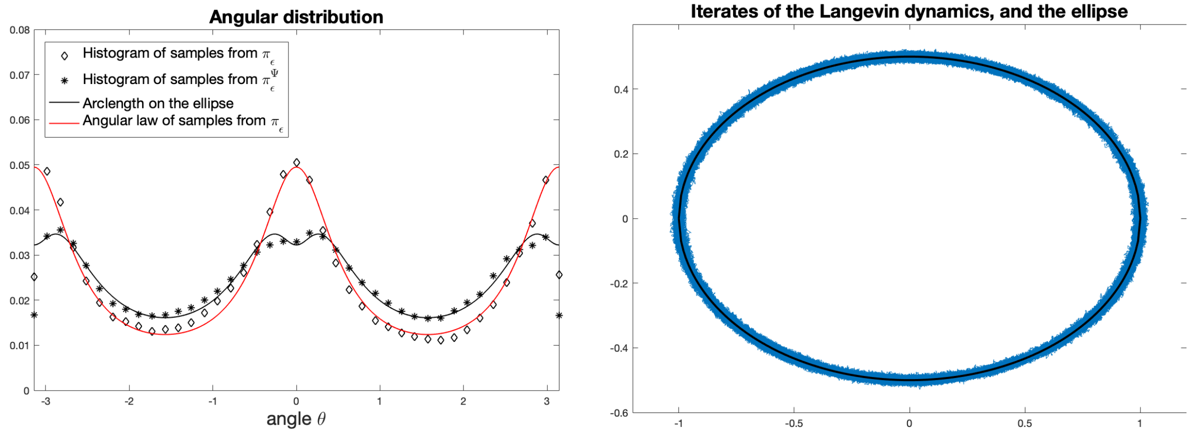

Figure 3: Left: histogram of angles of the samples $(X_n)_{n\in\mathbb{N}}$ and $(Y_n)_{n\in\mathbb{N}}$, showing that π_ε^Ψ is close to the uniform distribution on the ellipse for small values of $\varepsilon > 0$. Right: the trajectory of the first 10^6 samples of Y_n (with a burnin of 10^3 iterates).

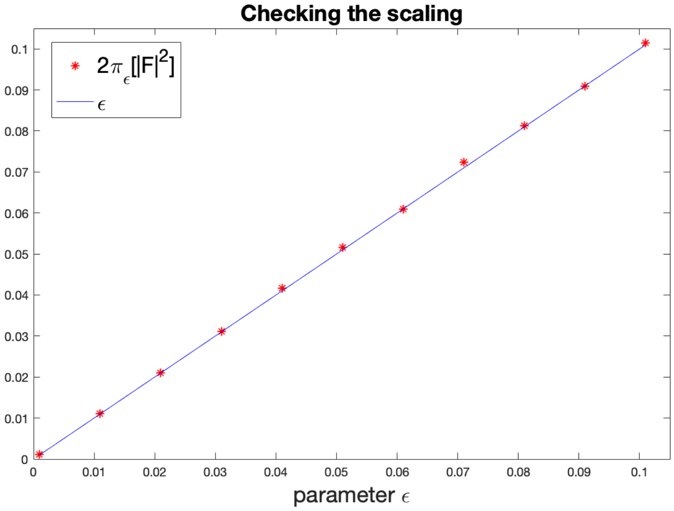

Figure 4: Checking the scaling relation of Proposition 6.

3.2 Non-convex minimization

3.2.1 Non-convex setting and related work

In this section, we consider the following minimization problem:

$$\text{find } x^\star \in \arg\min\{U(x) \,:\, x \in \mathbb{R}^d\} ,$$

with $U \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$ and $\arg\min\{U(x) \,:\, x \in \mathbb{R}^d\} \neq \emptyset$. Here, we do not assume that U is convex. Hence, classical first-order optimization schemes such as gradient descent might get trapped in saddle points. Adding isotropic Gaussian noise to this dynamics circumvents this issue, see Brandière and Duflo (1996); Pelletier (1998) for instance. The algorithm is then given by the following recursion: $X_0 \in \mathbb{R}^d$ and for any $k \in \mathbb{N}$

$$X_{k+1} = X_k - \gamma_k \nabla U(X_k) + Z_{k+1} , \tag{11}$$

where $(Z_k)_{k\in\mathbb{N}}$ is a sequence of independent Gaussian random variables with zero mean and covariance matrix $\sigma_k^2 \,\mathrm{Id}$, with $(\gamma_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+)^{\mathbb{N}}$ a sequence of stepsizes and $(\sigma_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+)^{\mathbb{N}}$. If $\sigma_k^2 = \gamma_k^2$ for any $k \in \mathbb{N}$, $\sum_{k\in\mathbb{N}} \gamma_k = +\infty$ and $\sum_{k\in\mathbb{N}} \gamma_k^2 < +\infty$ then the algorithm is in the *weakly disturbed* regime and under additional assumptions on U one can show that $(X_k)_{k\in\mathbb{N}}$ converges almost surely to a local minimizer of U, see Pelletier (1998) for instance. However, $(U(X_k)_{k\in\mathbb{N}})_{k\in\mathbb{N}}$ does not necessarily converge to the global minimum of U. The intuition behind this behavior is that the variance of the noise decreases too quickly for the sequence to explore efficiently the landscape of U.

In order to perform global optimization, one can consider *simulated annealing* algorithms where $\sigma_k^2 = \gamma_k T_k$ for any $k \in \mathbb{N}$, with $(T_k)_{k\in\mathbb{N}} \in (\mathbb{R}_+)^{\mathbb{N}}$ a sequence of temperatures which slowly decrease. These algorithms were introduced in the context of discrete optimization in Kirkpatrick (1984) and have been thoroughly investigated in Gidas (1985) (discrete state-space), Gelfand and Mitter (1991, 1993) (discrete-time algorithm), Geman and Hwang (1986); Chiang et al. (1987); Holley et al. (1989) (diffusion), Pelletier (1998) (CLT type results), Romeijn and Smith (1994) (constrained optimization), Yang (2000) (non Gaussian noise) and Andrieu et al. (2001) (control of the sequence in total variation) for instance. In Hajek (1988, 1985) a sufficient and necessary condition is given on the rate of decrease of $(T_k)_{k\in\mathbb{N}}$ so that $U(X_k)_{k\in\mathbb{N}}$ converges towards the minimum of U, more precisely the condition reads $\sum_{k\in\mathbb{N}} \exp[-b^\star/T_k] = +\infty$, where $b^\star > 0$ is a parameter which depends only on U called the *kinetic barrier* (or depth in Hajek (1988, 1985)). One of the main limiting factor of simulated annealing is the slow rate of convergence of $(T_k)_{k\in\mathbb{N}}$ towards 0, which is often set as $T_k = C/\log(k+1)$ for any $k \in \mathbb{N}$ and for some constant $C \geq 0$. Note that recently the convergence of modifications of the simulated annealing algorithms under faster cooling rates have been investigated in discrete spaces by either changing the Markov chain transitions (Choi, 2019) or the energy landscape (Choi, 2020), see also Catoni (1998). Finally, we emphasize that most of the results regarding the convergence of (11) can be extended to the case where ∇U is replaced by an unbiased estimator under additional conditions.

With the advent of neural networks, numerous schemes exploiting the annealing structure have been proposed to minimize the non-convex losses which arise in deep learning applications, see Ye et al. (2017) for instance. Drawing connections with unadjusted Langevin algorithms, and in particular Stochastic Gradient Langevin Dynamics (SGLD) (Welling and Teh, 2011), Raginsky et al. (2017); Zhang et al. (2017) replace ∇U by an unbiased estimator and let $\gamma_k = \gamma > 0$, $\sigma_k^2 = 2\gamma\varepsilon$ for any $k \in \mathbb{N}$ with $\varepsilon > 0$ in (11). Under curvature and regularity assumptions on the potential U, the authors derive quantitative bounds on $(\mathbb{E}[U(X_k) - \min_{\mathbb{R}^d} U])_{k\in\mathbb{N}}$. Since then several accelerations have been proposed in the literature (Chau and Rasonyi, 2019; Gao et al., 2018; Erdogdu et al., 2018; Nguyen et al., 2019; Zhang et al., 2019; Xu et al., 2017).

In the next section, we improve the results of Raginsky et al. (2017) by providing upper bounds w.r.t. the first order Wasserstein distance between the distribution of $(X_k)_{k\in\mathbb{N}}$ given by SGLD and an explicit limiting distribution. Our results complete the ones of Raginsky et al. (2017) which deal with the behavior of $(\mathbb{E}[U(X_k) - \min_{\mathbb{R}^d} U])_{k\in\mathbb{N}}$. To the best of our knowledge, this is the first result establishing quantitative bounds on the distance between the iterates of SGLD and a limiting measure concentrated on the minimizers of the target potential U.

3.2.2 Quantitative convergence for SGLD

Setting and notation In this section we start by recalling the setting considered in Raginsky et al. (2017). We assume that there exist a topological space $(\mathsf{Z}, \mathcal{B}(\mathsf{Z}))$, a probability measure $\mu \in \mathscr{P}(\mathsf{Z}, \mathcal{B}(\mathsf{Z}))$ and $u : \mathbb{R}^d \times \mathsf{Z} \to \mathbb{R}_+$ such that for any $z \in \mathsf{Z}$, $u(\cdot, z) \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}_+)$, and define for any $x \in \mathbb{R}^d$

$$U(x) = \textstyle\int_{\mathsf{Z}} u(x,z)\mathrm{d}\mu(z) \; . \tag{12}$$

We also denote by $U^\star$ the global minimum of U. We do not have access to μ and U directly but instead consider an empirical version of the target. Let $n \in \mathbb{N}$. We define $U_n : \mathbb{R}^d \times \mathsf{Z}^n \to \mathbb{R}_+$ given for any $z^{1:n} = \{z_i\}_{i=1}^n \in \mathsf{Z}^n$ and $x \in \mathbb{R}^d$ by

$$U_n(x, z^{1:n}) = (1/n) \textstyle\sum_{i=1}^n u(x, z_i) \; .$$

Let $(\mathsf{Y}, \mathcal{Y})$ be a measurable space, $\mathrm{R} : \mathsf{Z}^n \times \mathcal{Y} \to [0,1]$ a Markov kernel and $g : \mathbb{R}^d \times \mathsf{Y} \to \mathbb{R}^d$ such that for any $x \in \mathbb{R}^d$ and $z^{1:n} \in \mathsf{Z}^n$ we have

$$\nabla_x U_n(x, z^{1:n}) = \textstyle\int_{\mathsf{Y}} g(x,y)\mathrm{R}(z^{1:n}, \mathrm{d}y) \; . \tag{13}$$

Let Z be a random variable on Z^n with distribution $\mu^{\otimes n}$. Let $\{Y_k\}_{k\in\mathbb{N}}$ be a family of independent random variables on Y such that conditionally to Z we have for any $k \in \mathbb{N}$ that Y_k has distribution $\updelta_Z \mathrm{R}$. Finally, let $\varepsilon > 0$. We consider the Stochastic Gradient Langevin Dynamics (SGLD) sequence $(X_k)_{k\in\mathbb{N}}$ given by the following recursion: $X_0 \in \mathbb{R}^d$ and for any $k \in \mathbb{N}$

$$X_{k+1} = X_k - \gamma g(X_k, Y_k) + \sqrt{2\gamma\varepsilon} G_{k+1} \; , \tag{14}$$

where $(G_k)_{k\in\mathbb{N}}$ is a sequence of independent Gaussian random variables with zero mean and identity covariance matrix and $\gamma > 0$ is a stepsize. We also assume that $\{Y_k\}_{k\in\mathbb{N}}$ and $\{G_k\}_{k\in\mathbb{N}}$ are independent. For any $k \in \mathbb{N}$, we denote by $\mathrm{Q}_k : \mathsf{Z}^n \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$, the Markov kernel such that for any random variable Z on Z^n with distribution $\mu^{\otimes n}$, we have that X_k has distribution $\updelta_Z \mathrm{Q}_k$ conditionally to Z. Finally, we also define for any $\varepsilon > 0$ the Markov kernel $\mathrm{S}_\varepsilon : \mathsf{Z}^n \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$ associated with the Gibbs measure with temperature ε such that for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ and $z^{1:n} \in \mathsf{Z}^n$

$$\updelta_{z^{1:n}} \mathrm{S}_\varepsilon(\mathsf{A}) = \textstyle\int_{\mathsf{A}} \exp[-U_n(x, z^{1:n})/\varepsilon]\mathrm{d}x / \textstyle\int_{\mathbb{R}^d} \exp[-U_n(x, z^{1:n})/\varepsilon]\mathrm{d}x \; . \tag{15}$$

Notation	Source	Target	Description
μ	$\cdot$	Z	Data distribution, see (12).
R	Z^n	Y	Markov kernel defining the stochastic gradient conditionally to the data, see (13).
Q_k	Z^n	$\mathbb{R}^d$	Markov kernel associated with SGLD at step k conditionally to the data, see (14).
S_ε	Z^n	$\mathbb{R}^d$	Gibbs measure with temperature $\varepsilon > 0$ (conditional to the data) defined by (15).
S_0	Z^n	$\mathbb{R}^d$	Limiting measure at temperature zero (conditional to the data) defined by (16).

Table 1: Summary of Markov kernels and probability measures used in this section.

Similarly at temperature 0, we define the Markov kernel $\mathrm{S}_0 : \mathsf{Z}^n \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$ such that for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ and $z^{1:n} \in \mathsf{Z}^n$

$$\begin{aligned}\updelta_{z^{1:n}}\mathrm{S}_0(\mathsf{A}) = & \textstyle\int_{\mathsf{C}(z^{1:n})\cap\mathsf{A}} \det(\nabla_x^2 U_n(x, z^{1:n}))^{-1/2} \mathrm{d}\mathcal{H}^0(x) \\ & \textstyle / \int_{\mathsf{C}(z^{1:n})} \det(\nabla_x^2 U_n(x, z^{1:n}))^{-1/2} \mathrm{d}\mathcal{H}^0(x) \ , \end{aligned} \tag{16}$$

with $\mathsf{C}(z^{1:n}) = \arg\min\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\}$. We summarize the different probability measures and Markov kernels used in this section in Table 1.

Assumptions We first consider the following assumption, which is similar to the one of Raginsky et al. (2017).

H3 **(n)** *For any $z \in \mathsf{Z}$, $u(\cdot, z) \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R})$ and the following hold:*

(a) There exist $A, B \geq 0$ such that for any $z \in \mathsf{Z}$, $|u(0,z)| \leq A$ and $\|\nabla_x u(0,z)\| \leq B$.

(b) There exists $\mathtt{M} \geq 0$ such that for any $x_1, x_2 \in \mathbb{R}^d$, $z \in \mathsf{Z}$, $\|\nabla_x^k u(x_1, z) - \nabla_x^k u(x_2, z)\| \leq \mathtt{M}\|x_1 - x_2\|$ for any $k \in \{1,2,3\}$.

(c) There exist $\mathtt{m} > 0$, $\mathtt{c} \geq 0$ such that for any $x \in \mathbb{R}^d$, $z \in \mathsf{Z}$, $\langle \nabla_x u(x,z), x\rangle \geq \mathtt{m}\|x\|^2 - \mathtt{c}$.

(d) There exists $\kappa \geq 0$ such that for any $x \in \mathbb{R}^d$ and $z^{1:n} \in \mathsf{Z}^n$ we have

$$\textstyle\int_{\mathsf{Z}^n} \int_{\mathsf{Y}} \|\nabla_x U_n(x, z^{1:n}) - g(x,y)\|^2 \mathrm{R}(z^{1:n}, \mathrm{d}y) \mathrm{d}\mu^{\otimes n}(z^{1:n}) \leq \kappa(1 + \|x\|^2) \ .$$

Similar to Raginsky et al. (2017) we define, for any $\varepsilon > 0$ the uniform spectral gap

$$\lambda^\star(\varepsilon) = \inf_{z^{1:n} \in \mathsf{Z}^n, \ h \in \mathrm{C}^1(\mathbb{R}^d) \cap \mathrm{L}^2(\updelta_{z^{1:n}}\mathrm{S}_\varepsilon)} \{\updelta_{z^{1:n}}\mathrm{S}_\varepsilon[\|\nabla h\|^2] / \updelta_{z^{1:n}}\mathrm{S}_\varepsilon[|h|^2] \ : \ h \neq 0, \ \updelta_{z^{1:n}}\mathrm{S}_\varepsilon[h] = 0\} \ ,$$

and note that under **H**3(n), $\lambda^\star(\varepsilon) > 0$. We emphasize that **H**3(n) is satisfied in the case of a quadratic loss with a predictor given by a smooth and bounded neural network with bounded derivatives up to order 4 and a quadratic regularization. More precisely, we can consider $u(x,z) = \|f(x,z_1) - z_2\|^2 + \alpha\|x\|^2$ with $f \in \mathrm{C}^\infty(\mathbb{R}^d \times \mathsf{Z}_1, \mathsf{Z}_2)$ and $\mathsf{Z} = \mathsf{Z}_1 \times \mathsf{Z}_2$ with f and its derivatives bounded up to order 4 and $\alpha > 0$. In this setting f can be seen as a predictor of z_2 given z_1 and x the parameters of f. Finally, we also consider the following assumption which ensures that the limiting measures we consider are well-defined.

H4 **(**n**)** $u \in \mathrm{C}(\mathbb{R}^d \times \mathsf{Z}, \mathbb{R})$ *and the following hold:*

(a) Z *is compact.*

(b) For any $\beta > 0$, $\int_{\mathsf{Z}^n} \sigma^\star_\beta(z^{1:n}) \mathrm{d}\mu^{\otimes n}(z^{1:n}) < +\infty$*, where*

$$\sigma^\star_\beta(z^{1:n}) = \textstyle\sum_{x \in \mathsf{C}(z^{1:n})} \det(\nabla_x^2 U(x, z^{1:n}))^{-\beta} \ . \tag{17}$$

Note that **H**4 is satisfied if the number of minimizers is bounded w.r.t. $z \in \mathsf{Z}$ and if for any $z^{1:n} \in \mathsf{Z}^n$, $x^\star \in \mathsf{C}(z^{1:n})$, $\nabla_x^2 U(x^\star, z^{1:n}) \succeq \eta \,\mathrm{Id}$ with $\eta > 0$. Note that this condition is a slight strengthening of the condition that for any $z^{1:n} \in \mathsf{Z}^n$, $x^\star \in \mathsf{C}(z^{1:n})$, $\nabla_x^2 U(x^\star, z^{1:n}) \succ 0$. In particular, we impose that the landscape of $U(\cdot, z)$ is not too flat around the minimizers. We also introduce the *thermodynamic barrier* $c^\star : \ \mathsf{Z}^n \to \mathbb{R}$ such that for any $z^{1:n} \in \mathsf{Z}^n$, $c^\star(z^{1:n}) = +\infty$ if $U_n(\cdot, z^{1:n})$ does not admit a local minimizer which is not a global minimizer and

$$c^\star(z^{1:n}) = \inf\{U_n(x, z^{1:n}) \ : \ x \text{ local min. of } U_n(\cdot, z) \text{ but not global min.}\} - U_n^\star(z^{1:n}) \ ,$$

with $U_n^\star(z^{1:n}) = \inf\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\}$. We refer to Section 3.2.3 for a discussion on the *thermodynamic barrier* and its importance in non-convex optimization.

Main results We are now ready to state our main results. First, under **H**3(n) and **H** 4(n), we derive quantitative bounds for the sequence $(X_k)_{k \in \mathbb{N}}$.

Proposition 8 *Let* $n \in \mathbb{N}$. *Assume* **H**3(n) *and* **H**4(n). *Then there exist* $C \geq 0$, $\bar{\varepsilon}, \bar{\gamma}, \beta > 0$ *and* $k_0 \in \mathbb{N}$ *such that for any* $\varepsilon \in (0, \bar{\varepsilon}]$, $\gamma \in (0, \bar{\gamma}]$ *and* $k \in \mathbb{N}$ *with* $k \geq k_0$ *we have*

$$\begin{aligned} \mathbf{W}_1(\mu^{\otimes n}\mathrm{Q}_k, \mu^{\otimes n}\mathrm{S}_0) \leq C(1/\varepsilon + d)^2(\kappa^{1/4}\log(1/\gamma) + \gamma)/(\lambda^\star(\varepsilon)\varepsilon) \\ + C(1 + D_n)(\varepsilon^{1/2} + \varepsilon^{-d/2} \textstyle\int_{\mathsf{Z}^{1:n}} \exp[-c^\star(z^{1:n})/\varepsilon] \mathrm{d}\mu^{\otimes n}(z^{1:n})) \ , \end{aligned} \tag{18}$$

with $D_n = \int_{\mathsf{Z}^n} \sigma^\star_\beta(z^{1:n}) \mathrm{d}\mu^{\otimes n}(z^{1:n}) < +\infty$.

Proof We provide a sketch of the proof. The whole proof is postponed to Section 4.3.1. First, we assess the convergence of $(\mu^{\otimes n}\mathrm{Q}_k)_{k \in \mathbb{N}}$ by splitting the error in two parts. A first part is bounded using the geometric ergodicity of SGLD as in Raginsky et al. (2017) and controls the distance between $\mu^{\otimes n}\mathrm{Q}_k$ and $\mu^{\otimes n}\mathrm{S}_\varepsilon$. Then, using a parametric version of Theorem 1, see Theorem 33, we bound the distance between $\mu^{\otimes n}\mathrm{S}_\varepsilon$ and $\mu^{\otimes n}\mathrm{S}_0$. ■

A few remarks are in order:

(a) The condition that $\int_{\mathsf{Z}^n} \sigma^\star_1(z^{1:n}) \mathrm{d}\mu^{\otimes n}(z^{1:n}) < +\infty$ is necessary to ensure that $\mu^{\otimes n}\mathrm{S}_0$ is well-defined. In Theorem 8 we assume the condition that $\int_{\mathsf{Z}^n} \sigma^\star_\beta(z^{1:n}) \mathrm{d}\mu^{\otimes n}(z^{1:n}) < +\infty$ for any $\beta > 0$. In fact the condition could be relaxed to $\int_{\mathsf{Z}^n} \sigma^\star_\beta(z^{1:n}) \mathrm{d}\mu^{\otimes n}(z^{1:n}) < +\infty$ for any $\beta \in (0, \gamma]$ with $\gamma > 0$ an explicit constant. Obtaining such a result requires to derive explicit bounds in the quantitative Morse lemma (Le Loi and Phien, 2014, Theorem 3.2), see also Theorem 31. We leave this analysis to future works. Note also that this condition can be satisfied upon regularizing the function u and assuming that the number of global minimizers is bounded. Indeed, if we replace $u(x, z)$ by $u(x, z)+(\alpha/2)\|x\|^2$ (with $\alpha > 0$ some regularization parameter) then we get that for any $\beta > 0$, $\int_{\mathsf{Z}^n} \sigma^\star_\beta(z^{1:n})\mathrm{d}\mu^{\otimes n}(z^{1:n}) \leq N_0\alpha^{-\beta d}$ where N_0 is an upper-bound on the number of global minimizers.

(b) The upper-bound in (18) also depends on the *thermodynamic barrier*. This constant quantifies how close the local minima which are not global minima are to the global minima and is crucial to establish quantitative parametric Laplace-type results. We illustrate this situation in Section 3.2.3 with a simple example.

(c) Theorem 8 ensures that for a given precision level $r > 0$, there exist $\varepsilon, \gamma > 0$ small enough and $k \in \mathbb{N}$ large enough such that $\mathbf{W}_1(\mu^{\otimes n}\mathrm{Q}_k, \mu^{\otimes n}\mathrm{S}_0) \leq r$. However, note that $\mu^{\otimes n}\mathrm{S}_0$ is not concentrated on the minimizers of U. This highlights the fact that the minimization of the empirical risk does not guarantee that the population risk is small.

In order to verify that the population risk is small in expectation w.r.t. the target measure $\mu^{\otimes n}\mathrm{S}_0$ we use stability tools to establish the following proposition.

Proposition 9 *Assume that* $\mathbf{H}3(n)$ *and* $\mathbf{H}4(n)$ *hold uniformly w.r.t.* $n \in \mathbb{N}$. *Assume that*

$$\lim_{\varepsilon \to 0} \sup\{\varepsilon^{-d/2} \textstyle\int_{\mathsf{Z}^{1:n}} \exp[-c^\star(z^{1:n})/\varepsilon]\mathrm{d}\mu^{\otimes n}(z^{1:n}) \ : \ n \in \mathbb{N}\} = 0 \ ,$$

Then

$$\lim_{n \to +\infty} \{\mu^{\otimes n}\mathrm{S}_0[U] - U^\star\} = 0 \ .$$

In addition, assume that there exist $C_0, \alpha, \bar{\varepsilon} > 0$ *such that for any* $\varepsilon \in (0, \bar{\varepsilon}]$

$$\varepsilon^{-d/2} \textstyle\int_{\mathsf{Z}^{1:n}} \exp[-c^\star(z^{1:n})/\varepsilon]\mathrm{d}\mu^{\otimes n}(z^{1:n}) \leq C_0\varepsilon^\alpha \ , \tag{19}$$

Then for any $\eta \in (0, 1)$, *there exist* $C_\eta \geq 0$ *and* $n_0 \in \mathbb{N}$ *such that for any* $n \geq n_0$ *we have*

$$\mu^{\otimes n}\mathrm{S}_0[U] - U^\star \leq C_\eta / \log(n)^{s\eta} \ ,$$

with $s = \min(1/4, \alpha/2)$ *and where we recall that* $U^\star$ *is the global minimum of* U.

Proof The proof is postponed to Section 4.3.2. ■

Note that Theorem 9 implies that the sequence $(\mu^{\otimes n}\mathrm{S}_0[U])_{n\in\mathbb{N}}$ is tight since under $\mathbf{H}$ $3(n)$ we have that there exist $R \geq 0$ and $\alpha > 0$ such that for any $x \in \mathbb{R}^d$ with $\|x\| \geq R$ $U(x) \geq \mathtt{m}\|x\|^\alpha$. Furthermore, Theorem 9 implies that each limiting point of the sequence $(\mu^{\otimes n}\mathrm{S}_0[U])_{n\in\mathbb{N}}$ is concentrated on the minimizers of U when $n \to +\infty$, *i.e.* when the number of training points $\{z_i\}_{i=1}^n$ grows to infinity. However, we do not necessarily have that $(\mu^{\otimes n}\mathrm{S}_0[U])_{n\in\mathbb{N}}$ converges towards π_0 given for any $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ by

$$\pi_0(\mathsf{A}) = \textstyle\int_{\mathsf{A}\cap\mathsf{C}} \det(\nabla^2 U(x))^{-1/2}\mathrm{d}\mathcal{H}^0(x) / \int_{\mathsf{C}} \det(\nabla^2 U(x))^{-1/2}\mathrm{d}\mathcal{H}^0(x) \ ,$$

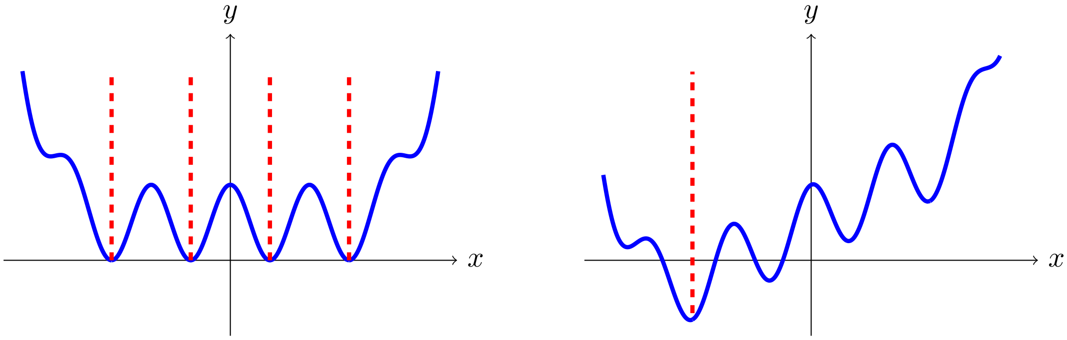

Figure 5: Left: the function $x \mapsto u(x,0)$ with u given in (20). Right: the function $x \mapsto u(x,0.5)$. The global minimizers are given by the dotted red lines.

where $\mathsf{C} = \arg\min\{U(x) \ : \ x \in \mathbb{R}^d\}$. Theorem 9 depends crucially on the *thermodynamic barrier*. In particular (19) allows us to derive quantitative stability results. The *thermodynamic barrier* and the associated conditions are discussed in Section 3.2.3.

In what follows, we describe a counter-example for which $(\mu^{\otimes n}\mathrm{S}_0)_{n\in\mathbb{N}}$ does not converge weakly towards π_0. Let $\mathsf{X} = \mathbb{R}$, $\mathsf{Z} = [-1/2, 1/2]$ and $u : \ \mathsf{X} \times \mathsf{Z} \to \mathbb{R}$ such that for any $x \in \mathbb{R}$ and $z \in [-1/2, 1/2]$ we have

$$u(x,z) = \begin{cases} (x+\pi)^4/(1+(x+\pi)^2) + \cos(3x) + zx \ , & \text{if } x < -\pi \ , \\ \cos(3x) + zx \ , & \text{if } -\pi \leq x \leq \pi \ , \\ (x-\pi)^4/(1+(x-\pi)^2) + \cos(3x) + zx \ , & \text{if } x > \pi \ . \end{cases} \tag{20}$$

We also let μ to be the uniform measure on $[-1/2, 1/2]$. We obtain that for any $n \in \mathbb{N}$, $z^{1:n} \in \mathsf{Z}^n$ and $x \in \mathsf{X}$

$$U(x) = u(x,0) \ , \qquad U_n(x, z^{1:n}) = u(x, (1/n)\textstyle\sum_{k=1}^n z^k) \ .$$

Note that for any $z \neq 0$, $x \mapsto u(x,z)$ admits a unique global minimizer, see the proof of Theorem 10, whereas if $z = 0$, $x \mapsto u(x,z)$ admits four global minimizers $\{x_i^\star\}_{i=1}^4 = \{-\pi, -\pi/3, \pi/3, \pi\}$, see Figure 5 for an illustration. In the next proposition, we show that $\lim_{n\to+\infty} \mu^{\otimes n}\mathrm{S}_0[U] = U^\star$ but that $(\mu^{\otimes n}\mathrm{S}_0)_{n\in\mathbb{N}}$ does *not* converge weakly towards π_0.

Proposition 10 *Let u be given by (20) and μ be the uniform measure on $[-1/2, 1/2]$. Then, we have that for any $n \in \mathbb{N}$*

$$\lim_{n\to+\infty} \mu^{\otimes n}\mathrm{S}_0 = (\delta_{-\pi} + \delta_{\pi})/2 \quad \textit{and} \quad \mu^{\otimes n}\mathrm{S}_0[U] - U^\star \leq (\pi/(6\sqrt{3}))n^{-1/2} \ .$$

Proof The proof is postponed to Section 4.3.3. ■

In particular, we have that $(\mu^{\otimes n}\mathrm{S}_0)_{n\in\mathbb{N}}$ converges towards a limiting probability measure supported on the set of minimizers of U.

3.2.3 The importance of the *thermodynamic barrier*

To conclude this section, we investigate the role of the *thermodynamic barrier* in order to establish quantitative parametric Laplace-type results. This quantity should not be confused with the concept of *kinetic barrier* which has been investigated in the context of simulated annealing, see Hajek (1988, 1985) for instance. We refer to Wang et al. (2016) for an introduction to the concept of *thermodynamic barrier* and *kinetic barrier* in the context of chemistry, see also Figure 6 for an illustration.

In a general setting, we consider a function $f : \mathsf{X} \times \mathsf{Z} \to \mathbb{R}$ where X and Z are topological spaces and $f(\cdot, z)$ admits a global minimizer for any $z \in \mathsf{Z}$. Let $z \in \mathsf{Z}$, if $f(\cdot, z)$ admits a local minimizer which is not a global minimizer we recall that the *thermodynamic barrier* $c^\star(z)$ is given by

$$c^\star(z) = \inf\{f(x,z) \, : \, x \text{ is a local minimizer of } f(\cdot,z) \text{ but not a global minimizer}\} - f^\star(z) \; ,$$

with $f^\star(z) = \inf\{f(x,z) \, : \, x \in \mathsf{X}\}$. The *thermodynamic barrier* quantifies how close the values of the local minima are to the global ones. Let $\mathsf{X} = \mathbb{R}^d$ and for any $\varepsilon > 0$, $z \in \mathsf{Z}$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$, define

$$\begin{aligned} &\delta_z \mathrm{S}_\varepsilon(\mathsf{A}) = \textstyle\int_{\mathsf{A}} \exp[-f(x,z)/\varepsilon] \mathrm{d}x / \textstyle\int_{\mathbb{R}^d} \exp[-f(x,z)/\varepsilon] \mathrm{d}x \; , \\ &\delta_z \mathrm{S}_0(\mathsf{A}) = \textstyle\int_{\mathsf{A} \cap \mathsf{C}(z)} \det(\nabla_x^2 f(x,z))^{-1/2} \mathrm{d}\mathcal{H}^0(x) / \textstyle\int_{\mathsf{C}(z)} \det(\nabla_x^2 f(x,z))^{-1/2} \mathrm{d}\mathcal{H}^0(x), \end{aligned}$$

with $\mathsf{C}(z) = \arg\min\{f(x,z) \, : \, x \in \mathbb{R}^d\}$. In Theorem 33 we show (under assumptions on f) that for any $z \in \mathsf{Z}$ and for any $\varphi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R})$ which satisfies the conditions of Theorem 33 there exist $A, \beta \geq 0$ and $\bar{\varepsilon} > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$

$$|\delta_z \mathrm{S}_\varepsilon[\varphi] - \delta_z \mathrm{S}_0[\varphi]| \leq A(1 + \sigma_\beta^\star(z))\{\varepsilon^{1/2} + \varepsilon^{-d/2} \exp[-c^\star(z)/\varepsilon]\} \; , \tag{21}$$

with A, β that do not depend on z. The dependency of the right-hand side w.r.t. $\sigma_\beta^\star(z)$ comes from the fact that $\delta_z \mathrm{S}_0$ is well-defined if and only if $\nabla_x^2 f(x,z)$ is invertible at the global minimizers of $f(\cdot, z)$.

We now investigate the dependency w.r.t. the *thermodynamic barrier* $c^\star$. We are going to build a simple example for which the *thermodynamic barrier* plays a crucial role. In particular, we will show that the dependency of the form $\exp[-c^\star(z)/\varepsilon]$ is tight in (21). Let $k \in \mathbb{N}$, $\mathsf{X} = \{0, 1\}$, $\mathsf{Z} = \mathbb{R}$ and $f : \mathsf{X} \times \mathsf{Z} \to \mathbb{R}$ such that for any $x \in \mathsf{X}$ and $z \in \mathsf{Z}$, $f(x,z) = x z^{2k+1}$. For any $\varepsilon > 0$ and $z \in \mathbb{R}$ we define $\delta_z \mathrm{S}_\varepsilon$ by

$$\begin{aligned} \delta_z \mathrm{S}_\varepsilon &= (\delta_0 \exp[-f(0,z)/\varepsilon] + \delta_1 \exp[-f(1,z)/\varepsilon])/(1 + \exp[-f(1,z)/\varepsilon]) \\ &= \delta_0 \mathrm{sigm}(z^{2k+1}/\varepsilon) + \delta_1 \mathrm{sigm}(-z^{2k+1}/\varepsilon) \; , \end{aligned}$$

where $\mathrm{sigm} : \mathbb{R} \to \mathbb{R}$ is the sigmoid function given for any $t \in \mathbb{R}$ by $\mathrm{sigm}(t) = (1 + \exp[-t])^{-1}$. When $z > 0$ the minimum of $f(\cdot, z)$ is 0 and is attained at $x = 0$. When $z = 0$, we have $f = 0$ (and the minimum is therefore attained at $x = 0$ and $x = 1$). When $z < 0$ the minimum of $f(\cdot, z)$ is $-z^{2k+1}$ and is attained at $x = 1$. Therefore we have that $\delta_z \mathrm{S}_0 = \delta_0$ if $z > 0$, $\delta_z \mathrm{S}_0 = \delta_1$ if $z < 0$ and $\delta_z \mathrm{S}_0 = (\delta_0 + \delta_1)/2$ if $z = 0$. Using that the Wasserstein distance of order 1 between two Bernoulli distributions with parameter p_1 and p_2 is given by $|p_1 - p_2|$ we get that for any $z \in \mathsf{Z}$ and $\varepsilon > 0$

$$\mathbf{W}_1(\delta_z \mathrm{S}_\varepsilon, \delta_z \mathrm{S}_0) = \mathrm{sigm}(-\left|z\right|^{2k+1}/\varepsilon) \; .$$

Hence, for a *fixed* value of $z \in \mathsf{Z}$, we get that $\mathbf{W}_1(\delta_z \mathrm{S}_\varepsilon, \delta_z \mathrm{S}_0)$ is of order $\mathcal{O}(\exp[-|z|^{2k+1}/\varepsilon])$. In particular, we get that for any $z \in \mathsf{Z}$, $\limsup_{\varepsilon \to 0} \mathbf{W}_1(\delta_z \mathrm{S}_\varepsilon, \delta_z \mathrm{S}_0)/\varepsilon$ is bounded. In what follows, we show that $\liminf_{\varepsilon \to 0} \mathbf{W}_1(\mu \mathrm{S}_\varepsilon, \mu \mathrm{S}_0)/\varepsilon = +\infty$, for some probability measure $\mu \in \mathscr{P}(\mathbb{R})$.

Let $\chi : \mathsf{X} \to \mathbb{R}$ with $\chi(0) = 1$ and $\chi(1) = 0$. First, note that for any $z > 0$ we have

$$\delta_z \mathrm{S}_\varepsilon[\chi] - \delta_z \mathrm{S}_0[\chi] = \mathrm{sigm}(-z^{2k+1}/\varepsilon) \ .$$

Hence, using that $c^\star(z) = |z|^{2k+1}$ for any $z \in \mathbb{R}$ and that $\mathrm{sigm}(-t) \geq \exp[-2t]$ for any $t > 0$ we have for any $z > 0$,

$$\delta_z \mathrm{S}_\varepsilon[\chi] - \delta_z \mathrm{S}_0[\chi] \geq \exp[-2z^{2k+1}/\varepsilon] \geq \exp[-2c^\star(z)/\varepsilon] \ . \tag{22}$$

Let $\mu \in \mathscr{P}(\mathbb{R}_+)$ such that $\mu(\{0\}) = 0$. Using (22) and that χ is 1-Lipschitz, we have

$$\mathbf{W}_1(\mu \mathrm{S}_\varepsilon, \mu \mathrm{S}_0) \geq \int_0^{+\infty} \exp[-2c^\star(z)/\varepsilon] \mathrm{d}\mu(z) \ .$$

Assume that μ is the uniform distribution on $[0,1]$. Then, we have for any $\varepsilon \in (0,1)$

$$\mathbf{W}_1(\mu \mathrm{S}_\varepsilon, \mu \mathrm{S}_0) \geq \int_0^1 \exp[-2z^{2k+1}/\varepsilon] \mathrm{d}x \geq 2^{-1/(2k+1)} (\int_0^1 \exp[-z^{2k+1}] \mathrm{d}z) \varepsilon^{1/(2k+1)} \ .$$

This shows that the order of $\mathbf{W}_1(\mu \mathrm{S}_\varepsilon, \mu \mathrm{S}_0)$ is at most $\mathcal{O}(\varepsilon^{1/(2k+1)})$. This is in stark contrast with the order identified for $\mathbf{W}_1(\delta_z \mathrm{S}_\varepsilon, \delta_z \mathrm{S}_0)$.

This latter observation highlights the crucial role of the *thermodynamic barrier* when establishing uniform Laplace-type results w.r.t. some parameter $z \in \mathsf{Z}$. Note that we recover that $\mathbf{W}_1(\mu \mathrm{S}_\varepsilon, \mu \mathrm{S}_0)$ is of order at least $\mathcal{O}(\varepsilon)$ if μ is supported on $[z_0, +\infty)$ with $z_0 > 0$. Similar conclusions hold if we show that $c^\star(z) \geq c_0$ for any $z \in \mathbb{R}$ with $c_0 > 0$. Hence, (assuming that $c^\star$ is continuous) the discrepancy between the order of $\mathbf{W}_1(\mu \mathrm{S}_\varepsilon, \mu \mathrm{S}_0)$ and the one of $\mathbf{W}_1(\delta_z \mathrm{S}_\varepsilon, \delta_z \mathrm{S}_0)$ might arise if: (a) At least one of the local minima (which is not a global minimum) converges towards a global minimum when $z \to z^\star$ for some value of $z^\star \in \mathsf{Z}$, (b) $z^\star$ belongs to the support of μ. If these two conditions are fulfilled then a more careful study of $\int_{\mathsf{Z}} \exp[-c^\star(z)/\varepsilon] \mathrm{d}\mu(z)$ is needed in order to obtain quantitative bounds.

4. Proofs

In this section, we gather the proofs of the previous sections. In Section 4.1 we prove Theorem 3. Then, in Section 4.2 we provide the proofs of the results of Section 3.1. Finally, the proofs of the results of Section 3.2 are given in Section 4.3.

4.1 Proof of Theorem 3

In this section, we prove Theorem 3. We recall that Theorem 1 is a straightforward consequence of Theorem 3 upon letting $\Psi = 1$. We let $k \in \mathbb{N}^\star$, $F : \mathbb{R}^d \to \mathbb{R}^p$ and $\Psi : \mathbb{R}^d \to \mathbb{R}_+$. For any $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ and $\varepsilon > 0$ we define

$$\begin{aligned} &\mathcal{I}_\varepsilon(\varphi) = C_\varepsilon^{-1} \int_{\mathbb{R}^d} \varphi(x) \Psi(x) \exp[-\|F(x)\|^k/\varepsilon] \mathrm{d}x \ , \quad \mathcal{J}_\varepsilon = \mathcal{I}_\varepsilon(1) \ , \\ &C_\varepsilon = \int_{\mathbb{R}^d} \exp[-\|x\|^k/\varepsilon] \mathrm{d}x = \varepsilon^{d/k} \int_{\mathbb{R}^d} \exp[-\|x\|^k] \mathrm{d}x = \varepsilon^{d/k} C_1 \ . \end{aligned} \tag{23}$$

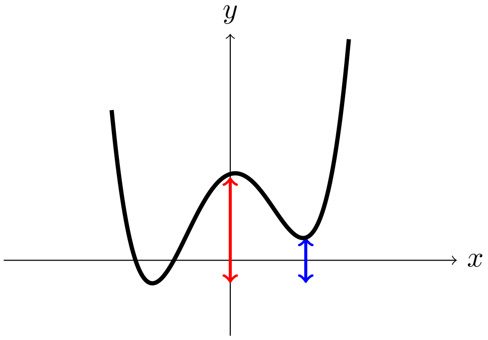

Figure 6: Difference between the *thermodynamic barrier* (blue) and the *kinetic barrier* (red).

In addition, we define

$$\mathcal{I}_0(\varphi) = \int_{F^{-1}(0)} \varphi(x)\Psi(x)\mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-\hat{d}}(x) \ , \quad \mathcal{J}_0 = \mathcal{I}_0(1) \ , \tag{24}$$

where $\hat{d} = \min(d, p)$ and we recall that for any $x \in \mathbb{R}^d$, $\mathrm{J}F(x) = \det(\mathrm{D}F(x)\mathrm{D}F(x)^\top)^{1/2}$ if $d \geq p$ and $\mathrm{J}F(x) = \det(\mathrm{D}F(x)^\top \mathrm{D}F(x))^{1/2}$ otherwise. If $\varphi : \mathbb{R}^d \to \mathbb{R}$ and $I_\varepsilon(|\varphi|) < +\infty$ for some $\varepsilon \geq 0$ we define $\mathcal{I}_\varepsilon(\varphi)$ similarly as in (23) and (24). Note that for any $\varepsilon \geq 0$ and $\varphi : \mathbb{R}^d \to \mathbb{R}^p$ such that it is defined we have $\pi_\varepsilon[\varphi] = \mathcal{I}_\varepsilon(\varphi)/\mathcal{J}_\varepsilon$.

The rest of this section is organized as follows. In Section 4.1.1, we prove our main result , *i.e.* a quantitative Laplace-type result in the case $d \geq p$ using the coarea formula. In Section 4.1.2, we prove similar results in the case $d \leq p$ using Laplace's method and Morse theory. Finally, we conclude with the proof of Theorem 3 in Section 4.1.3. Additional technical results are postponed to Appendix A.

4.1.1 The case $d \geq p$

In what follows, we assume that $d \geq p$ and for any $\varphi : \mathbb{R}^d \to \mathbb{R}_+$, $t \in \mathbb{R}^p$ we define $\mathcal{L}_t(\varphi)$ by

$$\mathcal{L}_t(\varphi) = \int_{F^{-1}(t)} \varphi(x)\Psi(x)\mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-p}(x) \ . \tag{25}$$

Note that $\mathcal{L}_0(\varphi) = \mathcal{I}_0(\varphi)$. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$. Then, if $\mathcal{L}_t(|\varphi|) < +\infty$, we define $\mathcal{L}_t(\varphi)$ similarly to (25). The following proposition establishes that $t \mapsto \mathcal{L}_t(\varphi)$ is Lipschitz under mild regularity conditions. We emphasize that this proposition is no longer true if $d \leq p$. Indeed, let us consider the following counterexample. Let $F : \mathbb{R} \to \mathbb{R}^2$ given for any $x \in \mathbb{R}$ by $F(x) = ((1-x^2)/(1+x^2), x(1-x^2)/(1+x^2))$. The set $F(\mathbb{R})$ defines a right strophoid. Then, for any $t \neq 0$ we have $\mathcal{H}^0(F^{-1}(t)) = 1$ or 0, but $\mathcal{H}^0(F^{-1}(0)) = 2$ and therefore $\mathcal{L}_t(\mathrm{J}F/\Psi) = 1$ or 0 near $t = 0$ but $\mathcal{L}_0(\mathrm{J}F/\Psi) = 2$ with $\Psi = 1$. Hence $t \mapsto \mathcal{L}_t(\mathrm{J}F/\Psi)$ is not even continuous.

Proposition 11 *Assume* **H**1, **H**2 *and that* $d \geq p$. *Let* $\mathsf{U} \subset \mathbb{R}^d$ *be open and such that* $F^{-1}(0) \subset \mathsf{U}$, *and let* $\varphi \in \mathrm{C}(\bar{\mathsf{U}}, \mathbb{R})$. *Then* $\lim_{t\to 0} \mathcal{L}_t(\varphi) = \mathcal{L}_0(\varphi)$. *In addition, assume that*

$\Psi, \varphi \in \mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$. *Then there exist* $B_0 \geq 0$ *and* $\eta > 0$ *such that for any* $t \in \bar{\mathrm{B}}(0, \eta)$,

$$|\mathcal{L}_t(\varphi) - \mathcal{L}_0(\varphi)| \leq B_0(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi}) \left\| t \right\| \ ,$$

with for any $i \in \{0, 1\}$ *and* $f \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$, $M_{i,f} = \sup\{\|\nabla^i f(x)\| \ : \ x \in F^{-1}(\bar{\mathrm{B}}(0, \eta))\}$, B_0 *and* η *do not depend on* φ *and* Ψ, *and* $F^{-1}(\bar{\mathrm{B}}(0, \eta)) \subset \mathsf{U}$.

Proof First, we show that there exists an explicit diffeomorphism between $F^{-1}(t)$ and $F^{-1}(0)$ for $\|t\|$ small enough. Then we use the coarea formula to express $\mathcal{L}_t$ as an integral over $F^{-1}(0)$ and use the dominated convergence theorem to conclude the first part of the proof. For the second part of the proof we differentiate the diffeomorphism w.r.t. the parameter t and provide explicit bounds for the derivative.

(a) The set $F^{-1}(0)$ is compact since $\lim_{\|x\| \to +\infty} \|F(x)\| = +\infty$. First, there exists $\eta_0 > 0$ such that $F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta_0)) \subset \mathsf{U}$. Indeed, since $F^{-1}(0)$ is compact, there exists $\varepsilon_0 > 0$ such that $F^{-1}(0) + \bar{\mathrm{B}}_\infty(0, \varepsilon_0) \subset \mathsf{U}$. We now show that for any $\varepsilon > 0$, there exists $\eta_\varepsilon > 0$ such that $F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta_\varepsilon)) \subset F^{-1}(0) + \bar{\mathrm{B}}_\infty(0, \varepsilon)$. If this is false, we let $\varepsilon > 0$ such that for any $\eta > 0$, $F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta)) \not\subset F^{-1}(0) + \bar{\mathrm{B}}_\infty(0, \varepsilon)$. Hence there exists a sequence $(x_k)_{k \in \mathbb{N}} \in (\mathbb{R}^d)^{\mathbb{N}}$ such that $\|F(x_k)\|_\infty \leq 1/(k+1)$ and $d(x_k, F^{-1}(0)) \geq \varepsilon$. But, up to taking a subsequence, there exists $x^\star \in F^{-1}(\bar{\mathrm{B}}_\infty(0, 1))$ such that $\lim_{k \to +\infty} x_k = x^\star$. Then, we have $F(x^\star) = 0$ and $d(x^\star, F^{-1}(0)) > \varepsilon$, which is absurd. Hence for any $\varepsilon > 0$, there exists $\eta_\varepsilon > 0$ such that $F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta_\varepsilon)) \subset F^{-1}(0) + \bar{\mathrm{B}}_\infty(0, \varepsilon)$. We let $\eta_0 = \eta_{\varepsilon_0}$.

Second, there exists $\eta_1 > 0$ such that for any $x \in F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta_1))$, $\mathrm{J}F(x) > 0$. Indeed, if this is not the case then there exists $(x_k)_{k \in \mathbb{N}}$ with $\lim_{k \to +\infty} F(x_k) = 0$ and $\mathrm{J}F(x_k) = 0$. Since $F^{-1}(\bar{\mathrm{B}}_\infty(0, 1))$ is compact there exists $x^\star$ such that, up to taking a subsequence, $\lim_{k \to +\infty} x_k = x^\star$. Then $F(x^\star) = 0$ and $\mathrm{J}F(x^\star) = 0$, which is absurd. We define $\mathsf{K}_0 = F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta_1))$ and $\mathsf{K}_1 = F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta))$ with $\eta = \min(\eta_1/2, \eta_0)$. Note that $\mathsf{K}_1 \subset \mathsf{U}$ and $\mathsf{K}_1 \subset \mathrm{int}(\mathsf{K}_0)$.

In what follows for any $x \in \mathbb{R}^d$ we define $G(x) = \mathrm{D}F(x)\mathrm{D}F(x)^\top$. Note that for any $x \in \mathbb{R}^d$, $\det(G(x))^{1/2} = \mathrm{J}F(x)$ since $d \geq p$. We also have that for any $x \in \mathsf{K}_0$, $G(x)$ is invertible since $\mathrm{J}F(x) > 0$. In addition, we have that for any $i, j \in \{1, \dots, p\}$ and $x \in \mathbb{R}^d$

$$G_{i,j}(x) = \langle \nabla F_i(x), \nabla F_j(x) \rangle \ .$$

We define $\{f_i\}_{i=1}^p$ such that for any $i \in \{1, \dots, p\}$, $f_i : \mathbb{R}^d \to \mathbb{R}^d$ is given for any $x \in \mathsf{K}_0$ by

$$f_i(x) = \textstyle\sum_{k=1}^p h_{i,k}(x) \nabla F_k(x) \ , \tag{26}$$

with $\{h_{i,j}(x)\}_{1 \leq i,j \leq p} = G(x)^{-1}$. For any $x \in \mathsf{K}_0$ and $i, j \in \{1, \dots, p\}$ we have

$$\langle f_i(x), \nabla F_j(x) \rangle = \textstyle\sum_{k=1}^p h_{i,k}(x) \langle \nabla F_k(x), \nabla F_j(x) \rangle = \updelta_i(j) \ ,$$

where $\updelta_i$ is the Dirac mass at i. In what follows, we let $\{g_i\}_{i=1}^p$ such that for any $i \in \{1, \dots, p\}$, $g_i : \mathbb{R}^d \to \mathbb{R}^d$ and $g_i \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ such that $g_i(x) = f_i(x)$ for any $x \in \mathsf{K}_1$, and $g_i(x) = 0$ for $x \in \mathrm{int}(\mathsf{K}_0)^\mathrm{c}$, such functions exist using Whitney extension theorem for instance, see Whitney (1934). In what follows, we fix $t = (t_1, \dots, t_p) \in \bar{\mathrm{B}}_\infty(0, \eta)$. For any

$i \in \{1, \dots, p\}$ let $\Phi_i : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ given by $\Phi_i(0, x) = x$ for any $x \in \mathbb{R}^d$ and for any $s \in \mathbb{R}$ and $x \in \mathbb{R}^d$

$$\partial_s \Phi_i(s, x) = -g_i(\Phi_i(s, x)) \ . \tag{27}$$

For any $i \in \{1, \dots, p\}$, Φ_i is well-defined using Theorem 36. Therefore, we have for any $i \in \{1, \dots, p\}$ and $s \in \mathbb{R}$, $x \in \mathbb{R}^d$ such that $\Phi_i(s, x) \in \mathsf{K}_1$

$$\partial_s F(\Phi_i(s, x)) = -(\langle g_i(\Phi_i(s, x)), \nabla F_1(\Phi_i(s, x))\rangle, \dots, \langle g_i(\Phi_i(s, x)), \nabla F_p(\Phi_i(s, x))\rangle) = -e_i \ ,$$

where we recall that $\{e_i\}_{i=1}^p$ is the canonical basis of $\mathbb{R}^p$. We define $\bar{\Phi}_t : \mathbb{R}^d \to \mathbb{R}^d$ such that for any $x \in \mathbb{R}^d$, $\bar{\Phi}_t(x) = x^{(p)}$ with $x^{(0)} = x$ and for any $i \in \{0, \dots, p-1\}$, $x^{(i+1)} = \Phi_{i+1}(t_{i+1}, x^{(i)})$. Note that $\bar{\Phi}_t \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$ and is a diffeomorphism, see Theorem 36. Using (27) we have that $\bar{\Phi}_t(F^{-1}(t)) = F^{-1}(0)$. In addition, $F^{-1}(t)$ is $\mathcal{H}^{d-p}$ countably rectifiable using Theorem 40. Using this result and the coarea formula, see Theorem 41, we have

$$\mathcal{L}_t(\varphi) = \int_{F^{-1}(0)} \varphi(\bar{\Phi}_t^{-1}(x))\Psi(\bar{\Phi}_t^{-1}(x))\mathrm{J}F(\bar{\Phi}_t^{-1}(x))^{-1}|\det(\mathrm{D}\bar{\Phi}_t^{-1}(x))|\mathrm{d}\mathcal{H}^{d-p}(x) \ .$$

Since $F^{-1}(0) \times \mathrm{B}_\infty(0, \eta)$ is compact and $(t, x) \mapsto \bar{\Phi}_t^{-1}(x)$ and $(t, x) \mapsto \mathrm{D}\bar{\Phi}_t^{-1}(x)$ are continuous with for any $x \in \mathbb{R}^d$, $\bar{\Phi}_0^{-1}(x) = x$ and $\mathrm{D}\bar{\Phi}_0^{-1}(x) = \mathrm{Id}$, we get that $\lim_{t\to 0} \mathcal{L}_t(\varphi) = \mathcal{L}_0(\varphi)$ using the dominated convergence theorem.

(b) For the second part of the proof we control the derivative of $t \mapsto \chi(t, x)$ where for any $x \in \mathbb{R}^d$ and $t \in \mathbb{R}^p$ we have

$$\chi(t, x) = \varphi(\bar{\Phi}_t^{-1}(x))\Psi(\bar{\Phi}_t^{-1})\mathrm{J}F(\bar{\Phi}_t^{-1}(x))^{-1}|\det(\mathrm{D}\bar{\Phi}_t^{-1}(x))| \ . \tag{28}$$

Using Theorem 27, there exists $P \in \mathtt{Poly}(4, \mathbb{R}_+)$ such that for $x \in F^{-1}(0)$ and $t \in \mathrm{B}_\infty(0, \eta)$ we have

$$\begin{aligned}\|\partial_t \chi(t, x)\| \leq &(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi}) \\ &\times P(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \exp[P(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F})] \ ,\end{aligned}$$

with P that does not depend on φ and Ψ. Hence we have that for any $t \in \mathrm{B}_\infty(0, \eta)$

$$\begin{aligned}|\mathcal{L}_t(\varphi) - \mathcal{L}_0(\varphi)| &\leq \int_{F^{-1}(0)} |\chi(t, x) - \chi(0, x)| \, \mathrm{d}\mathcal{H}^{d-p}(x) \\ &\leq (1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi}) P(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \\ &\qquad \exp[P(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F})]\mathcal{H}^{d-p}(F^{-1}(0)) \, \|t\| \ ,\end{aligned}$$

which concludes the proof since $\mathcal{H}^{d-p}(F^{-1}(0)) < +\infty$ by Theorem 40.

■

The smoothness of F can be relaxed to $\mathrm{C}^3(\mathbb{R}^d, \mathbb{R}^p)$ (at least). From our analysis, $F \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^p)$ (or at least $F \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^p)$ with Lipschitz derivative) seems to be necessary to obtain quantitative results.

Proposition 12 *Assume* **H**1, **H**2 *and that* $d \geq p$. *Let* $\mathsf{U} \subset \mathbb{R}^d$ *open and bounded such that* $F^{-1}(0) \subset \mathsf{U}$ *and* $\varphi \in \mathrm{C}(\bar{\mathsf{U}}, \mathbb{R})$. *Then there exists* $\eta > 0$ *such that* $\lim_{\varepsilon \to 0} \mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) = \mathcal{I}_0(\varphi)$, *where for any* $\varepsilon > 0$, $\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{F^{-1}(\mathrm{B}(0,\eta))})$. *In addition, assume that* $\varphi, \Psi \in \mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$, *then there exists* $A_2 \geq 0$ *such that for any* $\varepsilon > 0$

$$\left|\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0(\varphi)\right| \leq A_2(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} ,$$

with for any $i \in \{0, 1\}$ *and* $f \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$, $M_{i,f} = \sup\{\|\nabla^i f(x)\| \ : \ x \in F^{-1}(\bar{\mathrm{B}}(0,\eta))\}$, A_2 *and* η *do not depend on* φ *and* Ψ, *and* $F^{-1}(\bar{\mathrm{B}}(0,\eta)) \subset \mathsf{U}$.

Proof First, note that $\mathcal{I}_0(\varphi) = \mathcal{L}_0(\varphi)$, see (24) and (25). In what follows, we let $\eta > 0$ be given by Theorem 11 and define $\mathcal{I}_0^{\mathrm{in}}(\varphi) = C_\varepsilon^{-1} \int_{\mathrm{B}(0,\eta)} \exp[-\|t\|^k/\varepsilon] \mathcal{L}_0(\varphi) \mathrm{d}t$. We have

$$|\mathcal{I}_0(\varphi) - \mathcal{I}_0^{\mathrm{in}}(\varphi)| = |\mathcal{I}_0(\varphi)|\, C_\varepsilon^{-1} \int_{\mathrm{B}(0,\eta)^{\mathrm{c}}} \exp[-\|t\|^k/\varepsilon] \mathrm{d}t \leq 2^{d/k} |\mathcal{I}_0(\varphi)| \exp[-\eta^k/(2\varepsilon)] . \quad (29)$$

Using the coarea formula, see Theorem 41, we have for any $\varepsilon > 0$,

$$\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) = C_\varepsilon^{-1} \int_{F^{-1}(\mathrm{B}(0,\eta))} \Psi(x)\varphi(x) \exp[-\|F(x)\|^k/\varepsilon] \mathrm{d}x = C_\varepsilon^{-1} \int_{\mathrm{B}(0,\eta)} \exp[-\|t\|^k/\varepsilon] \mathcal{L}_t(\varphi) \mathrm{d}t .$$

Therefore, using this result and the change of variable $t \mapsto \varepsilon^{1/k} t$ we have

$$\begin{aligned} |\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0^{\mathrm{in}}(\varphi)| &\leq C_\varepsilon^{-1} \int_{\mathrm{B}(0,\eta)} \exp[-\|t\|^k/\varepsilon] |\mathcal{L}_t(\varphi) - \mathcal{L}_0(\varphi)| \mathrm{d}t \\ &\leq C_1^{-1} \int_{\mathrm{B}(0,\eta/\varepsilon^{1/k})} \exp[-\|t\|^k] |\mathcal{L}_{t\varepsilon^{1/k}}(\varphi) - \mathcal{L}_0(\varphi)| \mathrm{d}t . \end{aligned} \quad (30)$$

Hence, we get that $\lim_{\varepsilon \to 0} |\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0^{\mathrm{in}}(\varphi)| = 0$ using the dominated convergence theorem and that $\lim_{\varepsilon \to 0} |\mathcal{L}_{t\varepsilon^{1/k}}(\varphi) - \mathcal{L}_0(\varphi)| = 0$ according to Theorem 11. This concludes the first part of the proof upon combining this result and (29). In addition, assume that $\varphi \in \mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$ then using the second part of Theorem 11 and (30) we have

$$\left|\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0^{\mathrm{in}}(\varphi)\right| \leq B_0 C_1^{-1} \int_{\mathbb{R}^p} \|t\| \exp[-\|t\|^k] \mathrm{d}t (1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} ,$$

which concludes the proof upon combining this result and (29). ■

4.1.2 The case $d \leq p$

We now turn to the case $d \leq p$. The proof of this result is more classical and does not rely on geometric measure theory. Instead we build on the Morse theory approach for Laplace approximation, see Wong (2001) for example. The following proposition is a quantitative extension of (Wong, 2001, Theorem 3, p.495).

Proposition 13 *Assume* **H**1, **H**2 *and that* $d \leq p$. *Let* $\mathsf{U} \subset \mathbb{R}^d$ *open and bounded such that* $F^{-1}(0) \subset \mathsf{U}$ *and* $\varphi \in \mathrm{C}(\bar{\mathsf{U}}, \mathbb{R})$. *Then* $\lim_{\varepsilon \to 0} |\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0(\varphi)| = 0$, *with* $\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\mathsf{V}})$ *and* V *open such that* $F^{-1}(0) \subset \mathsf{V} \subset \mathsf{U}$. *In addition, assume that* $\varphi, \Psi \in \mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$. *Then there exists* $B_1 \geq 0$ *such that for any* $\varepsilon > 0$ *we have*

$$|\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0(\varphi)| \leq B_1(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} ,$$

with B_1 *that does not depend on* φ *and* Ψ, *and for any* $i \in \{0, 1\}$ *and* $f \in \mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$, $M_{i,f} = \sup\{\|\nabla^i f(x)\| \ : \ x \in \mathsf{U}\}$.

Proof Let $\{x_0^\ell\}_{\ell=1}^N$ and $\{\mathsf{W}_\ell\}_{\ell=1}^N$ be given by Theorem 26 such that $F^{-1}(0) = \cup_{\ell=1}^N \{x_0^\ell\}$ and $\mathrm{d}F(x)$ is injective for any $\ell \in \{1, \dots, N\}$ and $x \in \mathsf{W}_\ell$. In addition, for any $\ell, m \in \{1, \dots, N\}$, $\mathsf{W}_\ell \cap \mathsf{W}_m = \emptyset$. Let $\ell \in \{1, \dots, N\}$ and $U : \mathbb{R}^d \to \mathbb{R}_+$ such that for any $x \in \mathbb{R}^d$, $U(x) = \|F(x)\|^2$. Since $F \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^p)$ we have that $U \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R})$. We divide the rest of the proof into two parts.

(a) First, we have that $\nabla^2 U(x_0^\ell) = 2\mathrm{D}F(x_0^\ell)^\top \mathrm{D}F(x_0^\ell)$ which is invertible since $\mathrm{J}F(x_0^\ell) > 0$. Therefore, we can apply Morse's lemma (Nirenberg, 2001, Theorem 3.1.1) and there exists a diffeormorphism $\Phi_\ell \in \mathrm{C}^1(\bar{\Omega}_\ell, \bar{\mathsf{W}}_\ell)$ with $0 \in \Omega_\ell$, $x_0^\ell \in \mathsf{W}_\ell$ and $\Omega_\ell \subset \mathbb{R}^d$ open such that for any $x \in \Omega_\ell$, $U(\Phi_\ell(x)) = \|x\|^2$, $\Phi_\ell(0) = x_0^\ell$ and $\mathrm{D}\Phi_\ell(0) = (\mathrm{D}F(x_0^\ell)^\top \mathrm{D}F(x_0^\ell))^{-1/2}$. Note that $\det(\mathrm{D}\Phi_\ell(0)) = \mathrm{J}F(x_0^\ell)^{-1}$.

Let $r_\ell > 0$ such that $\bar{\mathrm{B}}(0, r_\ell) \subset \Omega_\ell$. We have

$$1 - \int_{\Omega_\ell/\varepsilon^{1/k}} \exp[-\|x\|^k]\mathrm{d}x/C_1 \leq \int_{\mathbb{R}^d} \exp[-\|x\|^k/2]\mathrm{d}x \exp[-r_\ell^k/(2\varepsilon)]/C_1 \; . \tag{31}$$

In what follows, we no longer consider $\ell \in \{1, \dots, N\}$ to be fixed. Let $\mathsf{V} = \cup_{\ell=1}^N \Phi_\ell(\Omega_\ell)$ and $\mathcal{I}_{0,\varepsilon}^{\mathrm{in}}(\varphi) = \sum_{\ell=1}^N \int_{\Omega_\ell/\varepsilon^{1/k}} \exp[-\|x\|^k]\mathrm{d}x \varphi(x_0^\ell)\Psi(x_0^\ell)\mathrm{J}F(x_0^\ell)^{-1}/C_1$. We recall that we have

$$\mathcal{I}_0(\varphi) = \textstyle\sum_{\ell=1}^N \varphi(x_0^\ell)\Psi(x_0^\ell)\mathrm{J}F(x_0^\ell)^{-1} \; .$$

Combining this result and (31) we get

$$\left|\mathcal{I}_{0,\varepsilon}^{\mathrm{in}}(\varphi) - \mathcal{I}_0(\varphi)\right| \leq NM \int_{\mathbb{R}^d} \exp[-\|x\|^k/2]\mathrm{d}x \exp[-r_{\min}^k/(2\varepsilon)]/C_1 \; , \tag{32}$$

where $r_{\min} = \min\{r_\ell : \ell \in \{1, \dots, N\}\}$ and $M = \max\{|\varphi(x_0^\ell)|\Psi(x_0^\ell)\mathrm{J}F(x_0^\ell)^{-1} : \ell \in \{1, \dots, N\}\}$. Using for any $\ell \in \{1, \dots, N\}$ the change of variable $x \mapsto \Phi_\ell(x)$ and $x \mapsto \varepsilon^{1/k}x$ we have

$$\begin{aligned}\left|\mathcal{I}_{0,\varepsilon}^{\mathrm{in}}(\varphi) - \mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi)\right| \leq \textstyle\sum_{\ell=1}^N \int_{\Omega_\ell/\varepsilon^{1/k}} &|\varphi(\Phi_\ell(\varepsilon^{1/k}x))\Psi(\Phi_\ell(\varepsilon^{1/k}x)) \det(\mathrm{D}\Phi_\ell(\varepsilon^{1/k}x)) \\ &- \varphi(x_0^\ell)\Psi(x_0^\ell)\mathrm{J}F(x_0^\ell)^{-1}| \exp[-\|x\|^k]\mathrm{d}x/C_1 \; ,\end{aligned} \tag{33}$$

For any $\ell \in \{1, \dots, N\}$, let $\chi_\ell : \Omega_\ell \to \mathbb{R}$ given for any $x \in \Omega_\ell$ by

$$\chi_\ell(x) = \varphi(\Phi_\ell(x))\Psi(\Phi_\ell(x)) \det(\mathrm{D}\Phi_\ell(x)) \; .$$

We conclude the first part of the proof using (32), the dominated convergence theorem in (33) and that for any $\ell \in \{1, \dots, N\}$, $\chi \in \mathrm{C}(\Omega_\ell, \mathbb{R})$.

(b) For the second part of the proof, since $\varphi, \Psi \in \mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$ and for any $\ell \in \{1, \dots, N\}$, $\Phi_\ell \in \mathrm{C}^2(\bar{\Omega}_\ell, \bar{\mathsf{W}}_\ell)$ since $F \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^p)$ using (Nirenberg, 2001, Theorem 3.1.1), we have that for any $\ell \in \{1, \dots, N\}$, $\chi_\ell \in \mathrm{C}^1(\bar{\Omega}_\ell, \mathbb{R})$ and there exists $B_1' \geq 0$ (which do not depend on φ and Ψ) such that for any $\ell \in \{1, \dots, N\}$ and $x \in \Omega_\ell$ we have

$$\|\mathrm{D}\chi_\ell(x)\| \leq B_1'(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi}) \; .$$

Using this result and (33) we get that

$$|\mathcal{I}_{0,\varepsilon}^{\mathrm{in}}(\varphi) - \mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi)| \leq NB_1'(1 + M_{0,\varphi} + M_{1,\varphi})\varepsilon^{1/k} \int_{\mathbb{R}^d} \|x\| \exp[-\|x\|^k]\mathrm{d}x/C_1 \; .$$

Combining this result and (32) concludes the proof.

■

4.1.3 Proof of Theorem 3

We start by proving the results of Theorem 3 in a smooth setting, then we deduce the general case using a smoothing lemma.

Proposition 14 *Assume* **H**1 *and* **H**2*. Let* $\mathsf{U} \subset \mathbb{R}^d$ *open and bounded such that* $F^{-1}(0) \subset \mathsf{U}$ *and* $\varphi \in \mathrm{C}(\bar{\mathsf{U}}, \mathbb{R})$ *which satisfies* (7)*. Then,* $\lim_{\varepsilon \to 0} \left|\pi_\varepsilon^\Psi[\varphi] - \pi_0^\Psi[\varphi]\right| = 0$*. In addition, assume that* $\varphi, \Psi \in \mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$*. Then there exists* $A \in \mathrm{C}(\mathbb{R}_+^3, \mathbb{R}_+)$ *such that for any* $\varepsilon \in (0, \mathtt{m}^k/(4\{C_\varphi + C_\Psi + C_\varphi C_\Psi + 1\}))$

$$\left|\pi_\varepsilon^\Psi[\varphi] - \pi_0^\Psi[\varphi]\right| \leq A(C_\varphi, C_\Psi, m_{0,\Psi})(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} ,$$

with for any $i \in \{0,1\}$ *and* $f \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$*,* $M_{i,f} = \sup\{\|\nabla^i f(x)\| \ : \ x \in \mathsf{U}\}$*,* $m_{0,\Psi} = \inf\{\Psi(x) \ : \ x \in F^{-1}(0)\}$ *and* A *function that does not depend on* φ *and* Ψ*. Finally,* A *is non-decreasing w.r.t. its first two variables and non-increasing w.r.t. its last variable.*

Proof Let $\bar{\varepsilon} = \mathtt{m}^k/(4 + 4C_{\varphi,\Psi})$, with $C_{\varphi,\Psi} = (C_\Psi + 1)(C_\varphi + 1)$ and note that we have

$$\bar{\varepsilon} < \min(\mathtt{m}^k/(1 + C_{\varphi,\Psi}), \mathtt{m}^k/(1 + C_{\varphi,1})) . \tag{34}$$

For any $\varepsilon \in (0, \bar{\varepsilon})$ we have

$$\begin{aligned} \left|\pi_\varepsilon^\Psi[\varphi] - \pi_0^\Psi[\varphi]\right| &\leq |\mathcal{I}_\varepsilon(\varphi)/\mathcal{J}_\varepsilon - \mathcal{I}_0(\varphi)/\mathcal{J}_0| \\ &\leq \mathcal{J}_\varepsilon^{-1} \left|\mathcal{I}_\varepsilon(\varphi) - \mathcal{I}_0(\varphi)\right| + \mathcal{I}_0(\varphi)/(\mathcal{J}_0 \mathcal{J}_\varepsilon) \left|\mathcal{I}_\varepsilon(1) - \mathcal{I}_0(1)\right| . \end{aligned} \tag{35}$$

Let $\eta > 0$ be given by Theorem 11 and let $\mathsf{V} = F^{-1}(\mathrm{B}(0, \eta))$ if $d \geq p$, and V given by Theorem 13 otherwise. Note that $F^{-1}(0) \subset \mathsf{V}$. For any $\varepsilon \in (0, \bar{\varepsilon})$ we define $\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\mathsf{V}^c})$ and $\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\mathsf{V}})$. We divide the rest of the proof into two parts. First, we control $\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi)$ using the technical bounds of Theorem 25. Then, we control $|\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0(\varphi)|$ using either Theorem 12 if $d \geq p$ or Theorem 13 if $d \leq p$. We conclude upon combining these results.

(a) Using (34) and Theorem 25 we get that for any $\varepsilon \in (0, \bar{\varepsilon})$ we have

$$\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi) \leq A_1(C_{\varphi,\Psi})\varepsilon^{-d/k} \exp[-\beta_1/\varepsilon] , \qquad \mathcal{I}_\varepsilon^{\mathrm{out}}(1) \leq A_1(C_{1,\Psi})\varepsilon^{-d/k} \exp[-\beta_1/\varepsilon] , \tag{36}$$

with $\beta_1 > 0$ and $A_1 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ that do not depend on φ and Ψ, and are non-decreasing.

(b) Using either Theorem 12 if $d \geq p$ or Theorem 13 if $d \leq p$ we have that for any $\varepsilon \in (0, \bar{\varepsilon})$

$$\begin{aligned} \left|\mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi) - \mathcal{I}_0(\varphi)\right| &\leq A_2(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} , \\ \left|\mathcal{I}_\varepsilon^{\mathrm{in}}(1) - \mathcal{I}_0(1)\right| &\leq 2A_2(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} , \end{aligned} \tag{37}$$

with for any $i \in \{0,1\}$ and $f \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$, $M_{i,f} = \sup\{\|\nabla^i f(x)\| \ : \ x \in \mathsf{U}\}$, and A_2 that does not depend on φ and Ψ.

Combining (36) and (37) we get that for any $\varepsilon \in (0, \bar{\varepsilon})$

$$|\mathcal{I}_\varepsilon(\varphi) - \mathcal{I}_0(\varphi)| \leq A_2(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} + A_1(C_{\varphi,\Psi})\varepsilon^{-d/k}\exp[-\beta_1/\varepsilon] ,$$
$$|\mathcal{I}_\varepsilon(1) - \mathcal{I}_0(1)| \leq 2A_2(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} + A_1(C_{1,\Psi})\varepsilon^{-d/k}\exp[-\beta_1/\varepsilon] .$$

Since for any $\varepsilon > 0$, $\varepsilon^{-d/k}\exp[-\beta/\varepsilon] \leq ((d+1)/(k\beta))^{1/k+d/k}\varepsilon^{1/k}$ (this inequality comes from first multiplying by $\varepsilon^{-1/k}$ on both sides, and then taking the value of ε that achieves the maximum of the left-hand term), there exists $\tilde{A} \in \mathrm{C}(\mathbb{R}_+^2, \mathbb{R}_+)$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$ we have

$$|\mathcal{I}_\varepsilon(\varphi) - \mathcal{I}_0(\varphi)| \leq \tilde{A}(C_\varphi, C_\Psi)(1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} , \tag{38}$$
$$|\mathcal{I}_\varepsilon(1) - \mathcal{I}_0(1)| \leq 2\tilde{A}(1, C_\Psi)(1 + M_{0,\Psi} + M_{1,\Psi})\varepsilon^{1/k} ,$$

with $\tilde{A}$ that does not depend on Ψ and φ, and $\tilde{A}$ non-decreasing w.r.t. to each of its variables. Using Theorem 24, there exists $A_0 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+^\star)$ such that for any $\varepsilon \in [0, \bar{\varepsilon}]$, $\mathcal{J}_\varepsilon \geq A_0(m_{0,\Psi})$ with $m_{0,\Psi} = \inf\{\Psi(x) \ : \ x \in F^{-1}(0)\}$, and A_0 that does not depend on Ψ and is non-increasing. Finally, note that $\mathcal{I}_0(\varphi) \leq \sup\{\mathrm{J}F(x)^{-1} \ : \ x \in F^{-1}(0)\} M_{0,\varphi} M_{0,\Psi} \mathcal{H}^{d-p}(F^{-1}(0))$ with $\mathcal{H}^{d-p}(F^{-1}(0)) < +\infty$ by Theorem 40. Combining these results, (35) and (38) concludes the proof. ■

Using this proposition along with a smoothing lemma, see Theorem 28, we conclude the proof of Theorem 3.

Proof [Proof of Theorem 3] We begin by introducing the families $\{\varphi^\delta \ : \ \delta \in \left(0, \bar{\delta}\right)\}$ and $\{\Psi^\delta \ : \ \delta \in \left(0, \bar{\delta}\right)\}$ which are smooth approximations of φ and Ψ respectively. Since $F^{-1}(0)$ is compact and U^c is closed there exists $r > 0$ such that $F^{-1}(0) + \mathrm{B}(0, r) \subset \mathsf{U}$. Let $\mathsf{U}_0 = F^{-1}(0) + \mathrm{B}(0, r/2)$ and note that $F^{-1}(0) \subset \mathsf{U}_0$ and $\mathsf{U}_0 + \mathrm{B}(0, r/2) \subset \mathsf{U}$. Let $\{\varphi^\delta \ : \ \delta \in \left(0, \bar{\delta}\right)\}$ and $\{\Psi^\delta \ : \ \delta \in \left(0, \bar{\delta}\right)\}$ with $\bar{\delta} > 0$ given by Theorem 28, $\mathsf{V} \leftarrow \mathsf{U}$ and $\mathsf{U} \leftarrow \mathsf{U}_0$. Then using the dominated convergence theorem and the fact that $F^{-1}(0)$ is compact we have that $\lim_{\delta \to 0} \pi_0^{\Psi^\delta}[\varphi^\delta] = \pi_0^\Psi[\varphi]$. Similarly, using the dominated convergence theorem we get that there exists $\bar{\varepsilon}_0 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ such that for any $\varepsilon \in (0, \bar{\varepsilon}_0(C_\varphi))$ we have $\lim_{\delta \to 0} \pi_\varepsilon^{\Psi^\delta}[\varphi^\delta] = \pi_\varepsilon^\Psi[\varphi]$. We conclude upon using Theorem 14 and Theorem 28.

■

4.2 Proofs of Section 3.1

In this section, we prove Theorem 6. Similarly to the proof of Theorem 3 we divide the proof into two parts depending on whether $d \geq p$ in Section 4.2.1 or $d \leq p$ in Section 4.2.2. Our main result, which is a generalization of Theorem 6 is presented in Section 4.2.3. We define $\mathcal{I}_\varepsilon$ and $\mathcal{I}_0$ as in (23) and (24).

4.2.1 The case $d \geq p$

Our first result corresponds to an adaptation of Theorem 12 to the case where $\varphi = \|F\|^k$. Indeed, in this case we have that $\mathcal{I}_0(\varphi) = 0$ and we can tighten our previous results.

Proposition 15 *Assume* $\mathbf{H}1$, $\mathbf{H}2$ *and that* $d \geq p$. *Then there exists* $\eta > 0$ *such that* $\lim_{\varepsilon \to 0} \mathcal{I}_{\varepsilon}^{\mathrm{in}}(\|F\|^k)/\varepsilon = \mathcal{C}_k \mathcal{I}_0(1)$, *where for any* $\varepsilon > 0$, $\mathcal{I}_{\varepsilon}^{\mathrm{in}}(\varphi) = \mathcal{I}_{\varepsilon}(\varphi \mathbb{1}_{F^{-1}(\mathrm{B}(0,\eta))})$ *for any* $\varphi \in \mathrm{C}(\bar{\mathsf{U}}, \mathbb{R})$ *and*

$$\mathcal{C}_k = \textstyle\int_{\mathbb{R}^p} \|t\|^k \exp[-\|t\|^k] \mathrm{d}t \Big/ \textstyle\int_{\mathbb{R}^p} \exp[-\|t\|^k] \mathrm{d}t \ .$$

Proof Let $\eta > 0$ be given by Theorem 11 with $\varphi = 1$. We recall that for any $\varepsilon > 0$ we have that

$$\mathcal{I}_{\varepsilon}^{\mathrm{in}}(\|F\|^k) = \textstyle\int_{F^{-1}(\mathrm{B}(0,\eta))} \|F(x)\|^k \Psi(x) \exp[-\|F(x)\|^k/\varepsilon] \mathrm{d}x / C_{\varepsilon} \ .$$

Using the coarea formula, see Theorem 41, we have for any $\varepsilon > 0$,

$$\mathcal{I}_{\varepsilon}^{\mathrm{in}}(\|F\|^k)/\varepsilon = \textstyle\int_{\mathrm{B}(0,\eta)} (\|t\|^k/\varepsilon) \exp[-\|t\|^k/\varepsilon] \mathcal{L}_t(1) \mathrm{d}t / C_{\varepsilon} \ ,$$

where $\mathcal{L}_t(1)$ is defined in (25) for any $t \in \mathrm{B}(0,\eta)$. Using the change of variable $t \mapsto \varepsilon^{1/k} t$ we have for any $\varepsilon > 0$

$$\mathcal{I}_{\varepsilon}^{\mathrm{in}}(\|F\|^k)/\varepsilon = \textstyle\int_{\mathrm{B}(0,\eta/\varepsilon^{1/k})} \|t\|^k \exp[-\|t\|^k] \mathcal{L}_{t\varepsilon^{1/k}}(1) \mathrm{d}t / C_1 \ .$$

For any $\varepsilon > 0$ let $g_{\varepsilon} : \mathbb{R}^p \to \mathbb{R}$ such that for any $t \in \mathbb{R}^p$, $g_{\varepsilon}(t) = \mathcal{L}_{t\varepsilon^{1/k}}(1) \mathbb{1}_{\mathrm{B}(0,\eta/\varepsilon^{1/k})}(t)$. Note that for any $t \in \mathbb{R}^d$ and $\varepsilon > 0$, $|g_{\varepsilon}(t)| \leq \sup_{\bar{\mathrm{B}}(0,\eta)} |\mathcal{L}_t(1)|$. In addition, we have that for any $t \in \mathbb{R}^d$, $\lim_{\varepsilon \to 0} g_{\varepsilon}(t) = \mathcal{L}_0(1)$ using Theorem 11. Therefore, we get that

$$\lim_{\varepsilon \to 0} \mathcal{I}_{\varepsilon}^{\mathrm{in}}(\|F\|^k)/\varepsilon = \mathcal{L}_0(1) \textstyle\int_{\mathbb{R}^p} \|t\|^k \exp[-\|t\|^k] \mathrm{d}t / C_1 \ .$$

We conclude the proof upon noting that $\mathcal{I}_0(1) = \mathcal{L}_0(1)$ and that $\mathcal{C}_k = \int_{\mathbb{R}^p} \|t\|^k \exp[-\|t\|^k] \mathrm{d}t / C_1$. ■

4.2.2 The case $d \leq p$

We now adapt Theorem 13 to the case where $\varphi = \|F\|^k$.

Proposition 16 *Assume* $\mathbf{H}1$, $\mathbf{H}2$ *and* $d \leq p$. *Let* $\mathsf{U} \subset \mathbb{R}^d$ *open and bounded such that* $F^{-1}(0) \subset \mathsf{U}$. *Then* $\lim_{\varepsilon \to 0} \mathcal{I}_{\varepsilon}^{\mathrm{in}}(\|F\|^k)/\varepsilon = \mathcal{C}_k \mathcal{I}_0(1)$, *where for any* $\varepsilon > 0$, $\mathcal{I}_{\varepsilon}^{\mathrm{in}}(\varphi) = \mathcal{I}_{\varepsilon}(\varphi \mathbb{1}_{\mathsf{V}})$ *for any* $\varphi \in \mathrm{C}(\bar{\mathsf{U}}, \mathbb{R})$, *with* V *open such that* $F^{-1}(0) \subset \mathsf{V} \subset \mathsf{U}$, *and*

$$\mathcal{C}_k = \textstyle\int_{\mathbb{R}^p} \|t\|^k \exp[-\|t\|^k] \mathrm{d}t \Big/ \textstyle\int_{\mathbb{R}^p} \exp[-\|t\|^k] \mathrm{d}t \ .$$

Proof Let $\{x_0^{\ell}\}_{\ell=1}^N$, $\{\mathsf{W}_{\ell}\}_{\ell=1}^N$, $\{\Phi_{\ell}\}_{\ell=1}^N$ and $\{\Omega_{\ell}\}_{\ell=1}^N$ be given as in the proof of Theorem 13. Let $\mathsf{V} = \cup_{\ell=1}^N \Phi_{\ell}(\Omega_{\ell})$. Using for any $\ell \in \{1, \dots, N\}$ the change of variable $x \mapsto \Phi_{\ell}(x)$ and $x \mapsto \varepsilon^{1/k} x$ we have

$$\mathcal{I}_{\varepsilon}^{\mathrm{in}}(\|F\|^k)/\varepsilon = \textstyle\sum_{\ell=1}^N \int_{\Omega_{\ell}/\varepsilon^{1/k}} \|x\|^k \exp[-\|x\|^k] \Psi(\Phi_{\ell}(\varepsilon^{1/k} x)) \det(\mathrm{D}\Phi_{\ell}(\varepsilon^{1/k} x)) \mathrm{d}x / C_1 \ .$$

For any $\ell \in \{1, \dots, N\}$ and $\varepsilon > 0$, let $g_{\ell,\varepsilon} : \mathbb{R}^p \to \mathbb{R}$ such that for any $\ell \in \{1, \dots, N\}$, $\varepsilon > 0$ and $x \in \mathbb{R}^p$ we have

$$g_{\ell,\varepsilon}(x) = \Psi(\Phi_{\ell}(\varepsilon^{1/k} x)) \det(\mathrm{D}\Phi_{\ell}(\varepsilon^{1/k} x)) \mathbb{1}_{\Omega_{\ell}/\varepsilon^{1/k}}(x) \ .$$

Note that for any $\ell \in \{1, \dots, N\}$, $\varepsilon > 0$ and $x \in \mathbb{R}^p$ we have

$$|g_{\ell,\varepsilon}(x)| \leq \sup\{|\Psi(\Phi_\ell(x))| \, |\det(\mathrm{D}\Phi_\ell(x))| \, : \, \ell \in \{1, \dots, N\}, \quad x \in \Omega_\ell\} \ .$$

In addition, we have that for any $\ell \in \{1, \dots, N\}$ and $x \in \mathbb{R}^p$, $\lim_{\varepsilon \to 0} g_{\ell,\varepsilon}(x) = \Psi(x_0^\ell) \mathrm{J}F(x_0^\ell)^{-1}$. We conclude upon using the dominated convergence theorem. ■

4.2.3 Main result

Proposition 17 *Assume* **H** 1 *and* **H** 2*. Let* $\mathsf{U} \subset \mathbb{R}^d$ *be open, bounded and such that* $F^{-1}(0) \subset \mathsf{U}$*. Then* $\lim_{\varepsilon \to 0} \pi_\varepsilon^\Psi(\|F\|^k)/\varepsilon = \mathcal{C}_k$*, where*

$$\mathcal{C}_k = \int_{\mathbb{R}^p} \|t\|^k \exp[-\|t\|^k] \mathrm{d}t \Big/ \int_{\mathbb{R}^p} \exp[-\|t\|^k] \mathrm{d}t \ .$$

Proof Let $\eta > 0$ be given by Theorem 11 with $\varphi = 1$ and let $\mathsf{V} = F^{-1}(\mathrm{B}(0, \eta))$ if $d \geq p$. If $d \leq p$, let V be given by Theorem 16. We have that $F^{-1}(0) \subset \mathsf{V}$. For any $\varepsilon \in (0, \bar{\varepsilon})$ we define $\mathcal{I}_\varepsilon^{\mathrm{out}}(\|F\|^k) = \mathcal{I}_\varepsilon(\|F\|^k \mathbb{1}_{\mathsf{V}^c})$ and $\mathcal{I}_\varepsilon^{\mathrm{in}}(\|F\|^k) = \mathcal{I}_\varepsilon(\|F\|^k \mathbb{1}_{\mathsf{V}})$. Using Theorem 25, we have that $\lim_{\varepsilon \to 0} \mathcal{I}_\varepsilon^{\mathrm{out}}(\|F\|^k)/\varepsilon = 0$. Hence, using Theorem 15 if $d \geq p$ and Theorem 16 if $d \leq p$, we get that $\lim_{\varepsilon \to 0} \mathcal{I}_\varepsilon(\|F\|^k)/\varepsilon = \mathcal{C}_k \mathcal{I}_0(1)$. Similarly, using Theorem 25 and Theorem 12 we have that $\lim_{\varepsilon \to 0} \mathcal{I}_\varepsilon(1) = \mathcal{I}_0(1)$, which concludes the proof upon remarking that for any $\varepsilon > 0$, $\pi_\varepsilon^\Psi(\|F\|^k) = \mathcal{I}_\varepsilon(\|F\|^k)/\mathcal{I}_\varepsilon(1)$. ■

We are now ready to prove a generalization of Theorem 6.

Proposition 18 *Let* $\mu \in \mathscr{P}(\mathbb{R}^d)$*,* $F : \mathbb{R}^d \to \mathbb{R}^p$ *and* $k \in \mathbb{N}^\star$*. Assume that the conditions of Theorem 5 with* $G_\varepsilon = \|F\|^k - \varepsilon$ *for any* $\varepsilon > 0$ *are satisfied and for any* $\varepsilon > 0$*, let* ρ_ε *the macrocanonical distribution with constraint* G_ε *and reference measure* μ*. Assume that there exists* $\Psi : \mathbb{R}^d \to \mathbb{R}_+$ *such that* μ *admits a density w.r.t. the Lebesgue measure given by* Ψ*. In addition, assume that* **H**1 *and* **H**2 *hold. Then, we have that* $\theta_\varepsilon \sim_{\varepsilon \to 0} \mathcal{C}_k/\varepsilon$*, where*

$$\mathcal{C}_k = \int_{\mathbb{R}^p} \|t\|^k \exp[-\|t\|^k] \mathrm{d}t \Big/ \int_{\mathbb{R}^p} \exp[-\|t\|^k] \mathrm{d}t \ .$$

Proof Recall that using Theorem 5 we have that for any $\varepsilon > 0$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$

$$\rho_\varepsilon(\mathsf{A}) = \int_{\mathsf{A}} \Psi(x) \exp[-\theta_\varepsilon \|F(x)\|^k] \mathrm{d}x \Big/ \int_{\mathbb{R}^d} \Psi(x) \exp[-\theta_\varepsilon \|F(x)\|^k] \mathrm{d}x \ .$$

Hence, using Theorem 17 we have that $\lim_{\varepsilon \to 0} \rho_\varepsilon(\|F\|^k) \theta_\varepsilon = \mathcal{C}_k$. Since $\rho_\varepsilon[G_\varepsilon] = 0$ we have also have that $\rho_\varepsilon[\|F\|^k] = \varepsilon$, which concludes the proof. ■

Note that Theorem 6 is obtained upon noting that $\mathcal{C}_2 = p/2$.

4.3 Proofs of Section 3.2

In Section 4.3.1, we establish Theorem 8. In Section 4.3.2 we use stability results from Raginsky et al. (2017); Bousquet and Elisseeff (2002) to obtain Theorem 9. Finally, we prove Theorem 10 in Section 4.3.3. Additional technical results are postponed to Appendix B.

4.3.1 Proof of Theorem 8

In this section, we prove Theorem 8 which is an application of a parametric version of the results presented in Section 4.1.2. We refer to Appendix B for a detailed presentation of these results. We will apply them in the context of the non-convex minimization setting presented in Section 3.2 which we recall here.

We aim at minimizing $U : \mathbb{R}^d \to \mathbb{R}$. We assume that there exist a topological space $(\mathsf{Z}, \mathcal{B}(\mathsf{Z}))$, a probability measure $\mu \in \mathscr{P}(\mathsf{Z}, \mathcal{Z})$ and $u : \mathbb{R}^d \times \mathsf{Z} \to \mathbb{R}_+$ such that for any $x \in \mathbb{R}^d$

$$U(x) = \textstyle\int_{\mathsf{Z}} u(x,z)\mathrm{d}\mu(z) \ .$$

For any $n \in \mathbb{N}$ we define $U_n : \mathbb{R}^d \times \mathsf{Z}^n \to \mathbb{R}$ such that for any $x \in \mathbb{R}^d$ and $z^{1:n} = \{z_i\}_{i=1}^n \in \mathsf{Z}^n$

$$U_n(x, z^{1:n}) = (1/n)\textstyle\sum_{i=1}^n u(x, z_i) \ .$$

For all $\varepsilon > 0$, when it is well-defined we denote by $\mathrm{S}_\varepsilon : \mathsf{Z}^n \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$ the Markov kernel such that for any $z^{1:n} \in \mathsf{Z}^n$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\updelta_{z^{1:n}}\mathrm{S}_\varepsilon(\mathsf{A}) = \textstyle\int_{\mathsf{A}} \exp[-U_n(x,z^{1:n})/\varepsilon]\mathrm{d}x / \textstyle\int_{\mathbb{R}^d} \exp[-U_n(x,z^{1:n})/\varepsilon]\mathrm{d}x \ .$$

Similarly, when it is well-defined, we denote by $\mathrm{S}_0 : \mathsf{Z}^n \times \mathcal{B}(\mathbb{R}^d) \to [0,1]$ the Markov kernel such that for any $z^{1:n} \in \mathsf{Z}^n$ and $\mathsf{A} \in \mathcal{B}(\mathbb{R}^d)$ we have

$$\updelta_{z^{1:n}}\mathrm{S}_0(\mathsf{A}) = \textstyle\int_{\mathsf{A}\cap\mathsf{C}_n(z^{1:n})} \det(\nabla_x^2 U_n(x,z^{1:n}))^{-1}\mathrm{d}\mathcal{H}^0(x) \Big/ \textstyle\int_{\mathsf{C}_n(z^{1:n})} \det(\nabla_x^2 U_n(x,z^{1:n}))^{-1}\mathrm{d}\mathcal{H}^0(x) \ ,$$

where $\mathsf{C}_n(z^{1:n}) = \arg\min U_n(\cdot, z^{1:n})$. We recall that for any $\upbeta > 0$, $\sigma_\upbeta^\star$ is defined in (17). We begin with the following proposition.

Proposition 19 *Let $n \in \mathbb{N}$ and assume* **H**3(n) *and* **H**4(n)*. Then there exist $C \geq 0$ and $\upbeta, \bar{\varepsilon} > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$*

$$\begin{aligned}&\textstyle\int_{\mathsf{Z}^n} \mathbf{W}_1(\updelta_{z^{1:n}}\mathrm{S}_\varepsilon, \updelta_{z^{1:n}}\mathrm{S}_0)\mathrm{d}\mu^{\otimes n}(z^{1:n}) \\ &\qquad \leq C(1+D_n)(\varepsilon^{1/2} + \varepsilon^{-d/2}\textstyle\int_{\mathsf{Z}^{1:n}} \exp[-c^\star(z^{1:n})/\varepsilon]\mathrm{d}\mu^{\otimes n}(z^{1:n})) \ ,\end{aligned}$$

with $D_n = \int_{\mathsf{Z}^n} \sigma_\upbeta^\star(z^{1:n})\mathrm{d}\mu^{\otimes n}(z^{1:n}) < +\infty$ and $C, \bar{\varepsilon}, \upbeta$ that do not depend on n.

Proof The proof of this result is a direct application of Theorem 33 which is a parametric version of Theorem 3. In order to apply Theorem 33, we check that **H**5 is satisfied for $\mathsf{Z} \leftarrow \mathsf{Z}^n$ and $u \leftarrow U_n$. We first check that there exists $\mathtt{m}_0, \upalpha_0 > 0$ and R_0 such that for any $x \in \mathbb{R}^d$ with $\|x\| \geq R_0$ and $z^{1:n} \in \mathsf{Z}^n$, $U_n(x, z^{1:n}) \geq \mathtt{m}_0 \|x\|^{\upalpha_0}$. We have that for any $x \in \mathbb{R}^d$ and $z \in \mathsf{Z}$

$$u(x,z) = u(0,z) + \textstyle\int_0^1 \langle \nabla_x u(tx,z), x\rangle\mathrm{d}t \geq -A - \mathtt{c} + (\mathtt{m}/2)\|x\|^2 \ .$$

Hence, for any $x \in \mathbb{R}^d$ with $\|x\| \geq 2((A+\mathtt{c})/\mathtt{m})^{1/2}$ we have that for any $z^{1:n} \in \mathsf{Z}^n$, $U_n(x, z^{1:n}) \geq (\mathtt{m}/4)\|x\|^2$. Let $z^{1:n} \in \mathsf{Z}$, we show that the number of minimizers of $U_n(\cdot, z^{1:n})$ is bounded. Since $x \in \mathbb{R}^d$ with $\|x\| \geq 2((A+\mathtt{c})/\mathtt{m})^{1/2}$ we have that $U_n(x, z^{1:n}) \geq (\mathtt{m}/4)\|x\|^2$ and $|U_n(0, z^{1:n})| \leq A$, there exists K compact such that $\arg\min\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\} \subset \mathsf{K}$

(see the remark following **H** 5). Assume that the number of minimizers is not bounded. In this case, there exists $(x_k)_{k\in\mathbb{N}} \in \arg\min\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\}^{\mathbb{N}}$ such that for any $k, \ell \in \mathbb{N}$, $x_k \neq x_\ell$. Up to extraction we can assume that there exists $x^\star \in \mathbb{R}^d$ such that $\lim_{k\to+\infty} x_k = x^\star$. Note that $x^\star \in \arg\min\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\}$ by continuity. In particular, $\det(\nabla_x^2 U_n(x^\star, z^{1:n})) > 0$ and there exists $r > 0$ such that for any $x \in \bar{\mathrm{B}}(0, r)$, $x \in \arg\min\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\}$ implies that $x = x^\star$. Hence, there exists $k_0 \in \mathbb{N}$ such that for any $k \in \mathbb{N}$ with $k \geq k_0$, $x_k = x^\star$ which is absurd. Hence, combining this result, **H** 3(n) and **H** 4(n), we can apply Theorem 33 which states that for any $\varphi : \ \mathbb{R}^d \to \mathbb{R}$ $M_{1,\varphi}$-Lipschitz function with $M_{1,\varphi}, C_\varphi \geq 0$ such that for any $x \in \mathbb{R}^d$, $|\varphi(x)| \leq C_\varphi \exp[C_\varphi \|x\|^\alpha]$ then, there exist $B_2 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ and $\beta > 0$ such that

$$\begin{aligned} &|\delta_{z^{1:n}}\mathrm{S}_\varepsilon[\varphi] - \delta_{z^{1:n}}\mathrm{S}_0[\varphi]| \\ &\qquad \leq B_2(C_\varphi)(1 + M_{0,\varphi} + M_{1,\varphi})(1 + \sigma_\beta^\star(z^{1:n}))\{\varepsilon^{1/2} + \varepsilon^{-d/2}\exp[-c^\star(z^{1:n})/\varepsilon]\} \ , \end{aligned} \tag{39}$$

with $M_{0,\varphi} = \sup\{|\varphi(x)| \ : \ x \in \mathsf{K}\}$, K, B_2 and β that do not depend on z, and B_2 non-decreasing. The rest of the proof is similar to the one of Theorem 2 upon replacing Theorem 1 by (39). ■

The proof of Theorem 8 is then a direct application of (Raginsky et al., 2017, Proposition 3.3), (Raginsky et al., 2017, Equation (3.3)) and Theorem 19.

4.3.2 Proof of Theorem 9

In this section, we prove Theorem 9. We start by recalling a proposition from (Raginsky et al., 2017, Proposition 3.5) about the uniform stability of the exponential measure with potential U_n, see Bousquet and Elisseeff (2002) for a definition of the uniform stability.

Lemma 20 *Assume that* **H** 3(n) *holds uniformly w.r.t.* n. *Then for any* $n \in \mathbb{N}$, $\varepsilon > 0$, $z_0^{1:n}, z_1^{1:n} \in \mathsf{Z}^n$ *which only differs along one index, we have*

$$|\delta_{z_0^{1:n}}\mathrm{S}_\varepsilon u(\cdot, z) - \delta_{z_1^{1:n}}\mathrm{S}_\varepsilon u(\cdot, z)| \leq 4(\mathtt{M}^2(\mathtt{c} + d\varepsilon)/\mathtt{m} + B^2)\mathtt{c}_{\mathtt{LS}}(\varepsilon)/(n\varepsilon) \ ,$$

where $\mathtt{c}_{\mathtt{LS}}(\varepsilon) \geq 0$ *is such that for any* $z^{1:n} \in \mathsf{Z}^n$, $\delta_{z^{1:n}}\mathrm{S}_\varepsilon$ *satisfies the logarithmic Sobolev inequality with constant* $\mathtt{c}_{\mathtt{LS}}(\varepsilon)$, *see (Raginsky et al., 2017, Proposition 3.2).*

For completeness, we recall that a probability measure $\nu \in \mathscr{P}(\mathbb{R}^d)$ is said to satisfy the logarithmic Sobolev inequality with constant $\mathtt{c}_{\mathtt{LS}}$ if for any $\pi \in \mathscr{P}(\mathbb{R}^d)$ with positive density w.r.t. ν given by $g \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$ we have

$$\mathrm{KL}\,(\pi|\nu) \leq 2\mathtt{c}_{\mathtt{LS}} \textstyle\int_{\mathbb{R}^d} \|\nabla \log(g(x))\|^2 \mathrm{d}\pi(x) \ .$$

We can relate the constant appearing in the logarithmic Sobolev inequality with the uniform spectral gap given for any $\varepsilon > 0$ by

$$\lambda^\star(\varepsilon) = \inf_{z^{1:n}\in\mathsf{Z}^n,\ h\in\mathrm{C}^1(\mathbb{R}^d)\cap\mathrm{L}^2(\delta_{z^{1:n}}\mathrm{S}_\varepsilon)}\{\delta_{z^{1:n}}\mathrm{S}_\varepsilon[\|\nabla h\|^2]/\delta_{z^{1:n}}\mathrm{S}_\varepsilon[|h|^2] \ : \ h \neq 0, \ \delta_{z^{1:n}}\mathrm{S}_\varepsilon[h] = 0\} \ .$$

More precisely, we have the following proposition (see (Raginsky et al., 2017, Proposition 3.2, Appendix B), see also Bakry et al. (2008)).

Proposition 21 *Assume that* $\mathbf{H}3(n)$ *holds uniformly w.r.t.* n. *Then, there exist* $A_0, A_1, \bar{\varepsilon} > 0$ *such that for any* $\varepsilon \in (0, \bar{\varepsilon}]$

$$\mathtt{c}_{\mathtt{LS}}(\varepsilon) \leq A_0(1 + (\lambda^\star(\varepsilon)\varepsilon)^{-1}) \; .$$

In addition, $\lambda^\star(\varepsilon) \geq (1/A_1)\exp[-A_1/\varepsilon]$.

Therefore, there exists $A \geq 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon}]$

$$\mathtt{c}_{\mathtt{LS}}(\varepsilon) \leq (A/\varepsilon)\exp[A/\varepsilon] \; . \tag{40}$$

We are now ready to show that the limiting measures are stable.

Proposition 22 *Assume that* $\mathbf{H}\,3(n)$ *and* $\mathbf{H}\,4(n)$ *hold uniformly w.r.t.* $n \in \mathbb{N}$. *Assume that*

$$\lim_{\varepsilon \to 0} \sup\{\varepsilon^{-d/2} \textstyle\int_{\mathsf{Z}^n} \exp[-c^\star(z^{1:n})/\varepsilon] \mathrm{d}\mu^{\otimes n}(z^{1:n}) \; : \; n \in \mathbb{N}\} = 0 \; . \tag{41}$$

Then for any $\delta > 0$, *there exists* $n_0 \in \mathbb{N}$ *such that for any* $z \in \mathsf{Z}$, $n \in \mathbb{N}$ *with* $n \geq n_0$ *and* $j \in \{1, \dots, n\}$, *we have*

$$\textstyle\int_{\mathsf{Z}^{n+1}} |\updelta_{z_0^{1:n}} \mathrm{S}_0 u(\cdot, z) - \updelta_{z_1^{1:n}} \mathrm{S}_0 u(\cdot, z)| \mathrm{d}\mu^{\otimes n}(z_0^{1:n}) \mathrm{d}\mu(z_{1,j}) \leq \delta \; , \tag{42}$$

where for any $z_0^{1:n} \in \mathsf{Z}^n$ *we let* $z_1^{1:n} \in \mathsf{Z}^n$ *with* $z_{0,i} = z_{1,i}$ *for any* $i \in \{1, \dots, N\}$ *such that* $i \neq j$. *In addition, assume that there exist* $C_0, \alpha, \bar{\varepsilon}_0 > 0$ *such that for any* $n \in \mathbb{N}$ *and* $\varepsilon \in (0, \bar{\varepsilon}_0]$

$$\varepsilon^{-d/2} \textstyle\int_{\mathsf{Z}^n} \exp[-c^\star(z^{1:n})/\varepsilon] \mathrm{d}\mu^{\otimes n}(z^{1:n}) \leq C_0 \varepsilon^\alpha \; , \tag{43}$$

Then there exists $n_0 \in \mathbb{N}$ *such that for any* $\eta \in (0, 1)$, *there exists* $C \geq 0$ *such that for any* $z \in \mathsf{Z}$, $n \in \mathbb{N}$ *with* $n \geq n_0$ *and* $j \in \{1, \dots, n\}$, *we have*

$$\textstyle\int_{\mathsf{Z}^{n+1}} |\updelta_{z_0^{1:n}} \mathrm{S}_0 u(\cdot, z) - \updelta_{z_1^{1:n}} \mathrm{S}_0 u(\cdot, z)| \mathrm{d}\mu^{\otimes n}(z_0^{1:n}) \mathrm{d}\mu(z_{1,j}) \leq C / \log(n)^{s\eta} \; , \tag{44}$$

where $s = \min(\alpha/2, 1/4)$ *and for any* $z_0^{1:n} \in \mathsf{Z}^n$ *we let* $z_1^{1:n} \in \mathsf{Z}^n$ *with* $z_{0,i} = z_{1,i}$ *for any* $i \in \{1, \dots, N\}$ *such that* $i \neq j$.

Proof Let $\delta > 0$, $\eta \in (0, 1)$. Using the triangle inequality we have for any $\varepsilon > 0$, $n \in \mathbb{N}$ and $z_0^{1:n}, z_1^{1:n}, z \in \mathsf{Z}^n$ where there exists $j \in \{1, \dots, n\}$ such that for any $i \in \{1, \dots, n\}$, $i \neq j$, $z_{0,i} = z_{1,i}$

$$\begin{aligned} |\updelta_{z_0^{1:n}} \mathrm{S}_0 u(\cdot, z) - \updelta_{z_1^{1:n}} \mathrm{S}_0 u(\cdot, z)| &\leq |\updelta_{z_0^{1:n}} \mathrm{S}_0 u(\cdot, z) - \updelta_{z_0^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \\ &+ |\updelta_{z_0^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z) - \updelta_{z_1^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| + |\updelta_{z_1^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z) - \updelta_{z_1^{1:n}} \mathrm{S}_0 u(\cdot, z)| \; . \end{aligned} \tag{45}$$

Using (Raginsky et al., 2017, Lemma 3.5) and Theorem 34, there exists $C_0 \geq 0$ such that for any $i \in \{0, 1\}$, $n \in \mathbb{N}$, $z_0^{1:n}, z_1^{1:n}, z \in \mathsf{Z}^n$ where there exists $j \in \{1, \dots, n\}$ such that for any $i \in \{1, \dots, n\}$, $i \neq j$, $z_{0,i} = z_{1,i}$ and $\varepsilon \in (0, \bar{\varepsilon}_1]$ (where $\bar{\varepsilon}_1$ is given by $\bar{\varepsilon}_1 \leftarrow \bar{\varepsilon}$ in Theorem 19) we have

$$|\updelta_{z_i^{1:n}} \mathrm{S}_0 u(\cdot, z) - \updelta_{z_i^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \leq C_0 \mathbf{W}_2(\updelta_{z_i^{1:n}} \mathrm{S}_0, \updelta_{z_i^{1:n}} \mathrm{S}_\varepsilon) \; .$$

Using this result, Theorem 34 and Theorem 35, there exists $C_1 \geq 0$ such that for any $i \in \{0,1\}$, $n \in \mathbb{N}$, $z_0^{1:n}, z_1^{1:n}, z \in \mathsf{Z}^n$ where there exists $j \in \{1, \ldots, n\}$ such that for any $i \in \{1, \ldots, n\}$, $i \neq j$, $z_{0,i} = z_{1,i}$ and $\varepsilon \in (0, \bar{\varepsilon}_1]$ (where $\bar{\varepsilon}_1$ is given by $\bar{\varepsilon}_1 \leftarrow \bar{\varepsilon}$ in Theorem 19) we have

$$|\delta_{z_i^{1:n}} \mathrm{S}_0 u(\cdot, z) - \delta_{z_i^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \leq C_1 \mathbf{W}_1^{\eta/2}(\delta_{z_i^{1:n}} \mathrm{S}_0, \delta_{z_i^{1:n}} \mathrm{S}_\varepsilon) \ . \tag{46}$$

We divide the rest of the proof into two parts. First, we start with our qualitative result by showing that (42) holds under (41). Then, we turn to our quantitative bounds by showing that (44) holds under (43)

(a) Using Theorem 19 and (41) we have that for any $i \in \{0,1\}$

$$\lim_{\varepsilon \to 0} \sup\{\textstyle\int_{\mathsf{Z}^n} \mathbf{W}_1(\delta_{z_i^{1:n}} \mathrm{S}_0, \delta_{z_i^{1:n}} \mathrm{S}_\varepsilon) \mathrm{d}\mu^{\otimes n}(z_i^{1:n}) \ : \ n \in \mathbb{N}\} = 0 \ .$$

Combining this result, (46) and that $t \mapsto t^{\eta/2}$ is concave, we get that for any $i \in \{0,1\}$

$$\lim_{\varepsilon \to 0} \sup\{\textstyle\int_{\mathsf{Z}^n} |\delta_{z_i^{1:n}} \mathrm{S}_0 u(\cdot, z) - \delta_{z_i^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \mathrm{d}\mu^{\otimes n}(z_i^{1:n}) \ : \ n \in \mathbb{N}\} = 0 \ . \tag{47}$$

In addition, using Theorem 20 we have for any $n \in \mathbb{N}$ and $\varepsilon > 0$

$$\textstyle\int_{\mathsf{Z}^{n+1}} |\delta_{z_0^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z) - \delta_{z_1^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \mathrm{d}\mu^{\otimes}(z_0^{1:n}) \mathrm{d}\mu(z_{1,j}) \leq 4(\mathtt{M}^2(\mathtt{c} + d\varepsilon)/\mathtt{m} + B^2)\mathtt{c}_{\mathtt{LS}}(\varepsilon)/(n\varepsilon) \ ,$$

Combining this result, (45) and (47) we get that there exists $\varepsilon > 0$ such that

$$\begin{aligned} &\textstyle\int_{\mathsf{Z}^{n+1}} |\delta_{z_0^{1:n}} \mathrm{S}_0 u(\cdot, z) - \delta_{z_1^{1:n}} \mathrm{S}_0 u(\cdot, z)| \mathrm{d}\mu^{\otimes}(z_0^{1:n}) \mathrm{d}\mu(z_{1,j}) \\ &\qquad \leq 4(\mathtt{M}^2(\mathtt{c} + d\bar{\varepsilon})/\mathtt{m} + B^2)\mathtt{c}_{\mathtt{LS}}(\varepsilon)/(n\varepsilon) \\ &\qquad\quad + 2\sup\{\textstyle\int_{\mathsf{Z}^n} |\delta_{z^{1:n}} \mathrm{S}_0 u(\cdot, z) - \delta_{z^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \mathrm{d}\mu^{\otimes n}(z^{1:n}) \ : \ n \in \mathbb{N}\} \\ &\qquad \leq 4(\mathtt{M}^2(\mathtt{c} + d\bar{\varepsilon})/\mathtt{m} + B^2)\mathtt{c}_{\mathtt{LS}}(\varepsilon)/(n\varepsilon) + \delta/2 \ . \end{aligned}$$

Hence, there exists $n_0 \in \mathbb{N}$ such that for any $n \in \mathbb{N}$ with $n \geq n_0$, $4(\mathtt{M}^2(\mathtt{c} + d\bar{\varepsilon})/\mathtt{m} + B^2)\mathtt{c}_{\mathtt{LS}}(\varepsilon)/(n\varepsilon) \leq \delta/2$, which concludes the first part of the proof.

(b) Let $n \in \mathbb{N}$ with $n \geq n_0$ and $(2/A)\log(n_0)^{-1} < \bar{\varepsilon} = \min(\bar{\varepsilon}_0, \bar{\varepsilon}_1, \bar{\varepsilon}_2)$ (where $\bar{\varepsilon}_1$ is given by $\bar{\varepsilon}_1 \leftarrow \bar{\varepsilon}$ in Theorem 19 and $A, \bar{\varepsilon}_2$ are given by $\bar{\varepsilon}_2 \leftarrow \bar{\varepsilon}$ in Theorem 21 and (40)). In addition, using Theorem 19, there exists $C_2 \geq 0$ (that does not depend on n) such that for any $i \in \{0,1\}$ and $\varepsilon \in (0, \bar{\varepsilon}]$

$$\textstyle\int_{\mathsf{Z}^n} \mathbf{W}_1(\delta_{z_i^{1:n}} \mathrm{S}_0, \delta_{z_i^{1:n}} \mathrm{S}_\varepsilon) \mathrm{d}\mu^{\otimes n}(z_i^{1:n}) \leq C_2 \max(\varepsilon^{1/2}, \varepsilon^{\alpha}) \ .$$

Combining this result, (46) and the fact that $t \mapsto t^{\eta/2}$ is concave, we get that there exists $C_3 \geq 0$ (that does not depend on n) such that for any $i \in \{0,1\}$ and $\varepsilon \in (0, \bar{\varepsilon}]$

$$\textstyle\int_{\mathsf{Z}^n} |\delta_{z_i^{1:n}} \mathrm{S}_0 u(\cdot, z) - \delta_{z_i^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \mathrm{d}\mu^{\otimes n}(z_i^{1:n}) \leq C_3 \max(\varepsilon^{\eta/4}, \varepsilon^{\alpha\eta/2}) \ , \tag{48}$$

In addition, using Theorem 20 we have for any $\varepsilon \in (0, \bar{\varepsilon}]$

$$\textstyle\int_{\mathsf{Z}^{n+1}} |\delta_{z_0^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z) - \delta_{z_1^{1:n}} \mathrm{S}_\varepsilon u(\cdot, z)| \mathrm{d}\mu^{\otimes}(z_0^{1:n}) \mathrm{d}\mu(z_{1,j}) \leq 4(\mathtt{M}^2(\mathtt{c} + d\varepsilon)/\mathtt{m} + B^2)\mathtt{c}_{\mathtt{LS}}(\varepsilon)/(n\varepsilon) \ .$$

Combining this result, (45) and (48) we get

$$\begin{aligned}\textstyle\int_{\mathsf{Z}^{n+1}} |\delta_{z_0^{1:n}}\mathrm{S}_0 u(\cdot, z) - \delta_{z_1^{1:n}}\mathrm{S}_0 u(\cdot, z)| \mathrm{d}\mu^{\otimes}(z_0^{1:n})\mathrm{d}\mu(z_{1,j}) \\ &\leq 4(\mathtt{M}^2(\mathtt{c} + d\bar{\varepsilon})/\mathtt{m} + B^2)\mathtt{c}_{\mathtt{LS}}(\varepsilon)/(n\varepsilon) + 2C_3\varepsilon^{s\eta} \\ &\leq 4A(\mathtt{M}^2(\mathtt{c} + d\bar{\varepsilon})/\mathtt{m} + B^2)\exp[A/\varepsilon]/(n\varepsilon^2) + 2C_3\varepsilon^{s\eta} \ ,\end{aligned}$$

with $s = \min(\alpha/2, 1/4)$. We conclude the proof upon letting $\varepsilon = (2/A)\log(n)^{-1}$.

■

The stability of the limiting measures allows us to establish Theorem 9 which provides quantitative bounds on $\mu^{\otimes n}\mathrm{S}_0[U] - U^\star$ for large values of $n \in \mathbb{N}$. Indeed, once Theorem 22 is established the proof of Theorem 9 is classical and follows the lines of (Raginsky et al., 2017, Section 3.7).

Proof Let $n \in \mathbb{N}$ and $n \geq n_0$ with n_0 given by Theorem 9. Using the definition of U and S_0 we have $\mu^{\otimes n}\mathrm{S}_0[U] = \textstyle\int_{\mathsf{Z}^n}\int_{\mathsf{Z}^n}\int_{\mathbb{R}^d} U_n(x, \tilde{z}^{1:n})\mathrm{S}_0(z^{1:n}, \mathrm{d}x)\mathrm{d}\mu^{\otimes n}(z^{1:n})\mathrm{d}\mu^{\otimes n}(\tilde{z}^{1:n})$. For any $z^{1:n} \in \mathsf{Z}^n$ we define $U_n^\star(z^{1:n}) = \inf\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\}$. Using that $U^\star \geq \textstyle\int_{\mathsf{Z}^n} U_n^\star(z^{1:n})\mathrm{d}\mu^{\otimes n}(z^{1:n})$ and that $\delta_{z^{1:n}}\mathrm{S}_0$ is concentrated on $\operatorname{argmin}\{U_n(x, z^{1:n}) \ : \ x \in \mathbb{R}^d\}$ we have

$$\begin{aligned}&\mu^{\otimes n}\mathrm{S}_0[U] - U^\star \leq \mu^{\otimes n}\mathrm{S}_0[U] - \textstyle\int_{\mathsf{Z}^n} U_n^\star(z^{1:n})\mathrm{d}\mu^{\otimes n}(z^{1:n}) \\ &\leq \mu^{\otimes n}\mathrm{S}_0[U] - \textstyle\int_{\mathsf{Z}^n}\int_{\mathbb{R}^d} U_n(x, z^{1:n})\mathrm{S}_0(z^{1:n}, \mathrm{d}x)\mathrm{d}\mu^{\otimes n}(z^{1:n}) \\ &\leq \textstyle\int_{\mathsf{Z}^n}\int_{\mathsf{Z}^n}\int_{\mathbb{R}^d} U_n(x, \tilde{z}^{1:n})\mathrm{S}_0(z^{1:n}, \mathrm{d}x)\mathrm{d}\mu^{\otimes n}(z^{1:n})\mathrm{d}\mu^{\otimes n}(\tilde{z}^{1:n}) \\ &\qquad - \textstyle\int_{\mathsf{Z}^n}\int_{\mathbb{R}^d} U_n(x, z^{1:n})\mathrm{S}_0(z^{1:n}, \mathrm{d}x)\mathrm{d}\mu^{\otimes n}(z^{1:n}) \\ &\leq (1/n)\textstyle\sum_{i=1}^n \int_{\mathsf{Z}}\int_{\mathsf{Z}^n}\int_{\mathbb{R}^d}\{u(x, \tilde{z}_i) - u(x, z_i)\}\mathrm{S}_0(z^{1:n}, \mathrm{d}x)\mathrm{d}\mu^{\otimes n}(z^{1:n})\mathrm{d}\mu(\tilde{z}_i) \\ &\leq (1/n)\textstyle\sum_{i=1}^n \int_{\mathsf{Z}}\int_{\mathsf{Z}^n}\{\int_{\mathbb{R}^d} u(x, z_i)\mathrm{S}_0(z^{1:n}, \mathrm{d}x) - \int_{\mathbb{R}^d} u(x, z_i)\mathrm{S}_0(\tilde{z}_i^{1:n}, \mathrm{d}x)\}\}\mathrm{d}\mu^{\otimes n}(z^{1:n})\mathrm{d}\mu(\tilde{z}_i) \ ,\end{aligned}$$

where for any $i \in \{1, \dots, n\}$, we have that for any $j \in \{1, \dots, n\}$, $\tilde{z}_{i,j} = z_i$ and $\tilde{z}_{i,i} = \tilde{z}_i$. We conclude using Theorem 9.

4.3.3 Proof of Theorem 10

We recall that u is given in (20). We divide the proof into two parts.

(a) First, we prove that $\lim_{n\to+\infty} \mu^{\otimes n}\mathrm{S}_0[\varphi] = (\varphi(-\pi) + \varphi(\pi))/2$. Let $n \in \mathbb{N}$. Assume that $z < 0$. Then the minimum of $x \mapsto u(x, z)$ is attained on $[\pi, +\infty)$. Denote $\bar{u} : \ \mathbb{R} \times [-1/2, 0)$ such that for any $x \in \mathbb{R}$ and $z \in [-1/2, 0)$, $\bar{u}(x, z) = h(x) + xz + \pi z + 1 - \cos(3x)$ with $h(x) = x^4/(1 + x^2)$. Note that for any $x \geq 0$ and $z \in [-1/2, 0)$, $\bar{u}(x, z) = u(x + \pi, z)$. There exists $a \in [0, \pi/3]$ such that for any $x \geq 0$, $h'(x) - 1 \leq 0$ if $x \leq a$ and $h'(x) - 1 > 0$ otherwise. Hence, we get that for any $x \in \mathbb{R}$ with $x \geq \pi/3$ and $z \in [-1/2, 0)$

$$\bar{u}(x, z) - \bar{u}(0, z) = \bar{u}(x, z) - \pi z \geq h(\pi/3) - \pi/6 > 0 \ .$$

Therefore, for any $z \in [-1/2, 0)$, the global minimum of $x \mapsto u(x, z)$ is attained on $(\pi, 4\pi/3)$. We have that for any $x \in [\pi/6, \pi/3]$ and $z \in [-1/2, 0)$

$$\partial_1\bar{u}(x, z) \geq h'(x) - 1/2 + 3(2 - (6/\pi)x) > 0 \ .$$

In addition, we have that for any $z \in [-1/2, 0)$, $\partial_1 \bar{u}(0, z) = -z$. Hence, there exists $\bar{x}(z) \in [0, \pi/6]$ such that $\partial_1 \bar{u}(\bar{x}(z), z) = 0$. In addition we have that for any $z \in [-1/2, 0)$, $x \mapsto \partial_1 \bar{u}(x, z)$ is increasing on $[0, \pi/6]$. Therefore, for any $z \in [-1/2, 0)$ there exists a unique minimizer of $x \mapsto u(x, z)$ on $[\pi, 5\pi/6]$ given by $x^\star(z) = \pi + \bar{x}(z)$. The same conclusion holds with $x^\star(z) \in [-5\pi/6, -\pi]$ if $z \in (0, 1/2]$. We have that $\lim_{z \to 0} \sup\{\|u(x, z) - u(x, 0)\| : x \in [-5\pi/6, 5\pi/6]\} = 0$. Therefore we have that every limit point of $\{x^\star(z)\}_{z<0}$ when $z \to 0$ is a global minimizer of $u(\cdot, 0)$. But recall that $\{x^\star(z)\}_{z<0} \subset [\pi, 5\pi/6]$. Therefore, every limit point of $\{x^\star(z)\}_{z<0}$ is equal to π and we have that $\lim_{z \to 0, z > 0} x^\star(z) = \pi$. For any $n \in \mathbb{N}$, denote by g_n the density of $T_\# \mu^{\otimes n}$ where $T : \mathsf{Z}^n \to \mathsf{Z}$ is given by $T(z^{1:n}) = (1/n) \sum_{i=1}^n z_i$. For any $r, \varepsilon > 0$ there exists $n_0 \in \mathbb{N}$ such that for any $n \in \mathbb{N}$ with $n \geq n_0$, $\int_{\mathrm{B}(0,r)^{\mathrm{c}}} g_n(z) \mathrm{d}z \leq \varepsilon$. Let $\varphi \in \mathrm{C}(\mathbb{R}, \mathbb{R})$ bounded and $\varepsilon > 0$. Let $r > 0$ such that for any $z \in [0, r]$, $|\varphi(x^\star(z)) - \varphi(-\pi)| \leq \varepsilon$ and for any $z \in [-r, 0]$, $|\varphi(x^\star(z)) - \varphi(\pi)| \leq \varepsilon$. Using this result, we have for any $n \in \mathbb{N}$ with $n \geq n_0$

$$
\begin{aligned}
&|\mu^{\otimes n} \mathrm{S}_0[\varphi] - (\varphi(-\pi) + \varphi(\pi))/2| \\
&\qquad \leq \textstyle\int_0^{+\infty} |\varphi(x^\star(z)) - \varphi(-\pi)| g_n(z) \mathrm{d}z + \int_{-\infty}^0 |\varphi(x^\star(z)) - \varphi(\pi)| g_n(z) \mathrm{d}z \\
&\qquad \leq (1 + 2\|\varphi\|_\infty)\varepsilon \; .
\end{aligned}
$$

Therefore, we get that $\lim_{n \to +\infty} \mu^{\otimes n} \mathrm{S}_0[\varphi] = (\varphi(-\pi) + \varphi(\pi))/2$, which concludes the first part of the proof.

(b) Second, we prove that for any $n \in \mathbb{N}$, $\mu^{\otimes n} \mathrm{S}_0[U] - U^\star \leq (\pi/(6\sqrt{3})) n^{-1/2}$. Note that for any $n \in \mathbb{N}$ and $z^{1:n} \in [-1/2, 1/2]^n$, with $\sum_{i=1}^n z_i \neq 0$, $\delta_{z^{1:n}} \mathrm{S}_0[U] = U(x^\star(\bar{z}^{1:n}))$ with $x^\star(\bar{z}^{1:n}) \in [-5\pi/6, 5\pi/6]$. We also have that for any $x \in \mathbb{R}$ and $z_1, z_2 \in [-1/2, 1/2]$, $|u(x, z_1) - u(x, z_2)| \leq |x||z_1 - z_2|$. In particular, we have that for any $z^{1:n} \in [-1/2, 1/2]^n$ and $x \in [-\pi/3, \pi/3]$

$$
|U(x) - U_n(x, z^{1:n})| \leq (\pi/3) |(1/n) \textstyle\sum_{i=1}^n z_i| \; .
$$

Hence, using this result and that $U^\star = U(\pi/3) = 0$, we have that for any $z^{1:n} \in [-1/2, 1/2]$

$$
U_n(x^\star(\bar{z}^{1:n}), z^{1:n}) \leq U_n(\pi/3, z^{1:n}) \leq U^\star + (\pi/3) |(1/n) \textstyle\sum_{i=1}^n z_i| \; .
$$

Combining this result and that $\int_{\mathbb{R}} z^2 \mathrm{d}\mu(z) = 1/12$ we have $\mu^{\otimes n} \mathrm{S}_0[U] - U^\star \leq (\pi/(6\sqrt{3})) n^{-1/2}$, which concludes the proof.

■

Acknowledgments

This work is part of the research program MISTIC, supported by the French Agence Nationale pour la Recherche (ANR-19-CE40-0005). V. De Bortoli was also supported by EPSRC grant EP/R034710/1. We thank Francesca Crucinio for pointing to us the use of *thermodynamic barriers* and *kinetic barriers* in chemistry.

Organization of the appendix

In this supplementary material we derive technical lemmas and additional results. In particular, we gather the technical lemmas of Section 2 in Appendix A and the ones of Section 4.3.1 in Appendix B. In Appendix C, we recall basic results from differential geometry and geometric measure theory.

Appendix A. Technical results for Section 2

In this section, we derive some technical lemmas used in Section 4.1 in order to prove Theorem 3 and other results from Section 2. We recall that for any $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ and $\varepsilon > 0$, when this is well-defined, we set

$$\begin{aligned} &\mathcal{I}_\varepsilon(\varphi) = C_\varepsilon^{-1} \textstyle\int_{\mathbb{R}^d} \varphi(x)\Psi(x)\exp[-\left\|F(x)\right\|^k/\varepsilon]\mathrm{d}x \ , \quad \mathcal{J}_\varepsilon = \mathcal{I}_\varepsilon(1) \ , \\ &C_\varepsilon = \textstyle\int_{\mathbb{R}^d} \exp[-\left\|x\right\|^k/\varepsilon]\mathrm{d}x = \varepsilon^{d/k}\textstyle\int_{\mathbb{R}^d}\exp[-\left\|x\right\|^k]\mathrm{d}x = \varepsilon^{d/k}C_1 \ . \end{aligned}$$

In addition, we define

$$\mathcal{I}_0(\varphi) = \textstyle\int_{F^{-1}(0)} \varphi(x)\Psi(x)\mathrm{J}F(x)^{-1}\mathrm{d}\mathcal{H}^{d-\hat{d}}(x) \ , \quad \mathcal{J}_0 = \mathcal{I}_0(1) \ ,$$

with $\hat{d} = \min(d, p)$.

In Appendix A.1 we establish a link between a Hessian computed on the normal bundle of a manifold and the generalized Jacobian. In Appendix A.2 we derive technical truncation bounds for the proof of Theorem 3. Explicit controls of some derivative are presented in Appendix A.3 in order to derive Theorem 11. Finally, we present a smoothing lemma in Appendix A.4 which is key to weaken the regularity assumptions of Theorem 14.

A.1 From normal Hessian to generalized Jacobian

Let $f \in \mathrm{C}^2(\mathbb{R}^d)$ and M a manifold in $\mathbb{R}^d$. For any $x \in \mathrm{M}$ we define $\nabla^2_\perp f(x)$ to be the projection of the Hessian on the orthogonal of the tangent space of M at x, see Hwang (1980).

Lemma 23 *Let $U : \mathbb{R}^d \to \mathbb{R}$ and $F \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^p)$ such that for any $x \in \mathbb{R}^d$, $U(x) = \|F(x)\|^2$. In addition, assume that $F^{-1}(0) \neq \emptyset$ and that for any $x \in F^{-1}(0)$, $\mathrm{J}F(x) > 0$. Then, $\arg\min\{U(x) : x \in \mathbb{R}^d\}$ is a smooth manifold and for any $x \in F^{-1}(0)$ we have that $\det(\nabla^2_\perp U(x)) = \mathrm{J}F(x)^2$.*

Proof First, we have that $\arg\min\{U(x) : x \in \mathbb{R}^d\} = F^{-1}(0)$. Hence, $\arg\min\{U(x) : x \in \mathbb{R}^d\}$ is a smooth manifold since $F \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R}^p)$. Let $x \in F^{-1}(0)$. We have that $\nabla^2 U(x) = \mathrm{D}F(x)^\top \mathrm{D}F(x)$. Note that $\mathrm{D}F(x)^\top = (\nabla F_1(x), \dots, \nabla F_p(x))$ is a basis of $\ker(\mathrm{D}F(x))^\perp$, where we recall that $\ker(\mathrm{D}F(x))$ is the tangent space to $F^{-1}(0)$ at x. Denote by $O(x) = (f_1(x), \dots, f_p(x))$ the orthonormal basis of $\ker(\mathrm{D}F(x))^\perp$ obtained from $\mathrm{D}F(x)^\top$ using the Gram-Schmidt process. There exists a triangular $p \times p$ matrix $T(x)$ such that $O(x) = \mathrm{D}F(x)^\top T(x)$. We also have

$$\mathrm{Id} = O(x)^\top O(x) = T(x)^\top \mathrm{D}F(x)\mathrm{D}F(x)^\top T(x) \ .$$

Hence, we get that $\det(T(x)) = \mathrm{J}F(x)^{-1}$. We also have that

$$\begin{aligned}\det(\nabla^2_{\perp} U(x)) &= \det(O(x)^{\top}\nabla^2 U(x) O(x)) \\ &= \det(T(x)^{\top}\mathrm{D}F(x)\mathrm{D}F(x)^{\top}\mathrm{D}F(x)\mathrm{D}F(x)^{\top}T(x)) = \mathrm{J}F(x)^2 \ ,\end{aligned}$$

which concludes the proof. ■

A.2 Truncation and lower bounds

Lemma 24 *Assume* $\mathbf{H}1$ *and* $\mathbf{H}2$. *Then, for any* $\bar{\varepsilon} \geq 0$ *there exists* $A_0 \geq 0$ *such that for any* $\varepsilon \in [0, \bar{\varepsilon}]$, $\mathcal{J}_{\varepsilon} \geq A_0 m_{0,\Psi}$ *with* $m_{0,\Psi} = \inf\{\Psi(x) \ : \ x \in F^{-1}(0)\}$ *and* A_0 *that does not depend on* Ψ.

Proof Since $F(0) = 0$ and $F \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^p)$, there exists $M \geq 0$ such that for any $x \in \bar{\mathrm{B}}(0,1)$, $\|F(x)\| \leq M\|x\|$. Note that $F^{-1}(0)$ is compact since $\lim_{\|x\| \to +\infty} \|F(x)\| = +\infty$ and $F \in \mathrm{C}(\mathbb{R}^d, \mathbb{R}^p)$. Hence, since for any $x \in F^{-1}(0)$, $\Psi(x) > 0$ and $\Psi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R}_+)$ there exists $\eta \in (0,1)$ such that for any $x \in \mathrm{B}(0,\eta) \cup F^{-1}(0)$, $\Psi(x) \geq m_{0,\Psi}/2$. Using this result we have for any $\varepsilon > 0$

$$\begin{aligned}\mathcal{J}_{\varepsilon} &= \varepsilon^{-d/k} C_1^{-1} \textstyle\int_{\mathbb{R}^d} \Psi(x) \exp[-\|F(x)\|^k/\varepsilon]\mathrm{d}x \\ &\geq \varepsilon^{-d/k}(m_{0,\Psi}/2) C_1^{-1} \textstyle\int_{\mathrm{B}(0,\eta)} \exp[-M^k\|x\|^k/\varepsilon]\mathrm{d}x \\ &\geq C_1^{-1}(m_{0,\Psi}/2) M^{-d} \textstyle\int_{\bar{\mathrm{B}}(0, M\eta/\bar{\varepsilon}^{1/k})} \exp[-\|x\|^k]\mathrm{d}x \ .\end{aligned}$$

Using that $F^{-1}(0)$ is compact, there exists $\mathtt{M} \geq 0$ such that for any $x \in F^{-1}(0)$, $\mathrm{J}F(x) \leq \mathtt{M}$. Therefore, we get that

$$\mathcal{J}_0 = \textstyle\int_{F^{-1}(0)} \Psi(x) \mathrm{J}F(x)^{-1} \mathrm{d}\mathcal{H}^{d-\hat{d}}(x) \geq m_{0,\Psi}\mathtt{M}^{-1}\mathcal{H}^{d-\hat{d}}(F^{-1}(0)) \ .$$

Since $\mathcal{H}^{d-\hat{d}}(F^{-1}(0)) < +\infty$ using Theorem 40 in the case where $d \geq p$ and the fact that $\mathcal{H}^0(F^{-1}(0)) < +\infty$ if $d \leq p$ (see the first part of the proof of Theorem 26), we have that for any $\varepsilon \in [0, \bar{\varepsilon}]$

$$\begin{aligned}&\mathcal{J}_{\varepsilon} \geq A_0 m_{0,\Psi} \ , \qquad A_0 = \min(A_0^1, A_0^2) \ , \\ &A_0^1 = (1/2) C_1^{-1} M^{-d} \textstyle\int_{\bar{\mathrm{B}}(0, M\eta/\bar{\varepsilon}^{1/k})} \exp[-\|x\|^k]\mathrm{d}x \ , \quad A_0^2 = M^{-1}\mathcal{H}^{d-\hat{d}}(F^{-1}(0)) \ ,\end{aligned}$$

which concludes the proof. ■

Lemma 25 *Assume* $\mathbf{H}1$ *and* $\mathbf{H}2$. *Let* $\varphi: \ \mathbb{R}^d \to \mathbb{R}$ *and* $C_{\varphi} \geq 0$ *such that for any* $x \in \mathbb{R}^d$

$$|\varphi(x)| \leq C_{\varphi} \exp[C_{\varphi}\|x\|^{\alpha k}] \ . \tag{49}$$

Then, for any $\bar{\varepsilon} \in (0, \mathtt{m}^k/(1 + C_{\varphi,\Psi}))$ *and* $\mathsf{V} \subset \mathbb{R}^d$ *open and bounded such that* $F^{-1}(0) \subset \mathsf{V}$ *there exist* $\beta_1 > 0$ *and* $A_1 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ *such that for any* $\varepsilon \in (0, \bar{\varepsilon})$

$$\mathcal{I}_{\varepsilon}^{\mathrm{out}}(\varphi) \leq A_1(C_{\varphi,\Psi})\varepsilon^{-d/k}\exp[-\beta_1/\varepsilon] \ ,$$

with $\mathcal{I}_{\varepsilon}^{\mathrm{out}}(\varphi) = \mathcal{I}_{\varepsilon}(\varphi \mathbb{1}_{\mathsf{V}^{\mathrm{c}}})$, $C_{\varphi,\Psi} = C_{\varphi} + C_{\Psi} + C_{\varphi}C_{\Psi}$ *and* A_1, β_1 *functions that do not depend on* φ *and* Ψ. *Finally,* A_1 *is non-decreasing.*

Proof First using **H**2 and (49) there exists $C_{\varphi,\Psi}$ such that for any $x \in \mathbb{R}^d$

$$|\varphi(x)| \, \Psi(x) \le C_{\varphi,\Psi} \exp[C_{\varphi,\Psi} \|x\|] \ , \qquad C_{\varphi,\Psi} = C_\varphi + C_\Psi + C_\varphi C_\Psi \ .$$

Since V is bounded there exists $R' \ge R$ (where R is given in **H**1) such that $\mathsf{V} \subset \bar{\mathrm{B}}(0, R')$. Note that for any $\varepsilon > 0$, we have

$$\begin{aligned} &\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi) = \mathcal{I}_\varepsilon^1(\varphi) + \mathcal{I}_\varepsilon^2(\varphi) \ , \\ &\mathcal{I}_\varepsilon^1(\varphi) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\bar{\mathrm{B}}(0,R')^{\mathrm{c}}}) \ , \quad \mathcal{I}_\varepsilon^2(\varphi) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\mathsf{V}^{\mathrm{c}} \cap \bar{\mathrm{B}}(0,R')}) \ . \end{aligned}$$

Let $\varepsilon \in (0, \bar{\varepsilon})$, we divide the rest of the proof into two parts. First, we bound $\mathcal{I}_\varepsilon^1(\varphi)$ and then $\mathcal{I}_\varepsilon^2(\varphi)$.

(a) Let $u = (\mathtt{m}^k/\varepsilon - C_{\varphi,\Psi})^{1/\alpha k}$ (which makes sense, since $\varepsilon < \mathtt{m}^k/(C_{\varphi,\Psi} + 1)$). Since $R' \ge R$ we have using (7) and that $u \ge 1$

$$\begin{aligned} \mathcal{I}_\varepsilon^1(\varphi) &= C_1^{-1} \varepsilon^{-d/k} \int_{\bar{\mathrm{B}}(0,R')^{\mathrm{c}}} \varphi(x) \Psi(x) \exp[-\|F(x)\|^k/\varepsilon] \mathrm{d}x \\ &\le C_1^{-1} C_{\varphi,\Psi} \varepsilon^{-d/k} \int_{\bar{\mathrm{B}}(0,R')^{\mathrm{c}}} \exp[-(\mathtt{m}^k/\varepsilon - C_{\varphi,\Psi}) \|x\|^{\alpha k}] \mathrm{d}x \\ &\le C_1^{-1} C_{\varphi,\Psi} \varepsilon^{-d/k} \int_{\bar{\mathrm{B}}(0,R'u)^{\mathrm{c}}} \exp[-\|x\|^{\alpha k}] \mathrm{d}x \ . \end{aligned}$$

Let $C_\alpha = \int_{\mathbb{R}^d} \exp[-\|x\|^{\alpha k}] \mathrm{d}x$. Using that $u = (\mathtt{m}^k/\varepsilon - C_{\varphi,\Psi})^{1/\alpha k}$, we have

$$\begin{aligned} \mathcal{I}_\varepsilon^1(\varphi) &\le C_1^{-1} C_{\varphi,\Psi} \varepsilon^{-d/k} \int_{\bar{\mathrm{B}}(0,R'u)^{\mathrm{c}}} \exp[-\|x\|^{\alpha k}] \mathrm{d}x \\ &\le C_1^{-1} C_{\varphi,\Psi} \varepsilon^{-d/k} \int_{\mathbb{R}^d} \exp[\|x\|^{\alpha k}/2] \exp[-\|x\|^{\alpha k}] \mathrm{d}x \exp[-(R'u)^{\alpha k}/2] \\ &\le 2^{d/\alpha k} C_\alpha C_1^{-1} C_{\varphi,\Psi} \exp[(R')^{\alpha k} C_{\varphi,\Psi}/2] \varepsilon^{-d/k} \exp[-(R')^{\alpha k} \mathtt{m}^k/(2\varepsilon)] \le A_1^1 \varepsilon^{-d/k} \exp[-\beta_1^1/\varepsilon] \ , \end{aligned} \tag{50}$$

with

$$A_1^1 = 2^{d/\alpha k} C_\alpha C_1^{-1} C_{\varphi,\Psi} \exp[(R')^{\alpha k} C_{\varphi,\Psi}/2] \ , \qquad \beta_1^1 = (R')^{\alpha k} \mathtt{m}^k/2 \ .$$

(b) Second, note that $\mathsf{K} = \mathsf{V}^{\mathrm{c}} \cap \bar{\mathrm{B}}(0, R')$ is bounded and closed, *i.e.* K is compact. Note that for any $x \in \mathsf{K}$, $\|F(x)\| > 0$, hence there exists $m > 0$ such that for any $x \in \mathsf{K}$, $\|F(x)\| \ge m$. In addition, we have that for any $x \in \mathsf{K}$,

$$|\varphi(x)| \, \Psi(x) \le C_{\varphi,\Psi} \exp[C_{\varphi,\Psi} \|x\|^{\alpha k}] \le C_\varphi \exp[C_{\varphi,\Psi} (R')^{\alpha k}] \ .$$

Therefore, we have

$$\mathcal{I}_\varepsilon^2(\varphi) \le C_1^{-1} C_{\varphi,\Psi} \varepsilon^{-d/k} \exp[C_{\varphi,\Psi} (R')^{\alpha k}] \exp[-m^k/\varepsilon] \lambda(\mathsf{K}) \ ,$$

where we recall that $\lambda(\mathsf{K})$ is the Lebesgue measure of K. Since $\mathsf{K} \subset \bar{\mathrm{B}}(0, R')$ we have

$$\mathcal{I}_\varepsilon^2(\varphi) \le \pi^{d/2} (R')^d \Gamma^{-1}(d/2+1) C_1^{-1} C_{\varphi,\Psi} \exp[(R')^{\alpha k}] \varepsilon^{-d/k} \exp[-m^k/\varepsilon] \le A_1^2 \varepsilon^{-d/k} \exp[-\beta_1^2/\varepsilon] \ , \tag{51}$$

where $\Gamma : \ (0, +\infty) \to \mathbb{R}_+$ is given for any $s \in (0, +\infty)$ by $\Gamma(s) = \int_0^{+\infty} t^{s-1} \exp[-t] \mathrm{d}t$ and

$$A_1^2 = \pi^{d/2} (R')^d \Gamma^{-1}(d/2+1) C_1^{-1} C_{\varphi,\Psi} \exp[C_{\varphi,\Psi} (R')^{\alpha k}] \ , \qquad \beta_1^2 = m^k \ .$$

We conclude the proof upon combining (50), (51), letting $\beta_1 = \min(\beta_1^1, \beta_1^2)$ and $A_1 = A_1^1 + A_1^2$. ■

Lemma 26 *Assume* $\mathbf{H}1$ *and that* $d \leq p$*. Then there exist* $N \in \mathbb{N}$, $\{x_0^k\}_{k=1}^N \in (\mathbb{R}^d)^N$ *and* $\mathsf{W}_k \subset \mathbb{R}^d$ *open such that for any* $k \in \{1, \dots, N\}$, $x_0^k \in \mathsf{W}_k$, $F : \bar{\mathsf{W}}_k \to F(\bar{\mathsf{W}}_k)$ *is a bi-Lipschitz homeomorphism, for any* $x \in \mathsf{W}_k$, $\mathrm{d}F(x)$ *is injective and for any* $j \in \{1, \dots, N\}$, $\bar{\mathsf{W}}_k \cap \bar{\mathsf{W}}_j = \emptyset$*. In addition,* $F^{-1}(0) = \cup_{k=1}^N \{x_0^k\}$*.*

Proof Since, for any $x \in F^{-1}(0)$, $\mathrm{J}F(x) > 0$ and $d \leq p$ there exists $r_x > 0$ such that for any $y \in \bar{\mathrm{B}}(x, r_x)$, $F(y) = F(x)$ implies that $y = x$. Since $\lim_{\|x\| \to +\infty} \|F(x)\| = +\infty$ we have that $F^{-1}(0)$ is compact. Assume that $\mathcal{H}^0(F^{-1}(0)) = +\infty$. Then, there exists $(x_k)_{k \in \mathbb{N}}$ such that for any $k \in \mathbb{N}$, $F(x_k) = 0$ and for any $j \in \{0, \dots, k-1\}$, $x_j \neq x_k$. Up to taking a subsequence, there exists $x^\star \in F^{-1}(0)$ such that $\lim_{k \to +\infty} x_k = x^\star$ and for any $k \in \mathbb{N}$, $x_k \neq x^\star$. Hence, there exists $k \in \mathbb{N}$ such that $x_k \in \bar{\mathrm{B}}(0, r_{x^\star})$ which is absurd. Hence $\mathcal{H}^0(F^{-1}(0)) < +\infty$. In what follows, we let $N = \mathcal{H}^0(F^{-1}(0))$ and denote $\{x_0^k\}_{k=1}^N \in (\mathbb{R}^d)^N$ such that $F^{-1}(0) = \{x_0^k\}_{k=1}^N$. There exists $\{r'_k\}_{k=1}^N \in (\mathbb{R}_+)^N$ such that for any $k, j \in \{1, \dots, N\}$, $\bar{\mathrm{B}}(x_0^k, r'_k) \cap \bar{\mathrm{B}}(x_0^j, r'_j) = \emptyset$. For any $k \in \{1, \dots, N\}$, we let $\mathsf{V}_k = \mathrm{B}(x_0^k, \min(r'_k, r_{x_0^k}/2))$. Let $k \in \{1, \dots, N\}$. By construction, $F : \bar{\mathsf{V}}_k \to F(\bar{\mathsf{V}}_k)$ is bijective and continuous. Since $\bar{\mathsf{V}}_k$ is compact, we have that $F : \bar{\mathsf{V}}_k \to F(\bar{\mathsf{V}}_k)$ is a homeomorphism. Since for any $k \in \{1, \dots, N\}$, $\mathrm{J}F(x_0^k) > 0$, there exists $m > 0$ such that for any $k \in \{1, \dots, N\}$ and $v \in \mathbb{R}^d$, $v^\top H(x_0^k, x_0^k) v \geq m \|v\|^2$ with $H(x, y) = \mathrm{D}F(x)^\top \mathrm{D}F(y)$ for any $x, y \in \mathbb{R}^d$. For any $k \in \{1, \dots, N\}$, there exists $\mathsf{W}_k \subset \mathsf{V}_k$ such that for any $x, y \in \mathsf{W}_k$ we have $\|H(x, y) - H(x_0^k, x_0^k)\|_2 \leq m/2$. Therefore we have for any $x, y \in \mathsf{W}_k$

$$\begin{aligned}
\|F(x) - F(y)\|^2 &= \textstyle\int_0^1 \int_0^1 \langle \mathrm{D}F(x + t(y-x))(y-x), \mathrm{D}F(x + s(y-x))(y-x) \rangle \mathrm{d}t\mathrm{d}s \\
&= \textstyle\int_0^1 \int_0^1 (y-x)^\top H(x_s, x_t)(y-x) \mathrm{d}t\mathrm{d}s \\
&= (y-x)^\top H(x_0^k, x_0^k)(y-x) + \textstyle\int_0^1 \int_0^1 (y-x)^\top (H(x_t, x_s) - H(x_0^k, x_0^k))(y-x) \mathrm{d}t\mathrm{d}s \\
&\geq (m/2) \|y - x\|^2 \ ,
\end{aligned}$$

where $x_t = x + t(y-x)$, which concludes the proof. ■

A.3 Quantitative control of the derivative

Lemma 27 *Under the same assumptions as Theorem 11, there exist* $\eta > 0$ *and* $P \in \mathtt{Poly}(4, \mathbb{R}_+)$ *such that for* $x \in F^{-1}(0)$ *and* $t \in \mathrm{B}_\infty(0, \eta)$ *we have*

$$\begin{aligned}
\|\mathrm{D}_t \chi(t, x)\| \leq (1 + M_{0,\varphi} + M_{1,\varphi})(1 + M_{0,\Psi} + M_{1,\Psi}) \\
\times P(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \exp[P(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F})] \ ,
\end{aligned}$$

where we recall that χ *is defined in* (28) *and for any* $\ell \in \mathbb{N}$, $i \in \{0, \dots, \ell\}$ *and* $f \in \mathrm{C}^\ell(\mathbb{R}^d, \mathbb{R}^p)$, $M_{i,f} = \sup\{\|\mathrm{D}^i f(x)\| \ : \ x \in F^{-1}(\mathrm{B}_\infty(0, \eta))\}$*.*

Proof In this proof, for any $f : \mathbb{R}^{m_0} \to \mathbb{R}^{m_1}$ with $m_0, m_1 \in \mathbb{N}$ differentiable, we denote $\mathrm{d}f$ its differential. Recall that $\bar{\Phi}_t : \mathbb{R}^d \to \mathbb{R}^d$ is defined such that for any $x \in \mathbb{R}^d$, $\bar{\Phi}_t(x) = x^{(p)}$ with $x^{(0)} = x$ and for any $i \in \{0, \dots, p-1\}$, $x^{(i+1)} = \Phi_{i+1}(t_{i+1}, x^{(i)})$, with Φ_{i+1} given by (27). The compact sets K_0 and K_1 are defined in the proof of Theorem 11. Since for any $x \in F^{-1}(0)$, $|\det(\mathrm{D}\bar{\Phi}_t^{-1}(x))| > 0$, we assume without loss of generality that $\det(\mathrm{D}\bar{\Phi}_t^{-1}(x)) > 0$. For ease of notation we denote $\varphi_\Psi = \varphi \times \Psi$. We have

$$M_{0,\varphi_\Psi} = M_{0,\varphi} M_{0,\Psi} \ , \qquad M_{1,\varphi_\Psi} = M_{0,\varphi} M_{1,\Psi} + M_{1,\varphi} M_{0,\Psi} \ .$$

In addition, we have for any $x \in \mathbb{R}^d$ and $t \in \mathsf{K}_0$

$$\chi(t,x) = \varphi_\Psi(\bar{\Phi}_t^{-1}(x)) \mathrm{J}F(\bar{\Phi}_t^{-1}(x))^{-1} \det(\mathrm{D}\bar{\Phi}_t^{-1}(x)) \ . \tag{52}$$

We now control the first derivative of χ. We divide the rest of the proof in three steps.

(a) We start by providing upper bounds for $x \mapsto h_{i,j}(x)$, $x \mapsto \mathrm{D}h_{i,j}(x)$ and $x \mapsto \mathrm{D}^2 h_{i,j}(x)$ for any $i, j \in \{1, \dots, p\}$ where we recall that for any $x \in \mathsf{K}_1$ we have

$$\{h_{i,j}(x)\}_{1 \leq i,j \leq p} = G(x)^{-1} = \mathrm{Adj}(G(x)) / \det(G(x)) = \mathrm{Adj}(G(x)) / \mathrm{J}F(x)^2 \ ,$$

where $\mathrm{Adj}(G(x))$ is the adjugate of $G(x) = \{\langle \nabla F_i(x), \nabla F_j(x) \rangle\}_{1 \leq i,j \leq p}$, where for any $x \in \mathsf{K}_1$ and $i, j \in \{1, \dots, p\}$

$$\mathrm{Adj}(G(x)) = \{(-1)^{i+j} \det(G^{i,j}(x))\}_{1 \leq i,j \leq p} \ ,$$

where for any $i, j \in \{1, \dots, p\}$, $\det(G^{i,j}(x))$ is the (i,j) minor of $G(x)$. Hence, using the Cauchy-Schwarz inequality there exists $D_0 \geq 0$ such that for any $x \in \mathsf{K}_1$ and $i, j \in \{1, \dots, p\}$

$$|h_{i,j}(x)| \leq D_0 M_{1,F}^{2p-2} / m_{1,F}^2 \ , \tag{53}$$

where we define

$$M_{1,F} = \sup\{\|\nabla F_i(x)\| \ : \ x \in \mathsf{K}_1, \ i \in \{1, \dots, p\}\} \ , \quad m_{1,F} = \inf\{\mathrm{J}F(x) \ : \ x \in \mathsf{K}_1\} \ .$$

Recall that $m_{1,F} > 0$ since K_1 is compact and for any $x \in \mathsf{K}_1$, $\mathrm{J}F(x) > 0$. We have that for any $x \in \mathsf{K}_1$, $u \in \mathbb{R}^d$ and $i, j \in \{1, \dots, p\}$

$$\mathrm{d}h_{i,j}(x)(u) = e_i^\top \mathrm{d}G(x)^{-1}(u) e_j = -e_i^\top G(x)^{-1} \mathrm{d}G(x)(u) G(x)^{-1} e_j \ . \tag{54}$$

In addition, we have for any $x, u \in \mathbb{R}^d$ and $i, j \in \{1, \dots, p\}$

$$\mathrm{d}G_{i,j}(x)(u) = \mathrm{d}^2 F_i(x)(\nabla F_j(x), u) + \mathrm{d}^2 F_j(x)(\nabla F_i(x), u) \ . \tag{55}$$

Combining this result and (54) we get that there exists $C_1 \geq 0$ such that for any $x \in \mathsf{K}_1$ and $i, j \in \{1, \dots, p\}$

$$\|\mathrm{D}h_{i,j}(x)\| \leq C_1 M_{1,F}^{4p-3} M_{2,F} / m_{1,F}^4 \ , \tag{56}$$

where $M_{2,F} = \sup\{\|\nabla^2 F_i(x)\| \ : \ x \in \mathsf{K}_1, \ i \in \{1, \dots, p\}\}$. Hence using (26), (53) and (56), there exist $C_2, C_3 \geq 0$ such that for any $x \in \mathsf{K}_1$ and $i \in \{1, \dots, p\}$

$$\|g_i(x)\| \leq C_2 M_{1,F}^{2p-1} / m_{1,F}^2 \ , \quad \|\mathrm{D}g_i(x)\| \leq C_3 M_{2,F} / m_{1,F}^2 \{M_{1,F}^{4p-2} / m_{1,F}^2 + M_{1,F}^{2p-2}\} \ . \tag{57}$$

Similarly, for any $x \in \mathsf{K}_1$, $u, v \in \mathbb{R}^d$ and $i, j \in \{1, \dots, p\}$ we have

$$\begin{aligned} \mathrm{d}^2 h_{i,j}(x)(u,v) &= e_i^\top \mathrm{d}^2 G(x)^{-1}(u,v) e_j \\ &= e_i^\top \Big\{ G(x)^{-1}\mathrm{d}G(x)(u)G(x)^{-1}\mathrm{d}G(x)(v)G(x)^{-1} \\ &\quad + G(x)^{-1}\mathrm{d}G(x)(v)G(x)^{-1}\mathrm{d}G(x)(u)G(x)^{-1} \\ &\quad - G(x)^{-1}\mathrm{d}^2G(x)(u,v)G(x)^{-1} \Big\} e_j \ . \end{aligned}$$

We also have for any $x, u, v \in \mathbb{R}^d$ and $i, j \in \{1, \dots, p\}$

$$\begin{aligned} \mathrm{d}^2 G_{i,j}(x) &= \mathrm{d}^3 F_i(x)(\nabla F_j(x), u, v) + \mathrm{d}^3 F_j(x)(\nabla F_i(x), u, v) \\ &\quad + \mathrm{d}^2 F_i(x)(\mathrm{d}^2 F_j(x)(v), u) + \mathrm{d}^2 F_j(x)(\mathrm{d}^2 F_i(x)(v), u) \ . \end{aligned}$$

Hence there exist $P_3 \in \mathtt{Poly}(4, \mathbb{R}_+)$ such that for any $x \in \mathsf{K}_1$ we have

$$\|\mathrm{D}^2 h_{i,j}(x)\| \leq P_3(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \ , \tag{58}$$

where $M_{3,F} = \sup\{\|\nabla^3 F_i(x)\| \ : \ x \in \mathsf{K}_1, \ i \in \{1, \dots, p\}\}$. Next, note that for any $i \in \{1, \dots, p\}$, we can choose g_i such that $g_i \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R}^d)$ (and therefore $\Phi_i \in \mathrm{C}^{3,2}(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}^d)$ by Theorem 36). Combining this result, (57) and (58), there exist $P_1 \in \mathtt{Poly}(2, \mathbb{R}_+)$, $P_2 \in \mathtt{Poly}(3, \mathbb{R}_+)$ and $P_3 \in \mathtt{Poly}(4, \mathbb{R}_+)$ such that for any $x \in \mathsf{K}_1$ and $i \in \{1, \dots, p\}$

$$\begin{aligned} &\|g_i(x)\| \leq P_1(M_{1,F}, 1/m_{1,F}) \ , \qquad \|\mathrm{D}g_i(x)\| \leq P_2(M_{1,F}, M_{2,F}, 1/m_{1,F}) \ , \\ &\|\mathrm{D}^2 g_i(x)\| \leq P_3(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \ . \end{aligned} \tag{59}$$

(b) For any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have $|t_i| \leq 2\eta$, since $F_i(\Phi_i(t_i, x)) = F_i(x) - t_i$. Hence combining this result, the fact that for any $x \in \mathbb{R}^d$, $\Phi_i(0, x) = x$, (27) and (57), for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\|\Phi_i(t_i, x)\| \leq \|x\| + \textstyle\int_0^{t_i} \|\mathrm{D}_t \Phi_i(s, x)\| \, \mathrm{d}s \leq \|x\| + 2\eta C_2 M_{1,F}^{2p-1}/m_{1,F}^2 \ .$$

Therefore, there exists $C_4 \geq 0$ such that for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\|\Phi_i(t_i, x)\| \leq C_4(1 + M_{1,F}^{2p-1}/m_{1,F}^2) \ , \quad \|\mathrm{D}_t \Phi_i(t_i, x)\| \leq C_2 M_{1,F}^{2p-1}/m_{1,F}^2 \ . \tag{60}$$

Similarly, for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\begin{aligned} \|\mathrm{D}_x \Phi_i(t_i, x)\| &\leq \|\mathrm{Id}\| + \textstyle\int_0^{t_i} \|\mathrm{D}_{t,x} \Phi_i(s, x)\| \, \|\mathrm{D}_x \Phi_i(s, x)\| \, \mathrm{d}s \\ &\leq \|\mathrm{Id}\| + \textstyle\int_0^{t_i} \|\mathrm{D}g_i(\Phi_i(t_i, x))\| \, \|\mathrm{D}_x \Phi_i(s, x)\| \, \mathrm{d}s \\ &\leq \|\mathrm{Id}\| + C_3(M_{2,F}/m_{1,F}^2)\{M_{1,F}^{4p-2}/m_{1,F}^2 + M_{1,F}^{2p-2}\} \textstyle\int_0^{t_i} \|\mathrm{D}_x \Phi_i(t_i, x)\| \, \mathrm{d}s \ . \end{aligned}$$

Hence, using Grönwall's lemma, for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\|\mathrm{D}_x \Phi_i(t_i, x)\| \leq \|\mathrm{Id}\| \exp[2\eta C_3(M_{2,F}/m_{1,F}^2)\{M_{1,F}^{4p-2}/m_{1,F}^2 + M_{1,F}^{2p-2}\}] \ .$$

Therefore, there exists $C_5 \geq 0$ such that for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ with $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\|\mathrm{D}_x \Phi_i(t_i, x)\| \leq C_5 \exp[C_5 (M_{2,F}/m_{1,F}^2) \{M_{1,F}^{4p-2}/m_{1,F}^2 + M_{1,F}^{2p-2}\}] \ . \tag{61}$$

Using this result, (27) and (59), there exists $C_6 \geq 0$ such that for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ with $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\|\mathrm{D}_{t,x} \Phi_i(t_i, x)\| \leq C_6 \exp[C_6 (M_{2,F}/m_{1,F}^2) \{M_{1,F}^{4p-2}/m_{1,F}^2 + M_{1,F}^{2p-2}\}] \ . \tag{62}$$

Hence combining (60), (61) and (62), there exist $P_4 \in \mathtt{Poly}(2, \mathbb{R}_+)$ and $P_5 \in \mathtt{Poly}(3, \mathbb{R}_+)$ such that for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$, $t_i \in \mathbb{R}$ with $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\begin{aligned} &\|\mathrm{D}_t \Phi_i(t_i, x)\| + \|\Phi_i(t_i, x)\| \leq P_4(M_{1,F}, 1/m_{1,F}) \ , \qquad (63) \\ &\|\mathrm{D}_x \Phi_i(t_i, x)\| + \|\mathrm{D}_{t,x} \Phi_i(t_i, x)\| \leq P_5(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_5(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ . \end{aligned}$$

In addition, using (27) we have for any $i \in \{1, \dots, p\}$, $s \in \mathbb{R}$, $x \in \mathbb{R}^d$

$$\mathrm{D}_t \mathrm{D}_x^2 \Phi_i(s, x) = \mathrm{D} g_i(\Phi_i(s, x)) \mathrm{D}_x^2 \Phi_i(s, x) + \mathrm{D}^2 g_i(\Phi_i(s, x)) \mathrm{D}_x \Phi_i(s, x) \ .$$

Therefore, using (59) and (63), we have for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$

$$\begin{aligned} \|\mathrm{D}_x^2 \Phi_i(t_i, x)\| &\leq \textstyle\int_0^{t_i} \{\|\mathrm{D} g_i(\Phi_i(s, x))\| \|\mathrm{D}_x^2 \Phi_i(s, x)\| + \|\mathrm{D}^2 g_i(\Phi_i(s, x))\| \|\mathrm{D}_x \Phi_i(s, x)\|\} \mathrm{d}s \\ &\leq 2\eta P_3(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) P_5(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_5(M_{1,F}, M_{2,F}, 1/m_{1,F})] \\ &\qquad + P_2(M_{1,F}, M_{2,F}, 1/m_{1,F}) \textstyle\int_0^{t_i} \|\mathrm{D}_x^2 \Phi_i(s, x)\| \mathrm{d}s \ . \end{aligned}$$

Hence, using Grönwall's lemma, there exists $P_6 \in \mathtt{Poly}(4, \mathbb{R}_+)$ such that for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$ we have

$$\|\mathrm{D}_x^2 \Phi_i(t_i, x)\| \leq P_6(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \exp[P_6(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F})] \ .$$

Hence, combining this result and (63), we have for any $x \in \mathsf{K}_1$, $i \in \{1, \dots, p\}$ and $t_i \in \mathbb{R}$ such that $\Phi_i(t_i, x) \in \mathsf{K}_1$

$$\begin{aligned} &\|\mathrm{D}_t \Phi_i(t_i, x)\| + \|\Phi_i(t_i, x)\| \leq P_4(M_{1,F}, 1/m_{1,F}) \ , \\ &\|\mathrm{D}_x \Phi_i(t_i, x)\| + \|\mathrm{D}_{t,x} \Phi_i(t_i, x)\| \leq P_5(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_5(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ , \\ &\|\mathrm{D}_x^2 \Phi_i(t_i, x)\| \leq P_6(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \exp[P_6(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F})] \ . \qquad (64) \end{aligned}$$

(c) In what follows, we fix $t \in \mathrm{B}_\infty(0, \eta)$ and use (64) to provide uniform bounds for $\mathrm{D}_t \bar{\Phi}_t^{-1}$, $\mathrm{D}_t \mathrm{D} \bar{\Phi}_t^{-1}$ and $\mathrm{D} \bar{\Phi}_t^{-1}$ on $F^{-1}(0)$. We introduce $\{\bar{\Phi}_{i,t}^{-1}\}_{i=1}^p$ such that for any $x \in F^{-1}(0)$ and $i \in \{1, \dots, p\}$

$$\bar{\Phi}_{t,i}^{-1}(x) = \Phi_i(-t_i, \bar{\Phi}_{t,i+1}^{-1}(x)) \ , \qquad \bar{\Phi}_{t,p+1}^{-1}(x) = x \ .$$

Note that $\bar{\Phi}_t^{-1} = \bar{\Phi}_{t,1}^{-1}$. Let $j \in \{1, \dots, p\}$. For any $i \in \{1, \dots, p\}$ we distinguish three cases:

(i) $i > j$, then for any $x \in F^{-1}(0)$, $\mathrm{D}_{t_j} \bar{\Phi}_{t,i}^{-1}(x) = 0$.

(ii) $i = j$, then for any $x \in F^{-1}(0)$, $\mathrm{D}_{t_j} \bar{\Phi}_{t,i}^{-1}(x) = -\mathrm{D}_t \Phi_j(-t_j, \bar{\Phi}_{t,j+1}^{-1}(x))$.

(iii) $i < j$, then for any $x \in F^{-1}(0)$, $\mathrm{D}_{t_j} \bar{\Phi}_{t,i}^{-1}(x) = \mathrm{D}_x \Phi_i(-t_i, \bar{\Phi}_{t,i+1}^{-1}(x)) \mathrm{D}_{t_j} \bar{\Phi}_{t,i+1}^{-1}(x)$.

Combining these results, (64) and the fact that $\bar{\Phi}_t^{-1} = \bar{\Phi}_{t,1}^{-1}$ we get that there exists $\bar{P}_7 \in$ `Poly`$(3, \mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$, and $j \in \{1, \dots, p\}$ we have

$$\|\mathrm{D}_{t_j} \bar{\Phi}_t^{-1}(x)\| \leq \bar{P}_7(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[\bar{P}_7(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ .$$

Hence, there exists $P_7 \in$ `Poly`$(3, \mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$ we have

$$\|\mathrm{D}_t \bar{\Phi}_t^{-1}(x)\| \leq P_7(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_7(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ . \tag{65}$$

Next, note that for any $x \in F^{-1}(0)$, $t \in \mathrm{B}_\infty(0, \eta)$ and $i \in \{1, \dots, p\}$ we have

$$\mathrm{D}\bar{\Phi}_{t,i}^{-1}(x) = \mathrm{D}_x \Phi_i(-t_i, \bar{\Phi}_{t,i+1}^{-1}(x)) \mathrm{D}\bar{\Phi}_{t,i+1}^{-1}(x) \ , \qquad \mathrm{D}\bar{\Phi}_{t,p+1}^{-1}(x) = \mathrm{Id} \ .$$

Combining this result and (64), there exists $P_8 \in$ `Poly`$(3, \mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$ and $i \in \{1, \dots, p\}$

$$\|\mathrm{D}\bar{\Phi}_{t,i}^{-1}(x)\| \leq P_8(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_8(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ . \tag{66}$$

In particular, since $\bar{\Phi}_t = \bar{\Phi}_{t,1}$ we have for any $x \in F^{-1}(0)$

$$\|\mathrm{D}\bar{\Phi}_t^{-1}(x)\| \leq P_8(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_8(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ . \tag{67}$$

Finally, we give a uniform upper-bound on $\mathrm{D}_t \mathrm{D}\bar{\Phi}_t^{-1}$. Let $j \in \{1, \dots, p\}$. For any $i \in \{1, \dots, p\}$ we distinguish three cases:

(i) $i > j$, then for any $x \in F^{-1}(0)$, $\mathrm{D}_{t_j} \mathrm{D}\bar{\Phi}_{t,i}^{-1}(x) = 0$.

(ii) $i = j$, then for any $x \in F^{-1}(0)$, $\mathrm{D}_{t_j} \mathrm{D}\bar{\Phi}_{t,i}^{-1}(x) = -\mathrm{D}_{t,x} \Phi_j(-t_j, \bar{\Phi}_{t,j+1}^{-1}(x)) \mathrm{D}\bar{\Phi}_{t,j+1}^{-1}(x)$.

(iii) $i < j$, then for any $x \in F^{-1}(0)$, we have

$$\begin{aligned} \mathrm{D}_{t_j} \mathrm{D}\bar{\Phi}_{t,i}^{-1}(x) &= \mathrm{D}_x \Phi_i(-t_i, \bar{\Phi}_{t,i+1}^{-1}(x)) \mathrm{D}_{t_j} \mathrm{D}\bar{\Phi}_{t,i+1}^{-1}(x) \\ &\quad + \mathrm{D}_x^2 \Phi_i(-t_i, \bar{\Phi}_{t,i+1}^{-1}(x))(\mathrm{D}\bar{\Phi}_{t,i+1}^{-1}(x), \mathrm{D}_{t_j} \mathrm{D}\bar{\Phi}_{t,i+1}^{-1}(x)) \ . \end{aligned}$$

Using (64) and (66), there exists $P_9 \in$ `Poly`$(4, \mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$ we have

$$\|\mathrm{D}_t \mathrm{D}\bar{\Phi}_t^{-1}(x)\| \leq P_9(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \exp[P_9(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F})] \ . \tag{68}$$

Therefore, summarizing (65), (67) and (68), there exist $P_7, P_8 \in$ `Poly`$(\mathbb{R}_+, 3)$ and $P_9 \in$ `Poly`$(4, \mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$ and $t \in \mathrm{B}_\infty(0, \eta)$ we have

$$\begin{aligned} &\|\mathrm{D}_t \bar{\Phi}_t^{-1}(x)\| \leq P_7(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_7(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ , \\ &\|\mathrm{D}\bar{\Phi}_t^{-1}(x)\| \leq P_8(M_{1,F}, M_{2,F}, 1/m_{1,F}) \exp[P_8(M_{1,F}, M_{2,F}, 1/m_{1,F})] \ , \\ &\|\mathrm{D}_t \mathrm{D}\bar{\Phi}_t^{-1}(x)\| \leq P_9(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F}) \exp[P_9(M_{1,F}, M_{2,F}, M_{3,F}, 1/m_{1,F})] \ . \end{aligned} \tag{69}$$

(d) Next, we use (69) to conclude the proof by providing uniform upper-bounds on the differential χ on $F^{-1}(0)$ w.r.t. t, where χ given in (52). For any $x \in F^{-1}(0)$ and $t \in \mathrm{B}_\infty(0,\eta)$ we have

$$\begin{aligned} \mathrm{D}_t\chi(t,x) = &\mathrm{D}_t\varphi_\Psi(\bar{\Phi}_t^{-1}(x))\mathrm{J}F(\bar{\Phi}_t^{-1}(t))^{-1}\det(G(\bar{\Phi}_t^{-1}(x))) \\ &+ \varphi_\Psi(\bar{\Phi}_t^{-1}(x))\mathrm{D}_t\mathrm{J}F(\bar{\Phi}_t^{-1}(t))^{-1}\det(G(\bar{\Phi}_t^{-1}(x))) \\ &+ \varphi_\Psi(\bar{\Phi}_t^{-1}(x))\mathrm{J}F(\bar{\Phi}_t^{-1}(t))^{-1}\mathrm{D}_t\det(G(\bar{\Phi}_t^{-1}(x))) \ . \end{aligned} \tag{70}$$

First, for any $x \in F^{-1}(0)$ and $t \in \mathrm{B}_\infty(0,\eta)$ we have

$$|\varphi_\Psi(\bar{\Phi}_t^{-1}(x))| \leq M_{0,\varphi_\Psi} \ , \qquad \mathrm{J}F(\bar{\Phi}_t^{-1}(x))^{-1} \leq 1/m_{1,F} \ . \tag{71}$$

In addition, using (69), there exists $P_{10} \in \texttt{Poly}(3,\mathbb{R}_+)$ such that for any $x \in \mathsf{K}_1$ and $t \in \mathrm{B}_\infty(0,\eta)$

$$|\det(\mathrm{D}\bar{\Phi}_t^{-1}(x))| \leq P_{10}(M_{1,F},M_{2,F},1/m_{1,F})\exp[P_{10}(M_{1,F},M_{2,F},1/m_{1,F})] \ . \tag{72}$$

For any $x \in F^{-1}(0)$ and $t \in \mathrm{B}_\infty(0,\eta)$ we have

$$\mathrm{D}_t\varphi_\Psi(\bar{\Phi}_t^{-1}(x)) = \mathrm{D}\varphi_\Psi(\bar{\Phi}_t^{-1}(x))\mathrm{D}_t\bar{\Phi}_t^{-1}(x) \ .$$

Combining this result and (69), there exists $P_{11} \in \texttt{Poly}(3,\mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$ and $t \in \mathrm{B}_\infty(0,\eta)$ we have

$$\|\mathrm{D}_t\varphi(\bar{\Phi}_t^{-1}(x))\| \leq M_{1,\varphi_\Psi}P_{11}(,M_{1,F},M_{2,F},1/m_{1,F})\exp[P_{11}(,M_{1,F},M_{2,F},1/m_{1,F})] \ . \tag{73}$$

For any $x \in F^{-1}(0)$, $t \in \mathrm{B}_\infty(0,\eta)$ we have

$$\mathrm{D}_t\mathrm{J}F(\bar{\Phi}_t^{-1}(x))^{-1} = \mathrm{D}_t\det(G(\bar{\Phi}_t^{-1}(x)))^{1/2} = \mathrm{J}F(\bar{\Phi}_t^{-1}(x))^{-1}\mathrm{D}_t\det(G(\bar{\Phi}_t^{-1}(x))^{-1}) \ . \tag{74}$$

In addition, for any $x \in F^{-1}(0)$, $t \in \mathrm{B}_\infty(0,\eta)$ and $h \in \mathbb{R}^p$ we have

$$\mathrm{D}_t\det(G(\bar{\Phi}_t^{-1}(x)))(h) = \mathrm{Tr}(\mathrm{Adj}(G(\bar{\Phi}_t^{-1}(x)))\mathrm{D}G(\bar{\Phi}_t^{-1}(x))(\mathrm{D}_t\bar{\Phi}_t^{-1}(x)(h))) \ .$$

Using this result, (74), (55) and (69), there exists $P_{12} \in \texttt{Poly}(3,\mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$, $t \in \mathrm{B}_\infty(0,\eta)$ we have

$$\|\mathrm{D}_t\mathrm{J}F(\bar{\Phi}_t^{-1}(x))^{-1}\| \leq P_{12}(M_{1,F},M_{2,F},1/m_{1,F})\exp[P_{12}(M_{1,F},M_{2,F},1/m_{1,F})] \ . \tag{75}$$

Finally, for any $x \in F^{-1}(0)$, $t \in \mathrm{B}_\infty(0,\eta)$ and $h \in \mathbb{R}^p$ we have

$$\mathrm{D}_t\det(\mathrm{D}\bar{\Phi}_t^{-1}(x)) = \mathrm{Tr}(\mathrm{Adj}(\mathrm{D}\bar{\Phi}_t^{-1}(x))\mathrm{D}_t\mathrm{D}\bar{\Phi}_t^{-1}(x)(h)) \ .$$

Hence, using (69), there exists $P_{13} \in \texttt{Poly}(4,\mathbb{R}_+)$ such that for any $x \in F^{-1}(0)$ and $t \in \mathrm{B}_\infty(0,\eta)$ we have

$$\begin{aligned} \|\mathrm{D}_t\det(\mathrm{D}\bar{\Phi}_t^{-1}(x))\| \leq &P_{13}(M_{1,F},M_{2,F},M_{3,F},1/m_{1,F}) \\ &\times\exp[P_{13}(M_{1,F},M_{2,F},M_{3,F},1/m_{1,F})] \ . \end{aligned} \tag{76}$$

We conclude the proof upon combining (70), (71), (72), (73), (75) and (76). ■

A.4 Regularity results

In the following section we prove a smoothing lemma which is key to extend Theorem 14 to the case where Ψ and φ are no longer in $\mathrm{C}^1(\bar{\mathsf{U}}, \mathbb{R})$ but are Lipschitz continuous.

Lemma 28 *Let U be open bounded, $r > 0$, $\mathsf{V} = \mathsf{U} + \mathrm{B}(0, r)$ and $\varphi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R})$ which satisfies (7) and is Lipschitz on V, i.e. there exists $M_{1,\varphi} \geq 0$ such that for any $x, y \in \mathsf{V}$ we have*

$$|\varphi(x) - \varphi(y)| \leq M_{1,\varphi} \|x - y\| \ .$$

In addition, let $M_{0,\varphi} = \sup\{|\varphi(x)| \ : \ x \in \mathsf{V}\}$. Then there exist $\bar{\delta} > 0$ and $\{\varphi^\delta \ : \ \delta \in (0, \bar{\delta})\}$ such that the following hold:

(a) For any $x \in \mathsf{U}$, $\lim_{\delta \to 0} \varphi^\delta(x) = \varphi(x)$.

(b) For any $\delta \in (0, \bar{\delta})$, $\varphi^\delta \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$ and there exists $\mathtt{L}_\delta \geq 0$ such that for any $x, y \in \mathsf{U}$,

$$\left|\varphi^\delta(x) - \varphi^\delta(y)\right| \leq \mathtt{L}_\delta \|x - y\| \ .$$

Let $\mathtt{L}_0 = \sup\{\mathtt{L}_\delta \ : \ \delta \in (0, \bar{\delta})\} < +\infty$ and we have $\mathtt{L}_0 \leq C_1(1 + M_{0,\varphi} + M_{1,\varphi})(1 + C_\varphi)$, with $C_1 \geq 0$ that does not depend on φ.

(c) For any $\delta \in (0, \bar{\delta})$ there exists $M_\delta \geq 0$ such that for any $x \in \mathsf{U}$, $|\varphi^\delta(x)| \leq M_\delta$ and $M_0 = \sup\{M_\delta \ : \ \delta \in (0, \bar{\delta})\} < +\infty$. In addition, $M_0 \leq C_2(1 + M_{0,\varphi})(1 + C_\varphi)$ with $C_2 \geq 0$ that does not depend on φ.

(d) For any $\delta \in (0, \bar{\delta})$ there exists $D_\delta \geq 0$ such that for any $x \in \mathbb{R}^d$ we have

$$|\varphi^\delta(x)| \leq D_\delta \exp[D_\delta \|x\|^{\alpha k}] \ ,$$

and $D_\delta \leq C_3(1 + C_\varphi) \exp[C_3 C_\varphi]$ with C_3 that does not depend on φ.

(e) Assume **H**1*, that $F^{-1}(0) \subset \mathsf{U}$ and that there exists $m_{0,\varphi} > 0$ such that for any $x \in F^{-1}(0)$, $\varphi(x) \geq m_{0,\varphi}$. Then there exists $\bar{\delta}^\star > 0$ such that for any $\delta \in (0, \bar{\delta}^\star)$ and $x \in F^{-1}(0)$, $\varphi^\delta(x) \geq m_{0,\varphi}/2$. In addition if $\varphi \geq 0$ then for any $\delta > 0$, $\varphi^\delta \geq 0$.*

In addition, $\bar{\delta} = f_1(C_\varphi)$ and $\bar{\delta}^\star = f_2(M_{0,\varphi}, M_{1,\varphi}, 1/m_{0,\varphi}, C_\varphi)$ with $f_1 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ and $f_2 \in \mathrm{C}(\mathbb{R}_+^4, \mathbb{R}_+)$ and are non-decreasing w.r.t. to each of their variables. Finally, there exists $C_4 \geq 0$ such that

$$\limsup_{\delta \to 0} M_\delta \leq C_4 M_{0,\varphi} \ , \quad \limsup_{\delta \to 0} \mathtt{L}_\delta \leq C_4 M_{1,\varphi} \ , \quad \limsup_{\delta \to 0} D_\delta \leq C_4 C_\varphi \ ,$$

with C_4 that does not depend on φ.

Proof Since U is bounded there exists $R \geq 0$ such that $\mathsf{U} \subset \bar{\mathrm{B}}(0, R)$. Let $p \in \mathbb{N}^\star$ such that $2p > \alpha k$ and for any $\delta > 0$ define $k_\delta : \ \mathbb{R}^d \to \mathbb{R}_+$ such that for any $x \in \mathbb{R}^d$ we have

$$k_\delta(x) = \exp[-\|x\|^{2p}/\delta] / \textstyle\int_{\mathbb{R}^d} \exp[-\|\tilde{x}\|^{2p}/\delta] \mathrm{d}\tilde{x} \ .$$

Let $C_1 = \int_{\mathbb{R}^d} \exp[-\|\tilde{x}\|^{2p}]\mathrm{d}\tilde{x}$ and we have for any $x \in \mathbb{R}^d$, $k_\delta(x) = C_1^{-1}\delta^{-d/2p}\exp[-\|x\|^{2p}/\delta]$. Note that since $2p > \alpha k$ and using (7) we have that for any $x \in \mathbb{R}^d$ $\int_{\mathbb{R}^d} |\varphi(x-y)|k_\delta(y)\mathrm{d}y < +\infty$. For any $\delta > 0$ we define φ^δ such that for any $x \in \mathbb{R}^d$

$$\varphi^\delta(x) = \int_{\mathbb{R}^d} \varphi(x-y)k_\delta(y)\mathrm{d}y \ . \tag{77}$$

We divide the rest of the proof into five parts.

(a) We have that for any $x \in \mathbb{R}^d$, $\lim_{\delta \to 0} \varphi^\delta(x) = \varphi(x)$, since $\varphi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R})$ and $\{k_\delta \ : \ \delta > 0\}$ is a mollifier. This concludes the proof of Theorem 28-(a).

(b) Using that $2p > \alpha k$ and (7) we have that for any $\beta \in \mathbb{N}$, $\int_{\mathbb{R}^d} \|y\|^\beta |\varphi(x-y)|k_\delta(y) < +\infty$. Hence, $\varphi^\delta \in \mathrm{C}^\infty(\mathbb{R}^d, \mathbb{R})$. Let $x \in \mathsf{U}$. Using that $\int_{\mathbb{R}^d} y\|y\|^{2p-1}k_\delta(y)\mathrm{d}y = 0$, we have that for any $\delta > 0$

$$\begin{aligned}
&\|\nabla\varphi^\delta(x)\| = (2p/\delta)\|\textstyle\int_{\mathbb{R}^d} \varphi(x-y)y\|y\|^{2p-2}k_\delta(y)\mathrm{d}y\| \\
&\quad = (2p/\delta)\textstyle\int_{\mathbb{R}^d} |\varphi(x-y) - \varphi(x)|\|y\|^{2p-1}k_\delta(y)\mathrm{d}y \\
&\quad \leq (2pM_{1,\varphi}/\delta)\textstyle\int_{\mathbb{R}^d} \|y\|^{2p}k_\delta(y)\mathrm{d}y + (2p/\delta)\int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(x-y) - \varphi(x)|\|y\|^{2p-1}k_\delta(y)\mathrm{d}y \\
&\quad \leq (2pM_{1,\varphi}/C_1)\textstyle\int_{\mathbb{R}^d} \|y\|^{2p}\exp[-\|y\|^{2p}]\mathrm{d}y \\
&\qquad\qquad + (2p/\delta)\textstyle\int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(x-y) - \varphi(x)|\|y\|^{2p-1}k_\delta(y)\mathrm{d}y \ .
\end{aligned} \tag{78}$$

We now bound the second term. Let $\bar{\alpha}_k = \lceil \alpha k \rceil$. Using (7), we have for any and $x \in \mathsf{U}$ and $y \in \mathbb{R}^d$

$$\begin{aligned}
|\varphi(x-y) - \varphi(x)| &\leq (1 + M_{0,\varphi})C_\varphi \exp[C_\varphi \|x-y\|^{\alpha k}] \\
&\leq (1 + M_{0,\varphi})C_\varphi \exp[3^{\bar{\alpha}_k - 1}C_\varphi(1 + \|x\|^{\bar{\alpha}_k} + \|y\|^{\bar{\alpha}_k})] \\
&\leq (1 + M_{0,\varphi})C_\varphi \exp[3^{\bar{\alpha}_k - 1}C_\varphi(1 + R^{\bar{\alpha}_k})]\exp[3^{\bar{\alpha}_k - 1}C_\varphi\|y\|^{\bar{\alpha}_k}] \ .
\end{aligned}$$

Therefore we get that for any $x \in \mathsf{U}$ and $y \in \bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}$

$$\begin{aligned}
&|\varphi(x-y) - \varphi(x)| \, \|y\|^{2p-1} \\
&\qquad \leq (1 + M_{0,\varphi})C_\varphi \exp[3^{\bar{\alpha}_k - 1}C_\varphi(1 + R^{\bar{\alpha}_k})]\exp[(3^{\bar{\alpha}_k}(r/2)^{\bar{\alpha}_k - 2p}C_\varphi + 1)\, \|y\|^{2p}] \ .
\end{aligned}$$

Therefore there exists $C_a \geq 0$ such that for any $x \in \mathsf{U}$ and $y \in \mathbb{R}^d$ we have

$$|\varphi(x-y) - \varphi(x)| \, \|y\|^{2p-1} \leq C_a(1 + M_{0,\varphi})C_\varphi \exp[C_a(1 + C_\varphi)\, \|y\|^{2p}] \ ,$$

with $C_a \geq 0$ that does not depend on φ. Using this result, we have for any $x \in \mathsf{U}$ and $\delta \in (0, 1/(2C_a(1 + C_\varphi)))$

$$\begin{aligned}
&(2p/\delta)\textstyle\int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(x-y) - \varphi(x)| \, \|y\|^{2p-1} \, k_\delta(y)\mathrm{d}y \\
&\qquad \leq 2pC_a(1 + M_{0,\varphi})C_\varphi\delta^{-d/2p-1}C_1^{-1}\textstyle\int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} \exp[(C_a(1 + C_\varphi) - 1/\delta)\, \|y\|^{2p}]\mathrm{d}y \\
&\qquad \leq 2^{1+d(1+2p)/2p}pC_a(1 + M_{0,\varphi})C_\varphi/(\delta C_1)\textstyle\int_{\bar{\mathrm{B}}(0,r/(2^{1+2p}\delta)^{1/2p})^{\mathrm{c}}} \exp[-\|y\|^{2p}]\mathrm{d}y \ .
\end{aligned}$$

We have for any $\delta \in (0, 1/(2C_a(1 + C_\varphi)))$

$$\int_{\bar{\mathrm{B}}(0,r/(2^{1+2p}\delta)^{1/2p})^{\mathrm{c}}} \exp[-\|y\|^{2p}]\mathrm{d}y/C_1 \leq \exp[-r^{2p}/(2^{2+2p}\delta)]\int_{\mathbb{R}^d} \exp[-\|y\|^{2p}/2]\mathrm{d}y/C_1 \ .$$

Therefore, there exists $C_b \geq 0$ (that does not depend on φ) such that for any $x \in \mathsf{U}$ and $\delta \in (0, 1/(2C_a(1+C_\varphi)))$ we have

$$(2p/\delta) \int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(x-y) - \varphi(x)| \, \|y\|^{2p-1} \, k_\delta(y)\mathrm{d}y \leq C_b(1+M_{0,\varphi})C_\varphi \delta^{-1} \exp[-r^{2p}/(2^{2+2p}\delta)] \ . \tag{79}$$

Combining this bound and (78), we get that for any $\delta \in (0, 1/(2C_a(1+C_\varphi)))$ there exists $\mathtt{L}_\delta \geq 0$ such that for any $x \in \mathsf{U}$, $\|\nabla \varphi^\delta(x)\| \leq \mathtt{L}_\delta$. Let $\mathtt{L}_0 = \sup\{\mathtt{L}_\delta \, : \, \delta \in (0, \bar{\delta})\} < +\infty$ and using, (78), (79) and that for any $t \geq 0$, $t\exp[-r^{2p}t/2^{2+2p}] \leq 2^{2+2p}/(\mathrm{e}r^{2p})$ we have

$$\mathtt{L}_0 \leq (2pM_{1,\varphi}/C_1) \int_{\mathbb{R}^d} \|y\|^{2p} \exp[-\|y\|^{2p}]\mathrm{d}y + 2^{2+2p}C_b(1+M_{0,\varphi})C_\varphi/(\mathrm{e}r^{2p}) \ .$$

This concludes the proof of Theorem 28-(b).

(c) For any $x \in \mathsf{U}$ and $\delta \in (1/(2C_a(1+C_\varphi)))$ we have

$$\begin{aligned} |\varphi(x)| &\leq \int_{\bar{\mathrm{B}}(0,r/2)} |\varphi(y-x)| k_\delta(y)\mathrm{d}y + \int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(y-x)| k_\delta(y)\mathrm{d}y \\ &\leq M_{0,\varphi} + \int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(y-x)| k_\delta(y)\mathrm{d}y \ . \end{aligned} \tag{80}$$

Similarly to (79), there exists $c \geq 0$ (that does not depend on φ) such that for $x \in \mathsf{U}$ and $\delta \in (1/(2C_a(1+C_\varphi)))$

$$\int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(y-x)| k_\delta(y)\mathrm{d}y \leq (C_\varphi/c)\exp[-c/\delta] \ .$$

Combining this result and (80) for any $\delta \in (0, 1/(2C_a(1+C_\varphi)))$ there exists $M_\delta > 0$ such that for any $x \in \mathsf{U}$, $|\varphi^\delta(x)| \leq M_\delta$. Let $M_0 = \sup\{M_\delta \, : \, \delta \in (0, \bar{\delta})\} < +\infty$. We have that

$$M_0 \leq M_{0,\varphi} + (C_\varphi/c)\exp[-c(1+C_\varphi)] \ ,$$

which concludes the proof of Theorem 28-(c).

(d) Using that for any $a, b \geq 0$ and $p \geq 0$, $(a+b)^p \leq 2^{\min(p-1,0)}(a^p + b^p)$ we have that for any $x \in \mathbb{R}^d$ and $\delta > 0$

$$|\varphi^\delta(x)| \leq C_\varphi \int_{\mathbb{R}^d} \exp[C_\varphi \beta_{\alpha,k}(\|x\|^{\alpha k} + \|y\|^{\alpha k})] k_\delta(y)\mathrm{d}y \ ,$$

where $\beta_{\alpha,k} = 2^{\min(\alpha k - 1, 0)}$. Hence, using this result we have for any $\delta \in (0, 1/(2\beta_{\alpha,k}C_\varphi))$

$$\begin{aligned} |\varphi^\delta(x)| &\leq C_\varphi/(C_1\delta^{d/2p})\exp[C_\varphi\beta_{\alpha,k}\|x\|^{\alpha k}] \int_{\mathbb{R}^d} \exp[(C_\varphi\beta_{\alpha,k} - 1/\delta)\|y\|^{2p}]\mathrm{d}y \\ &\leq C_\varphi/(C_1\delta^{d/2p})\exp[C_\varphi\beta_{\alpha,k}\|x\|^{\alpha k}] \int_{\mathbb{R}^d} \exp[-\|y\|^{2p}/(2\delta)]\mathrm{d}y \\ &\leq 2^{d/2p}C_\varphi \exp[C_\varphi\beta_{\alpha,k}\|x\|^{\alpha k}] \ . \end{aligned}$$

Therefore, for any $\delta \in (0, 1/(2\beta_{\alpha,k}C_\varphi))$ there exists $D_\delta \geq 0$ such that for any $x \in \mathbb{R}^d$, $\varphi(x) \leq D_\delta \exp[D_\delta \|x\|^{\alpha k}]$. In addition, there exists $C_d \geq 0$ (that does not depend on φ) such that $D_\delta \leq C_d(1+C_\varphi)\exp[C_dC_\varphi]$, which concludes the proof of Theorem 28-(d).

(e) If $\varphi \geq 0$ then for any $\delta > 0$, $\varphi^\delta \geq 0$ using (77). For any $x \in \mathsf{U}$ we have

$$\varphi^\delta(x) = \varphi(x) + \int_{\mathbb{R}^d}(\varphi(y-x) - \varphi(x))k_\delta(y)\mathrm{d}y \geq m_{0,\varphi} - \int_{\mathbb{R}^d} |\varphi(y-x) - \varphi(x)| \, k_\delta(y) \ . \tag{81}$$

For any $x \in \mathsf{U}$ we have

$$\int_{\mathbb{R}^d} |\varphi(y-x) - \varphi(x)| \, k_\delta(y) \leq M_{1,\varphi} \delta^{1/k} + \int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(y-x) - \varphi(x)| k_\delta(y) \mathrm{d}y \; .$$

Similarly to (79), there exists $C_d \geq 0$ that does not depend on φ and such that for any $x \in \mathsf{U}$

$$\int_{\bar{\mathrm{B}}(0,r/2)^{\mathrm{c}}} |\varphi(y-x) - \varphi(x)| k_\delta(y) \mathrm{d}y \leq C_d (1 + M_{0,\varphi}) C_\varphi \exp[-1/(C_d \delta)] \; .$$

Let $\bar{\delta}_1 = (m_{0,\varphi}/(4M_{1,\varphi}))^k$ and $\bar{\delta}_2 = C_d^{-1}(\log(4 C_d C_\varphi (1 + M_{0,\varphi})) - \min(0, \log(m_{0,\varphi})))^{-1}$. Then for any $\delta \in (0, \min(\bar{\delta}_1, \bar{\delta}_2))$ we have that for any $x \in \mathsf{U}$

$$\int_{\mathbb{R}^d} |\varphi(y-x) - \varphi(x)| \, k_\delta(y) \leq m_{0,\varphi}/2 \; ,$$

which concludes the proof of Theorem 28-(e) upon combining this result with (81).

■

Appendix B. Technical results for Section 4.3.1

In this section, we derive a quantitative parametric theory for Laplace-type expansion in Appendix B.1 and Appendix B.2. We start by deriving technical bounds in Appendix B.1. Our main result, Theorem 33, is presented in Appendix B.2 along with a quantitative Morse lemma. Finally, in Appendix B.3 we derive some moments bounds.

Let $u: \mathbb{R}^d \times \mathsf{Z} \to \mathbb{R}$ with Z a metric space. We consider the following assumption.

H5 $u \in \mathrm{C}(\mathbb{R}^d \times \mathsf{Z}, \mathbb{R})$, *for any* $z \in \mathsf{Z}$, $u(\cdot, z) \in \mathrm{C}^2(\mathbb{R}^d, \mathbb{R})$ *and the following hold:*

(a) Z *is compact.*

(b) There exists $A \geq 0$ *such that for any* $z \in \mathsf{Z}$, $|u(0,z)| \leq A$.

(c) There exists $\mathtt{M} \geq 0$ *such that for any* $x_1, x_2 \in \mathbb{R}^d$ *and* $z \in \mathsf{Z}$, $\|\nabla_x^k u(x_1, z) - \nabla_x^k u(x_2, z)\| \leq \mathtt{M} \|x_1 - x_2\|$.

(d) There exists $\mathtt{m}, \alpha > 0$ *and* $R \geq 0$ *such that for any* $x \in \mathbb{R}^d$ *with* $\|x\| \geq R$, $u(x,z) \geq \mathtt{m}\|x\|^\alpha$.

(e) For any $z \in \mathsf{Z}$ *the number of global minimizers is bounded.*

There exists $R_A \geq 0$ such that for any $x \in \mathbb{R}^d$ with $\|x\| \geq R_A$ and $z \in \mathsf{Z}$, $|u(x,z)| \geq A \geq |u(0,z)|$. Hence, for any $z \in \mathsf{Z}$, $\arg\min\{u(x,z) \; : \; x \in \mathbb{R}^d\} \subset \mathsf{C} = \bar{\mathrm{B}}(0, R_A)$.

We denote $u^\star: \mathsf{Z} \to \mathbb{R}$ such that for any $z \in \mathsf{Z}$, $u^\star(z) = \min\{u(x,z) \; : \; x \in \mathbb{R}^d\}$. We have that $u^\star \in \mathrm{C}(\mathsf{Z}, \mathbb{R})$. Indeed we have that u is uniformly continuous on $\mathsf{Z} \times \mathsf{C}$. Hence, for any $\varepsilon > 0$ there exists $\delta > 0$ such that for any $(z_0, x_0), (z_0, x_0) \in \mathsf{Z} \times \mathsf{C}$ with $d(z_0, z_1) + \|x_0 - x_1\| \leq \delta$, $|u(x_0, z_0) - u(x_1, z_1)| \leq \varepsilon$. Let $z_0 \in \mathsf{Z}$ and $x_0^\star \in \arg\min\{u(x, z_0) \; : \; x \in \mathbb{R}^d\} \subset \mathsf{C}$. Let $\varepsilon > 0$, $z \in \mathsf{Z}$ such that $d(z, z_0) \leq \delta$ and $x^\star \in \arg\min\{u(x, z_0) \; : \; x \in \mathbb{R}^d\} \subset \mathsf{C}$ then

$$u^\star(z) \leq u(x^\star, z) \leq u(x_0^\star, z) \leq u(x_0^\star, z) \leq u(x_0^\star, z_0) \leq u^\star(z_0) \; .$$

Similarly we have $u^\star(z_0) \leq u^\star(z) + \varepsilon$ which concludes the proof. For any $\varphi : \mathbb{R}^d \to \mathbb{R}_+$ and $\varepsilon > 0$ we define

$$\begin{aligned} &\mathcal{I}_\varepsilon(\varphi, z) = C_\varepsilon^{-1} \textstyle\int_{\mathbb{R}^d} \varphi(x) \exp[-(u(x,z) - u^\star(z))/\varepsilon] \mathrm{d}x \;, \quad \mathcal{J}_\varepsilon(z) = \mathcal{I}_\varepsilon(1, z) \;, \\ &C_\varepsilon = \textstyle\int_{\mathbb{R}^d} \exp[-\left\|x\right\|^2/\varepsilon] \mathrm{d}x = (\pi\varepsilon)^{d/2} \;. \end{aligned} \tag{82}$$

In addition, we define

$$\mathcal{I}_0(\varphi, z) = \textstyle\int_{\arg\min u(\cdot, z)} \varphi(x) \det(\nabla_x^2 u(x,z))^{-1/2} \mathrm{d}\mathcal{H}^0(x) \;, \quad \mathcal{J}_0(z) = \mathcal{I}_0(1, z) \;, \tag{83}$$

where $\det(\nabla_x^2 u(x,z))^{-1/2} = +\infty$ if $\nabla_x^2 u(x,z)$ is not invertible. If $\varphi : \mathbb{R}^d \to \mathbb{R}$ and $I_\varepsilon(|\varphi|) < +\infty$ for some $\varepsilon \geq 0$ we define $\mathcal{I}_\varepsilon(\varphi)$ similarly as in (82), (83). For any $\varepsilon \geq 0$ and $\varphi : \mathbb{R}^d \to \mathbb{R}^p$ such that it is defined we let $\updelta_z \mathrm{S}_\varepsilon[\varphi] = \mathcal{I}_\varepsilon(\varphi, z)/\mathcal{J}_\varepsilon(z)$. We emphasize that these definitions are the parametric counterparts to the ones introduced in Section 4.1.

In what follows, we define $\sigma : \mathbb{R}^d \times \mathsf{Z} \to [0, +\infty]$ such that for any $z \in \mathsf{Z}$ and $x \in \mathbb{R}^d$ $\sigma(x,z)$ is the inverse of the minimum eigenvalue of $\nabla_x^2 u(x,z)$. In addition, for any $\upbeta > 0$, we define $\sigma_\upbeta^\star : \mathsf{Z} \to [0, +\infty]$ such that for any $z \in \mathsf{Z}$, $\sigma_\upbeta^\star(z) = \int_{\arg\min u(\cdot,z)} \sigma(x,z)^\upbeta \mathrm{d}\mathcal{H}^0(x)$.

B.1 Parametric lower and truncation bounds

Lemma 29 *Assume* **H**5*. Then, for any $\bar\varepsilon \geq 0$ there exists $A_0 > 0$ such that for any $z \in \mathsf{Z}$ and $\varepsilon \in [0, \bar\varepsilon]$ we have $\mathcal{J}_\varepsilon(z) \geq A_0$.*

Proof Let $z \in \mathsf{Z}$ and $\varepsilon > 0$. Let $x^\star(z)$ be a global minimizer of $x \mapsto u(x,z)$ which exists since $\arg\min\{u(x,z) \,:\, x \in \mathbb{R}^d\} \neq \emptyset$. Using the change of variable $x \mapsto x + x^\star(z)$ we have

$$\mathcal{J}_\varepsilon(z) = C_\varepsilon^{-1} \textstyle\int_{\mathbb{R}^d} \exp[-(u(x,z) - u^\star(z))/\varepsilon] \mathrm{d}x = C_\varepsilon^{-1} \textstyle\int_{\mathbb{R}^d} \exp[-(u(x^\star(z) + x, z) - u^\star(z))/\varepsilon] \mathrm{d}x \;. \tag{84}$$

For any $x \in \mathbb{R}^d$ we have

$$u(x^\star(z) + x, z) - u^\star(z) = \textstyle\int_0^1 \langle \nabla_x u(x^\star(z) + tx, z) - \nabla_x u(x^\star(z), z), x\rangle \mathrm{d}t \leq \mathtt{M}\|x\|^2/2 \;.$$

Combining this result and (84) we get $\mathcal{J}_\varepsilon(z) \geq \int_{\mathbb{R}^d} \exp[-(\mathtt{M}/2)\|x\|^2/\varepsilon]\mathrm{d}x/C_\varepsilon \geq (2/\mathtt{M})^{d/2}$. In addition, we have that for any $x \in \mathbb{R}^d$, $\mathcal{J}_0(z) = \int_{\arg\min u(\cdot,z)} \det(\nabla_x^2 u(x,z))^{-1/2} \mathrm{d}\mathcal{H}^0(x)$. Note that if $\sigma_1^\star(z) = +\infty$, then there exists $\tilde{x}(z) \in \arg\min\{u(x,z) \;:\; x \in \mathbb{R}^d\}$ such that $\sigma(\tilde{x}(z), z) = +\infty$ and therefore $\mathcal{J}_0(z) = +\infty$. We have that for any $x \in \mathbb{R}^d$, $\det(\nabla_x^2 u(x,z)) \leq \mathtt{M}^d$ and therefore, $\mathcal{J}_0(z) \geq N\mathtt{M}^{-d/2}$. We conclude upon letting $A_0 = \min(N, 2^{d/2})/\mathtt{M}^{d/2}$. ■

Lemma 30 *Assume* **H**5*. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ and $C_\varphi \geq 0$ such that for any $x \in \mathbb{R}^d$*

$$|\varphi(x)| \leq C_\varphi \exp[C_\varphi \left\|x\right\|^\alpha] \;. \tag{85}$$

Let $\bar\varepsilon \in (0, \mathtt{m}/(1 + C_\varphi))$ and $z \in \mathsf{Z}$. Assume that there exists $\mathsf{V}(z) \subset \mathbb{R}^d$ open and bounded such that

$$\arg\min\{u(x,z) \;:\; x \in \mathbb{R}^d\} \subset \mathsf{V}(z) \subset \arg\min\{u(x,z) \;:\; x \in \mathbb{R}^d\} + \bar{\mathrm{B}}(0,1) \;.$$

Then, there exist $\beta_1 > 0$ *and* $A_1 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ *such that for any* $\varepsilon \in (0, \bar{\varepsilon})$

$$\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi, z) \leq A_1(C_\varphi)\varepsilon^{-d/2}\{\exp[-m(z)/\varepsilon] + \exp[-\beta_1/\varepsilon]\} ,$$

with $m(z) = \inf\{u(x, z) \ : \ x \in \mathbb{R}^d \backslash \mathsf{V}(z)\} - u^\star(z)$*,* $\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi, z) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\mathsf{V}(z)^{\mathrm{c}}}, z)$ *and* A_1, β_1 *that do not depend on* φ *and* z*, with* A_1 *non-decreasing.*

Proof First, we have using the remark following **H**5

$$\mathsf{V}(z) \subset \mathsf{C} + \bar{\mathrm{B}}(0, 1) .$$

Since $\mathsf{C} + \bar{\mathrm{B}}(0, 1)$ is compact, there exists $R' \geq 0$ (that does not depend on z) such that $R' \geq R$ (where R is given in **H**5) and $\mathsf{V}(z) \subset \bar{\mathrm{B}}(0, R')$. Note that for any $\varepsilon > 0$, we have

$$\begin{aligned}
&\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi, z) = \mathcal{I}_\varepsilon^1(\varphi, z) + \mathcal{I}_\varepsilon^2(\varphi, z) , \\
&\mathcal{I}_\varepsilon^1(\varphi, z) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\bar{\mathrm{B}}(0,R')^{\mathrm{c}}}, z) , \quad \mathcal{I}_\varepsilon^2(\varphi, z) = \mathcal{I}_\varepsilon(\varphi \mathbb{1}_{\mathsf{V}^{\mathrm{c}} \cap \bar{\mathrm{B}}(0,R')}, z) .
\end{aligned}$$

Let $\varepsilon \in (0, \bar{\varepsilon})$. The rest of the proof is similar to the one of Theorem 25 but is given for completeness. We divide the proof into two parts. First, we bound $\mathcal{I}_\varepsilon^1(\varphi, z)$ and then $\mathcal{I}_\varepsilon^2(\varphi, z)$.

(a) Let $w = (\mathtt{m}/\varepsilon - C_\varphi)^{1/\alpha}$ (which makes sense, since $\varepsilon < \mathtt{m}/(C_\varphi + 1)$). Since $R' \geq R$, using (85) and the fact that $u \geq 1$, we have

$$\begin{aligned}
\mathcal{I}_\varepsilon^1(\varphi, z) &= C_1^{-1}\varepsilon^{-d/2} \textstyle\int_{\bar{\mathrm{B}}(0,R')^{\mathrm{c}}} \varphi(x) \exp[-(u(x, z) - u^\star(z))/\varepsilon]\mathrm{d}x \\
&\leq C_1^{-1}C_\varphi\varepsilon^{-d/2} \exp[\bar{u}^\star] \textstyle\int_{\bar{\mathrm{B}}(0,R')^{\mathrm{c}}} \exp[-(\mathtt{m}/\varepsilon - C_\varphi) \|x\|^\alpha]\mathrm{d}x \\
&\leq C_1^{-1}C_\varphi \exp[\bar{u}^\star]\varepsilon^{-d/2} \textstyle\int_{\bar{\mathrm{B}}(0,R'w)^{\mathrm{c}}} \exp[- \|x\|^\alpha]\mathrm{d}x ,
\end{aligned}$$

where $\bar{u}^\star = \sup\{u^\star(z) \ : \ z \in \mathsf{Z}\}$. Let $C_\alpha = \int_{\mathbb{R}^d} \exp[- \|x\|^\alpha]\mathrm{d}x$. Using that $w = (\mathtt{m}/\varepsilon - C_\varphi)^{1/\alpha}$, we have

$$\begin{aligned}
\mathcal{I}_\varepsilon^1(\varphi, z) &\leq C_1^{-1}C_\varphi \exp[\bar{u}^\star]\varepsilon^{-d/2} \textstyle\int_{\bar{\mathrm{B}}(0,R'w)^{\mathrm{c}}} \exp[- \|x\|^\alpha]\mathrm{d}x \\
&\leq C_1^{-1}C_\varphi \exp[\bar{u}^\star]\varepsilon^{-d/2} \textstyle\int_{\mathbb{R}^d} \exp[- \|x\|^\alpha /2]\mathrm{d}x \exp[-(R'u)^\alpha/2] \\
&\leq 2^{d/\alpha}C_\alpha C_1^{-1}C_\varphi \exp[\bar{u}^\star] \exp[(R')^\alpha C_\varphi/2]\varepsilon^{-d/2} \exp[-(R')^\alpha\mathtt{m}/(2\varepsilon)] \\
&\leq A_1^1\varepsilon^{-d/2} \exp[-\beta_1^1/\varepsilon] , \qquad (86)
\end{aligned}$$

with

$$A_1^1 = 2^{d/\alpha}C_\alpha C_1^{-1}C_\varphi \exp[\bar{u}^\star] \exp[(R')^\alpha C_\varphi/2] , \qquad \beta_1^1 = (R')^\alpha\mathtt{m}/2 .$$

(b) Second, let $\mathsf{K}(z) = \mathsf{V}^{\mathrm{c}}(z) \cap \bar{\mathrm{B}}(0, R')$. Note that for any $x \in \mathsf{K}(z)$, $u(x, z) - u^\star(z) \geq m(z)$. Hence, we have

$$\mathcal{I}_\varepsilon^2(\varphi, z) \leq C_1^{-1}C_\varphi\varepsilon^{-d/2} \exp[C_\varphi(R')^\alpha] \exp[-m(z)/\varepsilon]\lambda(\mathsf{K}(z)) ,$$

where we recall that $\lambda(\mathsf{K}(z))$ is the Lebesgue measure of $\mathsf{K}(z)$. Since $\mathsf{K}(z) \subset \bar{\mathrm{B}}(0, R')$ we have

$$\begin{aligned}
\mathcal{I}_\varepsilon^2(\varphi, z) &\leq \pi^{d/2}(R')^d\Gamma^{-1}(d/2 + 1)C_1^{-1}C_\varphi \exp[(R')^\alpha]\varepsilon^{-d/2} \exp[-m(z)/\varepsilon] \\
&\leq A_1^2\varepsilon^{-d/2} \exp[-m(z)/\varepsilon] , \qquad (87)
\end{aligned}$$

where $\Gamma : (0,+\infty) \to \mathbb{R}_+$ is given for any $s \in (0,+\infty)$ by $\Gamma(s) = \int_0^{+\infty} t^{s-1} \exp[-t]\mathrm{d}t$ and

$$A_1^2 = \pi^{d/2}(R')^d \Gamma^{-1}(d/2+1) C_1^{-1} C_\varphi \exp[C_\varphi (R')^\alpha] .$$

We conclude the proof upon combining (86), (87) and letting $\beta_1 = \beta_1^1$ and $A_1 = A_1^1 + A_1^2$. ■

B.2 Quantitative Morse lemma and parametric Laplace-type results

We begin by recalling a quantitative version of the Morse lemma, see (Le Loi and Phien, 2014, Theorem 4.2).

Proposition 31 *Let $\mathsf{U} \subset \mathbb{R}^d$ be open with $x_0 \in \mathsf{U}$ and $f \in \mathrm{C}^k(\mathsf{U}, \mathbb{R})$ with $k \in \mathbb{N}$ and $k \geq 3$. Assume that $\nabla f(x_0) = 0$ and let $K \geq 0$ such that $K \geq \max_{j \in \{1,\dots,k\}} \sup\{\|\mathrm{D}^j f(x)\| : x \in \mathsf{U}\} < +\infty$. Let σ be the minimal eigenvalue of $\nabla^2 f(x_0)$ and assume that $\sigma > 0$. Let $\sigma_0 = \min(\sigma, 1)$. Then, there exist $c_0 > 0$, $\delta = c_0 \sigma_0^4$ and $\Phi : \mathrm{B}(0,\delta) \to \Phi(\mathrm{B}(0,\delta))$ such that the following hold:*

(a) $\Phi \in \mathrm{C}^{k-1}(\mathrm{B}(0,\delta), \Phi(\mathrm{B}(0,\delta)))$ is a diffeomorphism.

(b) For any $x \in \mathrm{B}(0,\delta)$, $f(\Phi(x)) = f(x_0) + \|x\|^2$.

(c) There exist $c_1, \upbeta > 0$ such that $\max_{j \in \{1,\dots,k-1\}} \sup\{\|\mathrm{D}^j \Phi(x)\| : x \in \mathrm{B}(0,\delta)\} \leq c_1 \sigma_0^{-\upbeta}$.

In addition, c_0, c_1 and $\upbeta$ depend only on k, d and K.

Note that in (Le Loi and Phien, 2014, Theorem 4.2), the constant $c_1 \sigma_0^{-\upbeta}$ is replaced by $M(K, \sigma_0, k)$ where $M : \mathbb{R}_+^3 \to \mathbb{R}_+$. However, a close examination of the proof reveals that the dependency of $M(K, \sigma_0, k)$ with respect to σ_0^{-1} is of order $\sigma_0^{-\upbeta}$ for some $\upbeta > 0$ which can be made explicit. Using Theorem 31 we derive the following parametric version of Theorem 13.

This proposition relies on the notion of *thermodynamic barrier* associated with u which we define as follows. Assume that $\arg\min\{u(x,z) : x \in \mathbb{R}^d\} \neq \emptyset$ for any $z \in \mathsf{Z}$. For any $z \in \mathsf{Z}$ we introduce $\mathsf{A}(z)$ such that $\mathsf{A}(z) = \emptyset$ if there are no other minimizers than the global minimizers and $\mathsf{A}(z) = \{u(x,z) : x \text{ is a local minimizer of } u(\cdot,z) \text{ but not a global minimizer}\}$ otherwise. Then we define $c^\star : \mathsf{Z} \to \mathbb{R}$ such that for any $z \in \mathsf{Z}$ we have

$$c^\star(z) = \inf \mathsf{A}(z) - \inf\{u(x,z) : x \in \mathbb{R}^d\} ,$$

with the convention that $\inf \emptyset = +\infty$. In words, the *thermodynamic barrier* constant quantifies how close the value of the local minimizers are from the global ones. We refer to Section 3.2.3 for a discussion on the importance of *thermodynamic barrier* when establishing parametric Laplace-type expansions.

Proposition 32 *Assume* **H**5. *Let $\varphi \in \mathrm{C}(\mathbb{R}^d, \mathbb{R})$ and $z \in \mathsf{Z}$, and assume that $\sigma_1^\star(z) < +\infty$. There exists $\mathsf{V}(z)$ open such that*

$$\arg\min\{u(x,z) : x \in \mathbb{R}^d\} \subset \mathsf{V}(z) \subset \arg\min\{u(x,z) : x \in \mathbb{R}^d\} + \bar{\mathrm{B}}(0,1) .$$

We have that $\lim_{\varepsilon \to 0} |\mathcal{I}^{\mathrm{in}}_{\varepsilon}(\varphi, z) - \mathcal{I}_0(\varphi, z)| = 0$, *with* $\mathcal{I}^{\mathrm{in}}_{\varepsilon}(\varphi, z) = \mathcal{I}_{\varepsilon}(\varphi \mathbb{1}_{\mathsf{V}(z)}, z)$. *Assume that* $\varphi \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$. *Then there exist* $B_1 \geq 0$ *and* $\beta > 0$ *such that for any* $\varepsilon > 0$ *we have*

$$|\mathcal{I}^{\mathrm{in}}_{\varepsilon}(\varphi, z) - \mathcal{I}_0(\varphi, z)| \leq B_1 \sigma^{\star}_{\beta}(z)(1 + M_{0,\varphi}(z) + M_{1,\varphi}(z))\varepsilon^{1/2} , \tag{88}$$

with B_1 *that does not depend on* φ *and* z, *and for any* $i \in \{0, 1\}$, $M_{i,\varphi}(z) = \sup\{\|\nabla^i \varphi(x)\| : x \in \mathsf{V}(z)\}$. *In addition, there exist* $c_1, \gamma > 0$ *(that do not depend on* z*) such that* $m(z) \geq \min(c_1/\sigma^{\star}_{\gamma}(z), c^{\star}(z))$, *with* $m(z) = \inf\{u(x, z) : x \in \mathbb{R}^d \backslash \mathsf{V}(z)\} - u^{\star}(z)$.

Proof Let $\{x^k_{\star}(z)\}_{k=1}^M = \arg\min\{u(x, z) : x \in \mathbb{R}^d\}$. Using **H**5 $\{x^k_{\star}(z)\}_{k=1}^M \subset \mathsf{C}$ with C compact. We define $K_{\mathrm{global}} \geq 0$ such that

$$K_{\mathrm{global}} = \sup\{\|\nabla^j_x u(x, z)\| : x \in \mathsf{K},\ z \in \mathsf{Z},\ j \in \{1, 2, 3\}\} ,$$

with $\mathsf{K} = \mathsf{C} + \bar{\mathrm{B}}(0, 1)$, where C is defined in the remark following **H**5. Let $i \in \{1, \dots, M\}$. Using that $\sigma^{\star}_1(z) < +\infty$ and Theorem 31 with $K \leftarrow K_{\mathrm{global}}$ and $k = 3$, there exist $\Phi^z_i \in \mathrm{C}^2(\mathrm{B}(0, \delta_i(z)), \Phi^z_i(\mathrm{B}(0, \delta_i(z))))$ which is a diffeomorphism, with $\Phi^z_i(0) = x^i_{\star}(z)$, $\delta_i(z) = c_0 \min(\sigma^4(x^i_{\star}(z), z), 1)$ and for any $x \in \mathrm{B}(0, \delta_i(z))$ we have $u(\Phi^z_i(x), z) = u^{\star} + \|x\|^2$.

We let $\delta_0(z) = \min\{\delta_i(z) : i \in \{1, \dots, M\}\}$. Using Theorem 31, there exist $c'_0 \geq 0$ and $\alpha > 0$ that do not depend on z such that for any $\ell \in \{1, \dots, M\}$ and $x \in \mathrm{B}(0, \delta_0(z))$, $\|\mathrm{d}\Phi^z_{\ell}(x)\| \leq c'_0 \sigma^{\star}_{\alpha}(z)$. Let $\delta(z) = \min(\delta_0(z), 1/(c'_0 \sigma^{\star}_{\alpha}(z)))$. We have that for any $\ell \in \{1, \dots, N\}$, $\Phi_{\ell}(\mathrm{B}(0, \delta(z))) \subset \mathrm{B}(x^{\ell}_{\star}(z), c'_0 \sigma^{\star}_{\alpha}(z)\delta(z)) \subset \arg\min\{u(x, z) : x \in \mathbb{R}^d\} + \bar{\mathrm{B}}(0, 1)$. We let $\mathsf{V}(z) = \cup_{i=1}^M \Phi_{\ell}(\mathrm{B}(0, \delta(z)))$.

We now show that for any $i, j \in \{1, \dots, M\}$, $\Phi^z_i(\mathrm{B}(0, \delta(z))) \cap \Phi^z_j(\mathrm{B}(0, \delta(z))) = \emptyset$. Let $i, j \in \{1, \dots, M\}$ and for ease of notation let $\mathsf{W}_{\ell} = \Phi^z_{\ell}(\mathrm{B}(0, \delta(z)))$ for any $\ell \in \{1, \dots, M\}$. Assume that $\mathsf{W}_i \cap \mathsf{W}_j \neq \emptyset$. Then, since W_i and W_j are connected, $\mathsf{W} = \mathsf{W}_i \cup \mathsf{W}_j$ is connected as well. In addition, note that $\Phi^z_i(0) \notin \mathsf{W}_j$ and $\Phi^z_j(0) \notin \mathsf{W}_i$. There exists $\gamma \in \mathrm{C}([0, 1], \mathsf{W})$ such that $\gamma(0) = \Phi^z_i(0)$ and $\gamma(1) = \Phi^z_j(0)$. Denote $t^{\star} = \inf\{t \in [0, 1] : \gamma(t) \in \mathsf{W}_j\}$. We have that $\gamma(t^{\star}) \in \bar{\mathsf{W}}_j \backslash \mathsf{W}_j$. Hence $u(\gamma(t^{\star}), z) = \delta(z)^2 + u^{\star}(z)$. But $\gamma(t^{\star}) \in \mathsf{W}_i$ and therefore $u(\gamma(t^{\star}), z) < \delta(z)^2 + u^{\star}(z)$. This is absurd hence for any $i, j \in \{1, \dots, M\}$, $\mathsf{W}_i \cap \mathsf{W}_j = \emptyset$.

Let $\varepsilon > 0$ and $\mathcal{I}^{\mathrm{in}}_{0,\varepsilon}$ be given by

$$\mathcal{I}^{\mathrm{in}}_{0,\varepsilon}(\varphi, z) = \textstyle\sum_{\ell=1}^M \int_{\mathrm{B}(0, \delta(z)/\varepsilon^{1/2})} \exp[-\|x\|^2] \mathrm{d}x \varphi(x^{\ell}_{\star}(z)) \det(\nabla^2_x u(x^{\ell}_{\star}(z), z))^{-1/2}/C_1 . \tag{89}$$

Recall that $\mathcal{I}_0(\varphi, z) = \sum_{\ell=1}^M \varphi(x^{\ell}_{\star}(z)) \det(\nabla^2_x u(x^{\ell}_{\star}(z), z))^{-1/2}$. Therefore, we have

$$\begin{aligned} \left| \mathcal{I}_0(\varphi, z) - \mathcal{I}^{\mathrm{in}}_{0,\varepsilon}(\varphi, z) \right| &\leq \sigma^{\star}_{1/2} M_{0,\varphi} \exp[-\delta(z)^2/(2\varepsilon)] \textstyle\int_{\mathbb{R}^d} \exp[-\|x\|^2/2] \mathrm{d}x/C_1 \\ &\leq 2^{d/2} \sigma^{\star}_{1/2} M_{0,\varphi} \exp[-\delta(z)^2/(2\varepsilon)] . \end{aligned} \tag{90}$$

Finally, using (89), that for any $\ell \in \{1, \dots, M\}$ and $x \in \mathrm{B}(0, \delta(z))$, $u(\Phi^z_{\ell}(x), z) = u^{\star}(z) + \|x\|^2$, that for any $\ell \in \{1, \dots, M\}$, $\det(\mathrm{d}\Phi^z_{\ell}(0)) = \det(\nabla^2_x u(x^{\ell}_{\star}(z), z))^{-1/2}$ and $\Phi^z_{\ell}(0) = x^{\ell}_{\star}(z)$, we have

$$\begin{aligned} |\mathcal{I}^{\mathrm{in}}_{0,\varepsilon}(\varphi, z) - \mathcal{I}^{\mathrm{in}}_{\varepsilon}(\varphi, z)| = \textstyle\sum_{\ell=1}^M \int_{\mathrm{B}(0, \delta(z)/\varepsilon^{1/2})} &\Big| \varphi(\Phi^z_{\ell}(0)) \det(\mathrm{d}\Phi^z_{\ell}(0)) \\ &- \varphi(\Phi^z_{\ell}(\varepsilon^{1/2}x)) \det(\mathrm{d}\Phi^z_{\ell}(\varepsilon^{1/2}x)) \Big| \exp[-\|x\|^2] \mathrm{d}x/C_1 , \end{aligned} \tag{91}$$

which concludes the first part of the proof upon combining this result and the dominated convergence theorem. Next assume that $\varphi \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$ and for any $\ell \in \{1, \dots, M\}$, let $\chi_\ell^z :\ \mathrm{B}(0, \delta(z)) \to \mathbb{R}$ given for any $x \in \mathrm{B}(0, \delta(z))$ by

$$\chi_\ell^z(x) = \varphi(\Phi_\ell^z(x)) \det(\mathrm{d}\Phi_\ell^z(x))$$

We have that for any $\ell \in \{1, \dots, M\}$, $\chi_\ell^z \in \mathrm{C}^1(\mathrm{B}(0, \delta))$. Therefore, we have that for any $\ell \in \{1, \dots, M\}$ and $x \in \mathrm{B}(0, \delta(z))$ and $h \in \mathbb{R}^d$

$$\mathrm{d}\chi_\ell(x)(h) = \mathrm{d}\varphi(\Phi_\ell^z(x))\mathrm{d}\Phi_\ell^z(x)(h) \det(\mathrm{d}\Phi_\ell^z(x)) + \varphi(\Phi_\ell^z(x)) \operatorname{Tr}(\operatorname{Adj}(\mathrm{d}\Phi_\ell^z(x))\mathrm{d}^2\Phi_\ell^z(x)(h)) \ .$$

Therefore, using Theorem 31 we have that there exist $C \geq 0$ and $\beta > 0$ such that for any $\ell \in \{1, \dots, M\}$ and $x \in \mathrm{B}(0, \delta(z))$, $\|\mathrm{d}\chi_\ell^z(x)\| \leq C(1 + M_{0,\varphi} + M_{1,\varphi})\sigma_\beta^\star(z)$. Using this result in (91) we get that

$$\left|\mathcal{I}_{0,\varepsilon}^{\mathrm{in}}(\varphi, z) - \mathcal{I}_\varepsilon^{\mathrm{in}}(\varphi, z)\right| \leq C\sigma_\beta^\star(z)\varepsilon^{1/2}(1 + M_{0,\varphi} + M_{1,\varphi}) \int_{\mathbb{R}^d} \|x\| \exp[-\|x\|^2]\mathrm{d}x/C_1 \ ,$$

which concludes the proof of (88) upon combining this result, (90), and the fact that $\exp[-t] \leq 1/t$ for any $t > 0$.

Next, we show that there exist $c_1, \beta > 0$ such that $m(z) \geq c_1 \min(1/\sigma_\beta^\star(z), c^\star(z))$ with $c_1, \beta > 0$ that do not depend on z. Since $\lim_{\|x\| \to +\infty} u(x, z) = +\infty$, there exists $\tilde{x}(z)$ which minimizes $x \mapsto u(x, z)$ on $\mathsf{E} = \mathbb{R}^d \backslash \mathsf{V}(z)$. We distinguish two cases. If $\tilde{x}(z) \in \operatorname{int}(\mathsf{E})$ then $\tilde{x}(z)$ is a local minimizer of $x \mapsto u(x, z)$. Hence $m(z) \geq c^\star(z)$. If $\tilde{x}(z) \in \mathsf{E} \backslash \operatorname{int}(\mathsf{E})$ then $\tilde{x}(z) \in \bar{\mathsf{V}}(z) \backslash \mathsf{V}(z)$ and we have that $m(z) \geq \delta(z)^2$, which concludes the proof. ■

Finally using Theorem 29, Theorem 30 and Theorem 32 we establish our main result.

Proposition 33 *Assume* **H**5. *Let* $\varphi :\ \mathbb{R}^d \to \mathbb{R}$ *be a* $M_{1,\varphi}$*-Lipschitz function,* $M_{1,\varphi}, C_\varphi \geq 0$ *and* $z \in \mathsf{Z}$. *Assume that* $\sigma_1^\star(z) < +\infty$ *and that for any* $x \in \mathbb{R}^d$, $|\varphi(x)| \leq C_\varphi \exp[C_\varphi \|x\|^\alpha]$. *Then, there exist* $B_2 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ *and* $\beta > 0$ *such that*

$$|\delta_z \mathrm{S}_\varepsilon[\varphi] - \delta_z \mathrm{S}_0[\varphi]| \leq B_2(C_\varphi)(1 + M_{0,\varphi} + M_{1,\varphi})(1 + \sigma_\beta^\star(z))\{\varepsilon^{1/2} + \varepsilon^{-d/2} \exp[-c^\star(z)/\varepsilon]\} \ ,$$

with $M_{0,\varphi} = \sup\{|\varphi(x)| \ :\ x \in \mathsf{K}\}$, K, B_2 *and* β *that do not depend on* z, *and* B_2 *non-decreasing.*

Proof In this proof we assume that $\varphi \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$. The extension to Lipschitz function is similar to the proof of Theorem 3, see Section 4.1.3, *i.e.* we use the smoothing Theorem 28. Let $\bar{\varepsilon} \in (0, \mathtt{m}/(1 + C_\varphi))$. First, let $\mathsf{V}(z)$ be given by Theorem 32 and set K such that $\mathsf{K} = \mathsf{C} + \bar{\mathrm{B}}(0, 1)$, with C given in the remark following **H**5. Note that $\mathsf{V}(z) \subset \mathsf{K}$. Applying Theorem 29 there exists $A_0 \geq 0$ such that for any $\varepsilon \in [0, \bar{\varepsilon}]$,

$$\mathcal{J}_\varepsilon(z) \geq A_0 \ . \tag{92}$$

In addition, using Theorem 30 we have that there exist $\beta_1 > 0$ and $A_1 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$ we have

$$\mathcal{I}_\varepsilon^{\mathrm{out}}(\varphi, z) \leq A_1(C_\varphi)\varepsilon^{-d/2}\{\exp[-m(z)/\varepsilon] + \exp[-\beta_1/\varepsilon]\} \ . \tag{93}$$

Using Theorem 32 there exist $c_1, \gamma > 0$ such that $m(z) \geq \min(c_1/\sigma^\star_\gamma(z), c^\star(z))$ (with $c_1, \gamma > 0$ that do not depend on z). Hence, combining this result, (93) and the fact that there exists c_3 such that for any $t > 0$, $\exp[-1/t] \leq c_3 t^{(d+1)/2}$, we get

$$\mathcal{I}^{\mathrm{out}}_\varepsilon(\varphi, z) \leq A_1(C_\varphi)\varepsilon^{-d/2}\{\exp[-c^\star(z)/\varepsilon] + (\sigma^\star_\beta(z))^{(d+1)/2}\varepsilon^{(d+1)/2}/c_3 + \exp[-\beta_1/\varepsilon]\} \ .$$

Therefore, there exist $\beta' > 0$ and $A_2 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ (non-increasing) such that for any $\varepsilon \in (0, \bar{\varepsilon})$ we have

$$\mathcal{I}^{\mathrm{out}}_\varepsilon(\varphi, z) \leq A_2(C_\varphi)\{(1 + \sigma^\star_{\beta'}(z))\varepsilon^{1/2} + \varepsilon^{-d/2}\exp[-c^\star(z)/\varepsilon]\} \ . \tag{94}$$

In addition, using Theorem 32, there exist $B_1 \geq 0$ and $\beta'' > 0$ such that for any $\varepsilon > 0$ we have

$$|\mathcal{I}^{\mathrm{in}}_\varepsilon(\varphi, z) - \mathcal{I}_0(\varphi, z)| \leq B_1 \sigma^\star_{\beta''}(z)(1 + M_{0,\varphi}(z) + M_{1,\varphi}(z))\varepsilon^{1/2} \ , \tag{95}$$

Note that $M_{i,\varphi}(z) \leq M_{i,\varphi}$ for $i \in \{0, 1\}$. Hence, combining (94) and (95), there exist $A_3 \in \mathrm{C}(\mathbb{R}_+, \mathbb{R}_+)$ (non-increasing) and $\beta_0 > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$ we have

$$|\mathcal{I}_\varepsilon(\varphi, z) - \mathcal{I}_0(\varphi, z)| \leq A_3(C_\varphi)(1 + M_{0,\varphi} + M_{1,\varphi})\{(1 + \sigma^\star_{\beta_0}(z))\varepsilon^{1/2} + \varepsilon^{-d/2}\exp[-c^\star(z)/\varepsilon]\} \ ,$$

Similar results hold if φ is replaced by 1 and we get that there exist $A_4 \geq 0$, $\beta_1 > 0$ such that for any $\varepsilon \in (0, \bar{\varepsilon})$ we have

$$\begin{aligned}
&|\mathcal{I}_\varepsilon(\varphi, z) - \mathcal{I}_0(\varphi, z)| \leq A_3(C_\varphi)(1 + M_{0,\varphi} + M_{1,\varphi})\{(1 + \sigma^\star_{\beta_1}(z))\varepsilon^{1/2} + \varepsilon^{-d/2}\exp[-c^\star(z)/\varepsilon]\} \ , \\
&|\mathcal{J}_\varepsilon(z) - \mathcal{J}_0(z)| \leq A_4\{(1 + \sigma^\star_{\beta_1}(z))\varepsilon^{1/2} + \varepsilon^{-d/2}\exp[-c^\star(z)/\varepsilon]\} \ . \qquad (96)
\end{aligned}$$

In addition, we have that

$$|\delta_z \mathrm{S}_\varepsilon(\varphi) - \delta_z \mathrm{S}_0(\varphi)| \leq (\mathcal{I}_\varepsilon(\varphi, z) - \mathcal{I}_0(\varphi, z))/\mathcal{J}_\varepsilon(z) + \mathcal{I}_0(\varphi, z)(\mathcal{J}_0(z) - \mathcal{J}_\varepsilon(z))/(\mathcal{J}_0(z)\mathcal{J}_\varepsilon(z)) \ . \tag{97}$$

Finally, we have that $\mathcal{I}_0(\varphi, z) \leq M_{0,\varphi}\sigma^\star_1(z)$. Combining this result (92), (96) and (97) concludes the proof.

■

B.3 Control of the moments

In order to derive the uniform stability of the limiting measure, we first need to control the moments of $\delta_{z^{1:n}}\mathrm{S}_\varepsilon$ uniformly in w.r.t. ε, z and n. Let $z^{1:n} \in \mathsf{Z}^n$ and $\varepsilon > 0$. We consider the Langevin diffusion $(\mathbf{X}^\varepsilon_t(z^{1:n}))_{t \geq 0}$ given by the following Stochastic Differential Equation (SDE): $\mathbf{X}^\varepsilon_0(z^{1:n}) \in \mathbb{R}^d$ and

$$\mathrm{d}\mathbf{X}^\varepsilon_t(z^{1:n}) = -\nabla_x U_n(\mathbf{X}^\varepsilon_t(z^{1:n}), z^{1:n})\mathrm{d}t + \sqrt{2\varepsilon}\mathrm{d}\mathbf{B}_t \ ,$$

where $(\mathbf{B}_t)_{t \geq 0}$ is a d-dimensional Brownian motion with filtration $(\mathcal{F}_t)_{t \geq 0}$. We recall that for any $n \in \mathbb{N}$, $z^{1:n} \in \mathsf{Z}^n$ and $x \in \mathbb{R}^d$ we have $U_n(x, z^{1:n}) = (1/n)\sum_{i=1}^n u(x, z_i)$. Therefore, under $\mathbf{H}3(n)$ we have that $(\mathbf{X}^\varepsilon_t(z^{1:n}))_{t \geq 0}$ is well-defined and admits $\delta_{z^{1:n}}\mathrm{S}_\varepsilon$ as an invariant measure, see Roberts and Tweedie (1996) for instance.

Lemma 34 *Let $n \in \mathbb{N}$ and assume* $\mathbf{H}3(n)$ *and* $\mathbf{H}4(n)$. *Let $z^{1:n} \in \mathsf{Z}^n$ and assume that $\sigma_1^{\star}(z^{1:n}) < +\infty$. Then there exist $\bar{\varepsilon} > 0$ such that for any $k \in \mathbb{N}$ there exists $C_k \geq 0$ such that for any $\varepsilon \in [0, \bar{\varepsilon})$*

$$\textstyle\int_{\mathbb{R}^d} \|x\|^{2k} \, \mathrm{S}_{\varepsilon}(z^{1:n}, \mathrm{d}x) \leq C_k \ ,$$

with C_k and $\bar{\varepsilon}$ that do not depend on $n \in \mathbb{N}$ and $z^{1:n} \in \mathsf{Z}^n$.

Proof Let $\varepsilon > 0$. First, since $x \mapsto \nabla_x U_n(x, z^{1:n})$ is Lipschitz continuous we have that $(\mathbf{X}_t(z^{1:n})^{\varepsilon})_{t \geq 0}$ is well-defined and is a continuous semi-martingale using (Ikeda and Watanabe, 2014, Theorem 2.3, Theorem 2.4 , Chapter 4) such that for any $t \geq 0$

$$\mathbf{X}_t^{\varepsilon}(z^{1:n}) = \mathbf{X}_0^{\varepsilon}(z^{1:n}) - \textstyle\int_0^t \nabla_x U_n(\mathbf{X}_s^{\varepsilon}(z^{1:n}), z^{1:n}) \mathrm{d}s + \sqrt{2\varepsilon} \mathbf{B}_t \ .$$

Let $\mathtt{m}_0 > 0$. Using Itô's formula, see (Ikeda and Watanabe, 2014, Theorem 5.1) we have that for any $t \geq 0$ and $k \in \mathbb{N}$ with $k \geq 2$

$$\begin{aligned}
&\|\mathbf{X}_t^{\varepsilon}(z^{1:n})\|^{2k} \exp[\mathtt{m}_0 t] \\
&= \mathbf{X}_0^{\varepsilon}(z^{1:n}) + 2k \textstyle\int_0^t \langle \mathbf{X}_s^{\varepsilon}(z^{1:n}), \nabla_x U_n(\mathbf{X}_s^{\varepsilon}(z^{1:n}), z^{1:n}) \rangle \|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2(k-1)} \exp[\mathtt{m}_0 s] \mathrm{d}s \\
&\quad + \mathtt{m}_0 \textstyle\int_0^t \|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2k} \exp[\mathtt{m}_0 s] \mathrm{d}s + 4\varepsilon k(2k-1) \textstyle\int_0^t \|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2(k-1)} \exp[\mathtt{m}_0 s] \mathrm{d}s + \mathbf{M}_t^{\varepsilon}(z^{1:n}) \ ,
\end{aligned}$$

with $(\mathbf{M}_t^{\varepsilon}(z^{1:n}))_{t \geq 0}$ a $\mathcal{F}_t$-martingale such that $\mathbf{M}_0^{\varepsilon}(z^{1:n}) = 0$. First, using Fubini-Tonelli theorem, we have that for any $t \geq 0$ and $k \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}[\|\mathbf{X}_t^{\varepsilon}(z^{1:n})\|^{2k}] \exp[\mathtt{m}_0 t] &\leq \mathbf{X}_0^{\varepsilon}(z^{1:n}) + |2k\mathtt{m} - \mathtt{m}_0| \textstyle\int_0^t \mathbb{E}[\|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2k}] \exp[\mathtt{m}_0 s] \mathrm{d}s \\
&\quad + (4\varepsilon k(2k-1) + 2k\mathtt{c})(1 + \textstyle\int_0^t \mathbb{E}[\|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2k}] \exp[\mathtt{m}_0 s] \mathrm{d}s) \ .
\end{aligned}$$

Using Grönwall's lemma we have that for any $t \geq 0$ and $k \in \mathbb{N}$, $\mathbb{E}[\|\mathbf{X}_t^{\varepsilon}(z^{1:n})\|^{2k}] < +\infty$. Hence, using this result, the Fubini theorem and $\mathbf{H}3(n)$, we have that for any $t \geq 0$ and $k \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}[\|\mathbf{X}_t^{\varepsilon}(z^{1:n})\|^{2k}] \exp[\mathtt{m}_0 t] &\leq \mathbf{X}_0^{\varepsilon}(z^{1:n}) - (2k\mathtt{m} - \mathtt{m}_0) \textstyle\int_0^t \mathbb{E}[\|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2k}] \exp[\mathtt{m}_0 s] \mathrm{d}s \\
&\quad + (4\varepsilon k(2k-1) + 2k\mathtt{c}) \textstyle\int_0^t \mathbb{E}[\|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2(k-1)}] \exp[\mathtt{m}_0 s] \mathrm{d}s \ .
\end{aligned}$$

Combining this result and the fact that for any $t, a, b > 0$ and $k \in \mathbb{N}$, $at^{2(k-1)} \leq bt^{2k} + a^k/b^{k-1}$, we have for any $t \geq 0$ and $k \in \mathbb{N}$

$$\begin{aligned}
\mathbb{E}[\|\mathbf{X}_t^{\varepsilon}(z^{1:n})\|^{2k}] \exp[\mathtt{m}_0 t] &\leq \mathbf{X}_0^{\varepsilon}(z^{1:n}) - (k\mathtt{m} - \mathtt{m}_0) \textstyle\int_0^t \mathbb{E}[\|\mathbf{X}_s^{\varepsilon}(z^{1:n})\|^{2k}] \exp[\mathtt{m}_0 s] \mathrm{d}s \\
&\quad + (4\varepsilon k(2k-1) + 2k\mathtt{c})^k / ((k\mathtt{m})^{k-1} \mathtt{m}_0) \exp[\mathtt{m}_0 t] \ .
\end{aligned} \tag{98}$$

Therefore, for any $k \in \mathbb{N}$, there exists $C_k \geq 0$ such that for any $t \geq 0$, $\mathbb{E}[\|\mathbf{X}_t^{\varepsilon}(z^{1:n})\|^{2k}] \leq C_k$ upon letting $\mathtt{m}_0 = k\mathtt{m}$ in (98) with C_k that does not depend on ε and $z^{1:n}$. Therefore, using that the sequence of distributions associated with $(\mathbf{X}_t^{\varepsilon}(z^{1:n}))_{t \geq 0}$ weakly converges towards $\updelta_{z^{1:n}} \mathrm{S}_{\varepsilon}$, see Roberts and Tweedie (1996) for instance, and the monotone convergence theorem we get that $\int_{\mathbb{R}^d} \|x\|^{2k} \mathrm{S}_{\varepsilon}(z^{1:n}, \mathrm{d}x) \leq C_k$. We conclude upon using Theorem 33 in the case where $\varepsilon = 0$. ■

Finally, we will make use of the following lemma.

Lemma 35 *Let* $\mu, \nu \in \mathscr{P}(\mathbb{R}^d)$ *such that for any* $k \in \mathbb{N}$ *there exists* $C_k \geq 0$ *such that* $\int_{\mathbb{R}^d} \|x\|^{2k} \mathrm{d}\mu(x) + \int_{\mathbb{R}^d} \|x\|^{2k} \mathrm{d}\nu(x) \leq C_k$. *Then, for any* $k \in \mathbb{N}$ *we have that*

$$\mathbf{W}_2(\mu, \nu) \leq \sqrt{2C_k} \mathbf{W}_1(\mu, \nu)^{(k-1)/(2k-1)} .$$

Proof Let $k \in \mathbb{N}$ and (X, Y) be the optimal coupling between μ and ν w.r.t. the Wasserstein distance of order one. Let $Z = X - Y$. Using Hölder's inequality we have

$$\mathbb{E}[\|Z\|^2] \leq \mathbb{E}[\|Z\|^{2k}]^{1/(2k-1)} \mathbb{E}[\|Z\|]^{(2k-2)/(2k-1)} \leq \mathbb{E}[\|Z\|^{2k}]^{1/(2k-1)} \mathbf{W}_1(\mu, \nu)^{(2k-2)/(2k-1)} .$$

We conclude upon combining this result and the fact that $\mathbb{E}[\|Z\|^{2k}]^{1/(2k-1)} \leq 2C_k$. ■

Appendix C. Basics on flows and geometric measure theory

In this section, we recall basic facts from geometric measure theory. We refer to Federer (1969); Morgan (2016); Ambrosio et al. (2000) for a complete exposition of geometric measure theory concepts. We begin with a proposition establishing the existence of flows. Then, we prove useful facts on rectifiable sets. Finally, we state area and coarea formulas which are central to our analysis.

Lemma 36 *Let* $k \in \mathbb{N}^\star$ *and* $X \in \mathrm{C}^k(\mathbb{R}^d, \mathbb{R}^d)$. *Assume that* X *is compactly supported. Then, there exists a unique mapping* $\Phi \in \mathrm{C}^{k+1,k}(\mathbb{R} \times \mathbb{R}^d, \mathbb{R}^d)$ *such that for any* $x \in \mathbb{R}^d$, $\Phi(0, x) = x$ *and* $\partial_s \Phi(s, \mathbb{R}^d) = X(\Phi(s, \mathbb{R}^d))$. *In addition, for any* $s \in \mathbb{R}$, $x \mapsto \Phi(s, x)$ *is a diffeomorphism.*

Proof The first part of the proposition is an application of (Lang, 2002, Theorem 2.5, Theorem 2.6). The second part is an application of (Milnor, 1963, Lemma 2.4). ■

Note that the previous lemma can be extended to smooth manifolds.

Definition 37 *(Ambrosio et al., 2000, Definition 2.57) Let* $\mathsf{E} \subset \mathbb{R}^d$ *and* $k \in \mathbb{N}$ *with* $k \leq d$. E *is countably* $\mathcal{H}^k$*-rectifiable if there exists* $(\psi_i)_{i \in \mathbb{N}}$ *such that for any* $i \in \mathbb{N}$, $\psi_i : \mathbb{R}^k \to \mathbb{R}^d$ *is Lipschitz continuous and*

$$\mathcal{H}^k \left(\mathsf{E} \setminus \cup_{i \in \mathbb{N}} \psi_i(\mathbb{R}^k) \right) = 0 .$$

In what follows, we provide an easy criterion to verify if a given level set is a countably $\mathcal{H}^k$-rectifiable set. We start by recalling the following proposition.

Proposition 38 *(Lang, 2007, Theorem 10.5) Let* $k \in \mathbb{N}$ *with* $k \leq d$ *and* $F : \mathbb{R}^d \to \mathbb{R}^k$ *be Lipschitz continuous. Then for* λ*-almost every* $y \in \mathbb{R}^k$, $F^{-1}(\{y\})$ *is countably* $\mathcal{H}^{d-k}$*-rectifiable.*

The following lemma ensures the stability of countably $\mathcal{H}^k$-rectifiable sets through diffeomorphisms.

Lemma 39 *Let* $k \in \mathbb{N}$ *with* $k \leq d$, $\mathsf{E} \subset \mathbb{R}^d$ *be a compact countably* $\mathcal{H}^k$*-rectifiable set and* $\Phi \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ *be a diffeomorphism. Then* $\Phi(\mathsf{E})$ *is countably* $\mathcal{H}^k$*-rectifiable.*

Proof Since $\mathsf{E} \subset \mathbb{R}^d$ is countably $\mathcal{H}^k$-rectifiable there exists $(\psi_i)_{i\in\mathbb{N}}$ such that for any $i \in \mathbb{N}$, $\psi_i : \mathbb{R}^k \to \mathbb{R}^d$ is Lipschitz continuous and

$$\mathcal{H}^k\left(\mathsf{E} \setminus \cup_{i\in\mathbb{N}} \psi_i(\mathbb{R}^k)\right) = 0 \; .$$

Let $y \in \Phi(\mathsf{E}) \setminus \cup_{i\in\mathbb{N}} (\Phi \circ \psi_i)(\mathbb{R}^k)$. Then there exists $x \in \mathsf{E}$ such that $y = \Phi(x)$ and for any $i \in \mathbb{N}$ and $y_i \in \psi_i(\mathbb{R}^k)$, $\Phi(x) \neq \Phi(y_i)$, *i.e.* $x \neq y_i$. Therefore $\Phi(\mathsf{E}) \setminus \cup_{i\in\mathbb{N}} (\Phi \circ \psi_i)(\mathbb{R}^k) \subset \Phi(\mathsf{E} \setminus \cup_{i\in\mathbb{N}} \psi_i(\mathbb{R}^k))$. Φ is Lipschitz-continuous on E with constant $\mathsf{L}_{\mathsf{E}} \geq 0$ and therefore using (Ambrosio et al., 2000, Proposition 2.49 (iv)) we get

$$0 \leq \mathcal{H}^k\left(\Phi(\mathsf{E}) \setminus \cup_{i\in\mathbb{N}} (\Phi \circ \psi_i)(\mathbb{R}^k))\right) \leq \mathsf{L}_{\mathsf{E}}^k \mathcal{H}^k\left(\mathsf{E} \setminus \cup_{i\in\mathbb{N}} \psi_i(\mathbb{R}^k)\right) \leq 0 \; ,$$

which concludes the proof. ■

Finally, we show that if a function is regular enough then its level-sets are countably $\mathcal{H}^k$-rectifiable.

Lemma 40 *Let* $F \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^p)$ *with* $d \geq p$ *and* $\mathsf{A} \subset \mathbb{R}^p$ *compact such that* $F^{-1}(\mathsf{A})$ *is compact and* $0 \in \mathrm{int}(\mathsf{A})$*. In addition, assume that for any* $x \in F^{-1}(\{0\})$, $|\mathrm{J}F(x)| > 0$*. Then, there exists* $\eta > 0$ *such that for any* $t \in \bar{\mathrm{B}}_\infty(0, \eta)$, $F^{-1}(t)$ *is countably* $\mathcal{H}^{d-p}$*-rectifiable and* $\mathcal{H}^{d-p}(F^{-1}(t)) < +\infty$*.*

Proof In this proof, we show that the level sets $F^{-1}(t)$ and $F^{-1}(s)$ are diffeomorphic for $t, s \in \bar{\mathrm{B}}(0, \eta)$. Then, we conclude upon combining Theorem 38 and Theorem 39. First, note that $F^{-1}(0)$ is compact since it is closed in $F^{-1}(\mathsf{A})$ which is compact. There exists $\eta_0 > 0$ such that for any $x \in F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta_0))$, $|\mathrm{J}F(x)| > 0$. Indeed, if this is not the case then there exists $(x_k)_{k\in\mathbb{N}}$ with $\lim_{k\to+\infty} F(x_k) = 0$ and $|\mathrm{J}F(x_k)| = 0$. Since $0 \in \mathrm{int}(\mathsf{A})$ there exists $\eta' > 0$ such that $F^{-1}(\bar{\mathrm{B}}(0, \eta'))$ is compact and therefore there exists $x^\star$ such that, up to taking a subsequence, $\lim_{k\to+\infty} x_k = x^\star$. Then $F(x^\star) = 0$ and $\mathrm{J}F(x^\star) = 0$, which is absurd. We define $\mathsf{K}_0 = F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta_0))$ and $\mathsf{K}_1 = F^{-1}(\bar{\mathrm{B}}_\infty(0, \eta))$ with $\eta = \min(\eta_0/2, \eta')$. Note that $\mathsf{K}_1 \subset \mathrm{int}(\mathsf{K}_0)$.

Let $G(x) = \mathrm{D}F(x)\mathrm{D}F(x)^\top$. Note that for any $x \in \mathsf{K}_0$, $G(x)$ is invertible. We define $\{f_i\}_{i=1}^p$ such that for any $i \in \{1, \dots, p\}$, $f_i : \mathbb{R}^d \to \mathbb{R}^d$ with for any $x \in \mathsf{K}_0$

$$f_i(x) = \textstyle\sum_{k=1}^p h_{i,k}(x) \nabla F_k(x) \; ,$$

with $\{h_{i,j}(x)\}_{1\leq i,j\leq p} = G^{-1}(x)$. In addition, we assume that $f_i(x) = 0$ for any $x \notin \mathsf{K}_1$. For any $x \in \mathsf{K}_1$ and $i, j \in \{1, \dots, p\}$ we have

$$\langle f_i(x), \nabla F_j(x) \rangle = \textstyle\sum_{k=1}^p h_{i,k}(x) \langle \nabla F_k(x), \nabla F_j(x) \rangle = \delta_i(j) \; .$$

In what follows, we let $\{g_i\}_{i=1}^p$ such that for any $i \in \{1, \dots, p\}$, $g_i : \mathbb{R}^d \to \mathbb{R}^d$ and $g_i \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ such that for any $x \in \mathsf{K}_1$, $g_i(x) = f_i(x)$ and for any $x \notin \mathrm{int}(\mathsf{K}_1)$, $g_i(x) = 0$, such functions exist using Whitney extension theorem for instance, see Whitney (1934). In what follows, we fix $t = (t_0, \dots, t_p) \in \bar{\mathrm{B}}_\infty(0, \eta)$. For any $i \in \{1, \dots, p\}$ let $\Phi_i : \mathbb{R} \times \mathbb{R}^d \to \mathbb{R}^d$ given by $\Phi_i(0, \cdot) = \mathrm{Id}$ and for any $s \in \mathbb{R}$ and $x \in \mathbb{R}^d$

$$\partial_s \Phi_i(s, x) = -g_i(\Phi_i(s, x)) \ .$$

For any $i \in \{1, \dots, p\}$, Φ_i is well-defined using Theorem 36. Therefore, we have for any $i \in \{1, \dots, p\}$ and $s \in \mathbb{R}$, $x \in \mathbb{R}^d$ such that $\Phi(s, x) \in \mathsf{K}_1$

$$\partial_s F(\Phi_i(s, x)) = -(\langle g_i(\Phi_i(s, x)), \nabla F_1(\Phi_i(s, x))\rangle, \dots, \langle g_i(\Phi_i(s, x)), \nabla F_p(\Phi_i(s, x))\rangle) = -e_i \ ,$$

where we recall that $\{e_i\}_{i=1}^p$ is the canonical basis of $\mathbb{R}^p$. We define $\bar{\Phi}_t : \mathbb{R}^d \to \mathbb{R}^d$ such that for any $x \in \mathbb{R}^d$, $\bar{\Phi}_t(x) = x^{(p)}$ with $x^{(0)} = x$ and for any $i \in \{0, \dots, p-1\}$, $x^{(i+1)} = \Phi_{i+1}(t_{i+1}, x^{(i)})$. Note that $\bar{\Phi}_t \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R})$ is a diffeomorphism using Theorem 36. Using Theorem 38, there exists $t_0 \in \bar{\mathrm{B}}_\infty(0, \eta,)$ such that $F^{-1}(t_0)$ is countably $\mathcal{H}^{d-p}$ rectifiable. Let $s \in \bar{\mathrm{B}}_\infty(0, \eta)$. Using Theorem 39 and that $\bar{\Phi}_{-s} \circ \bar{\Phi}_{t_0}$ is a diffeomorphism between $F^{-1}(t_0)$ and $F^{-1}(s)$ we get that $F^{-1}(s)$ is countably $\mathcal{H}^{d-p}$ rectifiable, which concludes the first part of the proof.

For the second part of the proof, we let $R \geq 0$ such that $F^{-1}(\mathsf{A}) \subset \bar{\mathrm{B}}(0, R)$. Since $\bar{\mathrm{B}}(0, R)$ is $\mathcal{H}^d$-rectifiable we obtain using the coarea formula Theorem 41

$$\int_{\bar{\mathrm{B}}(0,R)} |\mathrm{J}F(x)| \mathrm{d}x = \int_{\mathbb{R}^p} \mathcal{H}^{d-p}(\bar{\mathrm{B}}(0, R) \cap F^{-1}(t)) \mathrm{d}t \geq \int_{\bar{\mathrm{B}}_\infty(0,\eta)} \mathcal{H}^{d-p}(F^{-1}(t)) \mathrm{d}t \ .$$

Therefore, there exists $t_0 \in \bar{\mathrm{B}}_\infty(0, \eta)$ such that $\mathcal{H}^{d-p}(F^{-1}(t_0)) < +\infty$. Let $\Psi = \bar{\Phi}_{-t} \circ \bar{\Phi}_{t_0}$ and $\mathtt{L} \geq 0$ such that for any $x \in F^{-1}(t_0)$ we have $\|\mathrm{d}\Psi(x)\| \leq \mathtt{L}$. Then, using (Ambrosio et al., 2000, Proposition 2.49 (iv)) we have

$$\mathcal{H}^{d-p}(F^{-1}(t)) = \mathcal{H}^{d-p}(\Psi(F^{-1}(t_0))) \leq \mathtt{L}^{d-p} \mathcal{H}^{d-p}(F^{-1}(t_0)) < +\infty \ ,$$

which concludes the proof. ■

We conclude this section with the area and coarea formulae.

Theorem 41 *Let $F : \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^p)$ be Lipschitz continuous, $k \in \mathbb{N}$ with $k \leq d$ and $\mathsf{E} \subset \mathbb{R}^d$ be a countably $\mathcal{H}^k$-rectifiable set. Let $\varphi : \mathbb{R}^d \to \mathbb{R}$ measurable such that*

$$\int_{\mathsf{E}} |\varphi(x)| \, \mathrm{J}F(x) \mathrm{d}\mathcal{H}^k(x) < +\infty \ ,$$

or assume that $\varphi : \mathbb{R}^d \to [0, +\infty)$. Then, the following hold:

- *(Area formula) If $d \leq p$ then*

$$\int_{\mathsf{E}} \varphi(x) \mathrm{J}F(x) \mathrm{d}\mathcal{H}^k(x) = \int_{\mathbb{R}^p} \left(\int_{\mathsf{E} \cap F^{-1}(y)} \varphi(x) \mathrm{d}\mathcal{H}^0(x) \right) \mathrm{d}\mathcal{H}^k(y) \ .$$

- *(Coarea formula) If $k \geq p$ then*

$$\int_{\mathsf{E}} \varphi(x) \mathrm{J}F(x) \mathrm{d}\mathcal{H}^k(x) = \int_{\mathbb{R}^p} \left(\int_{\mathsf{E} \cap F^{-1}(y)} \varphi(x) \mathrm{d}\mathcal{H}^{k-p}(x) \right) \mathrm{d}\mathcal{H}^p(y) \ .$$

Proof These results follow from (Ambrosio et al., 2000, Theorem 2.91, Theorem 2.93) combined with (Ambrosio et al., 2000, Exercise 2.12). ■

In particular, note that if $F \in \mathrm{C}^1(\mathbb{R}^d, \mathbb{R}^d)$ is a Lipschitz diffeomorphism we have using Theorem 41 that for any $\varphi : \mathbb{R}^d \to \mathbb{R}$ measurable such that

$$\int_{\mathsf{E}} |\varphi(x)| \, |J\Phi(x)| \, \mathrm{d}\mathcal{H}^k(x) < +\infty \ ,$$

the following change of variable formula with respect to $\mathcal{H}^k$ holds

$$\int_{\mathsf{E}} \varphi(x) \mathrm{J}F(x) \mathrm{d}\mathcal{H}^k(x) = \int_{\mathbb{R}^d} \left(\int_{\mathsf{E} \cap F^{-1}(y)} \varphi(x) \mathrm{d}\mathcal{H}^0(x) \right) \mathrm{d}\mathcal{H}^k(y) = \int_{F(\mathsf{E})} \varphi(F^{-1}(y)) \mathrm{d}\mathcal{H}^k(y) \ . \tag{99}$$

Note that in order for (99) to hold, E needs to be countably $\mathcal{H}^k$-rectifiable.

References

Luigi Ambrosio, Nicola Fusco, and Diego Pallara. *Functions of bounded variation and free discontinuity problems*. Oxford Mathematical Monographs. The Clarendon Press, Oxford University Press, New York, 2000. ISBN 0-19-850245-1.

Christophe Andrieu, Laird A Breyer, and Arnaud Doucet. Convergence of simulated annealing using foster-lyapunov criteria. *Journal of Applied Probability*, 38(4):975–994, 2001.

Vladimir Igorevich Arnold, Aleksandr Nikolaevich Varchenko, and Sabir Medzhidovich Gusein-Zade. *Singularities of Differentiable Maps: Volume II Monodromy and Asymptotic Integrals*, volume 83. Springer Science & Business Media, 2012.

Dominique Bakry, Franck Barthe, Patrick Cattiaux, and Arnaud Guillin. A simple proof of the poincaré inequality for a large class of probability measures. *Electronic Communications in Probability*, 13:60–66, 2008.

Philippe Barbe. Approximation of integrals over asymptotic sets with applications to probability and statistics. *arXiv preprint math/0312132*, 2003.

Carl M Bender and Steven A Orszag. *Advanced mathematical methods for scientists and engineers I: Asymptotic methods and perturbation theory*. Springer Science & Business Media, 2013.

Norman Bleistein and Richard A Handelsman. *Asymptotic expansions of integrals*. Ardent Media, 1975.

Jean-Michel Bony, Fabrizio Broglia, Ferruccio Colombini, and Ludovico Pernazza. Nonnegative functions as squares or sums of squares. *Journal of Functional Analysis*, 232(1): 137–147, 2006.

Olivier Bousquet and André Elisseeff. Stability and generalization. *J. Mach. Learn. Res.*, 2(3):499–526, 2002. ISSN 1532-4435. doi: 10.1162/153244302760200704. URL `https://doi.org/10.1162/153244302760200704`.

Odile Brandière and Marie Duflo. Les algorithmes stochastiques contournent-ils les pièges? *Ann. Inst. H. Poincaré Probab. Statist.*, 32(3):395–427, 1996. ISSN 0246-0203. URL `http://www.numdam.org/item?id=AIHPB_1996__32_3_395_0`.

Pierre Bras and Gilles Pagès. Convergence of Langevin-simulated annealing algorithms with multiplicative noise. *arXiv preprint arXiv:2109.11669*, 2021.

Karl W Breitung. *Asymptotic approximations for probability integrals.* Springer, 2006.

Joan Bruna and Stéphane Mallat. Multiscale sparse microcanonical models. *Mathematical Statistics and Learning*, 1, 01 2018. doi: 10.4171/MSL/7.

Olivier Catoni. The energy transformation method for the Metropolis algorithm compared with simulated annealing. *Probability theory and related fields*, 110(1):69–89, 1998.

Huy N Chau and Miklos Rasonyi. Stochastic Gradient Hamiltonian Monte Carlo for Non-Convex Learning. *arXiv preprint arXiv:1903.10328*, 2019.

Tzuu-Shuh Chiang, Chii-Ruey Hwang, and Shuenn Jyi Sheu. Diffusion for global optimization in $\mathbf{R}^n$. *SIAM Journal on Control and Optimization*, 25(3):737–753, 1987.

Michael CH Choi. An improved variant of simulated annealing that converges under fast cooling. *arXiv preprint arXiv:1901.10269*, 2019.

Michael CH Choi. On the convergence of an improved discrete simulated annealing via landscape modification. *arXiv preprint arXiv:2011.09680*, 2020.

Edmond Combet. *Intégrales exponentielles: développements asymptotiques, propriétés lagrangiennes*, volume 937. Springer, 2006.

Imre Csiszár. I-divergence geometry of probability distributions and minimization problems. *The annals of probability*, pages 146–158, 1975.

Arnak S Dalalyan. Theoretical guarantees for approximate sampling from smooth and log-concave densities. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 79(3):651–676, 2017.

Valentin De Bortoli, Agnès Desolneux, Alain Durmus, Bruno Galerne, and Arthur Leclaire. Maximum entropy methods for texture synthesis: theory and practice. *SIAM Journal on Mathematics of Data Science*, 3(1):52–82, 2021.

Nicolaas Govert De Bruijn. *Asymptotic methods in analysis*, volume 4. Courier Corporation, 1981.

Agnès Desolneux and Arthur Leclaire. Stochastic image reconstruction from local histograms of gradient orientation. In Franccois Lauze, Yiqiu Dong, and Anders Bjorholm Dahl, editors, *Scale Space and Variational Methods in Computer Vision - 6th International Conference, SSVM 2017, Kolding, Denmark, June 4-8, 2017, Proceedings*, volume 10302 of *Lecture Notes in Computer Science*, pages 133–145. Springer, 2017.

Alain Durmus and Eric Moulines. Nonasymptotic convergence analysis for the unadjusted Langevin algorithm. *The Annals of Applied Probability*, 27(3):1551–1587, 2017.

Richard S. Ellis and Jay S. Rosen. Laplace's Method for Gaussian Integrals with an Application to Statistical Mechanics. *The Annals of Probability*, 10(1):47 – 66, 1982. doi: 10.1214/aop/1176993913. URL `https://doi.org/10.1214/aop/1176993913`.

Arthur Erdélyi. *Asymptotic expansions*. Number 3. Courier Corporation, 1956.

Murat A Erdogdu, Lester Mackey, and Ohad Shamir. Global non-convex optimization with discretized diffusions. *arXiv preprint arXiv:1810.12361*, 2018.

Marat Andreevich Evgrafov. *Asymptotic estimates and entire functions*. Courier Dover Publications, 2020.

Herbert Federer. *Geometric measure theory*. Die Grundlehren der mathematischen Wissenschaften, Band 153. Springer-Verlag New York Inc., New York, 1969.

MV Fedoryuk. Asymptotic methods in analysis. In *Analysis I*, pages 83–191. Springer, 1989.

C Fefferman and Duong Hong Phong. On positivity of pseudo-differential operators. *Proceedings of the National Academy of Sciences of the United States of America*, 75(10): 4673, 1978.

Xuefeng Gao, Mert Gürbüzbalaban, and Lingjiong Zhu. Global convergence of stochastic gradient Hamiltonian Monte Carlo for non-convex stochastic optimization: Non-asymptotic performance bounds and momentum-based acceleration. *arXiv preprint arXiv:1809.04618*, 2018.

Leon Gatys, Alexander S Ecker, and Matthias Bethge. Texture synthesis using convolutional neural networks. *Advances in neural information processing systems*, 28, 2015.

Saul B Gelfand and Sanjoy K Mitter. Recursive stochastic algorithms for global optimization in Rˆd. *SIAM Journal on Control and Optimization*, 29(5):999–1018, 1991.

Saul B. Gelfand and Sanjoy K. Mitter. Metropolis-type annealing algorithms for global optimization in $\mathbf{R}^d$. *SIAM J. Control Optim.*, 31(1):111–131, 1993. ISSN 0363-0129. doi: 10.1137/0331009. URL `https://doi.org/10.1137/0331009`.

Stuart Geman and Donald Geman. Stochastic Relaxation, Gibbs Distributions, and the Bayesian Restoration of Images. *IEEE Trans. Pattern Anal. Mach. Intell.*, 6(6):721–741, 1984.

Stuart Geman and Chii-Ruey Hwang. Diffusions for global optimization. *SIAM Journal on Control and Optimization*, 24(5):1031–1043, 1986.

Basilis Gidas. Nonstationary Markov chains and convergence of the annealing algorithm. *J. Statist. Phys.*, 39(1-2):73–131, 1985. ISSN 0022-4715. doi: 10.1007/BF01007975. URL `https://doi.org/10.1007/BF01007975`.

Amos Golan. *Information and entropy econometrics: a review and synthesis*, volume 3. now publishers inc, 2008.

Bruce Hajek. A tutorial survey of theory and applications of simulated annealing. In *1985 24th IEEE Conference on Decision and Control*, pages 755–760. IEEE, 1985.

Bruce Hajek. Cooling schedules for optimal annealing. *Mathematics of operations research*, 13(2):311–329, 1988.

Dominique MA Haughton. On the choice of a model to fit data from an exponential family. *The annals of statistics*, pages 342–355, 1988.

Heisuke Hironaka. Resolution of singularities of an algebraic variety over a field of characteristic zero: Ii. *Annals of Mathematics*, pages 205–326, 1964.

Richard A. Holley, Shigeo Kusuoka, and Daniel W. Stroock. Asymptotics of the spectral gap with applications to the theory of simulated annealing. *J. Funct. Anal.*, 83(2):333–347, 1989. ISSN 0022-1236. doi: 10.1016/0022-1236(89)90023-2. URL `https://doi.org/10.1016/0022-1236(89)90023-2`.

Chii-Ruey Hwang. Laplace's method revisited: Weak convergence of probability measures. *Annals of Probability*, 8(6):1177–1182, 1980.

Nobuyuki Ikeda and Shinzo Watanabe. *Stochastic differential equations and diffusion processes.* Elsevier, 2014.

E. T. Jaynes. Information theory and statistical mechanics. *Phys. Rev.*, 1957.

Scott Kirkpatrick. Optimization by simulated annealing: quantitative studies. *J. Statist. Phys.*, 34(5-6):975–986, 1984. ISSN 0022-4715. doi: 10.1007/BF01009452. URL `https://doi.org/10.1007/BF01009452`.

Dmitry Alekseevich Korshunov, Vladimir Il'ich Piterbarg, and E Hashorva. On the asymptotic Laplace method and its application to random chaos. *Mathematical Notes*, 97(5): 878–891, 2015.

Serge Lang. *Introduction to differentiable manifolds.* Universitext. Springer-Verlag, New York, second edition, 2002. ISBN 0-387-95477-5.

Urs Lang. Notes on rectifiability. *https://people. math. ethz. ch/˜ lang/rect_notes. pdf*, 2007.

Ta Le Loi and Phan Phien. A numerical approach to some basic theorems in singularity theory. *Mathematische Nachrichten*, 287(7):764–781, 2014.

Yang Lu, Song-Chun Zhu, and Ying Nian Wu. Learning FRAME models using CNN filters. In *Proceedings of the Thirtieth AAAI Conference on Artificial Intelligence*, pages 1902–1910, 2016.

J. Milnor. *Morse theory.* Based on lecture notes by M. Spivak and R. Wells. Annals of Mathematics Studies, No. 51. Princeton University Press, Princeton, N.J., 1963.

Frank Morgan. *Geometric measure theory.* Elsevier/Academic Press, Amsterdam, fifth edition, 2016. ISBN 978-0-12-804489-6. A beginner's guide, Illustrated by James F. Bredt.

Than Huy Nguyen, Umut Simsekli, and Gaël Richard. Non-Asymptotic Analysis of Fractional Langevin Monte Carlo for Non-Convex Optimization. In *International Conference on Machine Learning*, pages 4810–4819. PMLR, 2019.

Louis Nirenberg. *Topics in nonlinear functional analysis*, volume 6 of *Courant Lecture Notes in Mathematics.* New York University, Courant Institute of Mathematical Sciences, New York; American Mathematical Society, Providence, RI, 2001. ISBN 0-8218-2819-3. doi: 10.1090/cln/006. URL `https://doi.org/10.1090/cln/006`. Chapter 6 by E. Zehnder, Notes by R. A. Artino, Revised reprint of the 1974 original.

Frank Olver. *Asymptotics and special functions.* CRC Press, 1997.

Mariane Pelletier. Weak convergence rates for stochastic approximation with application to multiple targets and simulated annealing. *Ann. Appl. Probab.*, 8(1):10–44, 1998. ISSN 1050-5164. doi: 10.1214/aoap/1027961032. URL `https://doi.org/10.1214/aoap/1027961032`.

Javier Portilla and Eero P. Simoncelli. A parametric texture model based on joint statistics of complex wavelet coefficients. *Int. J. Comput. Vis.*, 40(1):49–70, 2000.

Maxim Raginsky, Alexander Rakhlin, and Matus Telgarsky. Non-convex learning via stochastic gradient langevin dynamics: a nonasymptotic analysis. In *Conference on Learning Theory*, pages 1674–1703. PMLR, 2017.

Gareth O Roberts and Richard L Tweedie. Exponential convergence of Langevin distributions and their discrete approximations. *Bernoulli*, pages 341–363, 1996.

H Edwin Romeijn and Robert L Smith. Simulated annealing for constrained global optimization. *Journal of Global Optimization*, 5(2):101–126, 1994.

Håvard Rue, Sara Martino, and Nicolas Chopin. Approximate Bayesian inference for latent Gaussian models by using integrated nested Laplace approximations. *Journal of the royal statistical society: Series b (statistical methodology)*, 71(2):319–392, 2009.

Anastasiia Igorevna Rytova and Elena Borisovna Yarovaya. Multidimensional Watson lemma and its applications. *Mathematical Notes*, 99(3):406–412, 2016.

Luke Tierney and Joseph B Kadane. Accurate approximations for posterior moments and marginal densities. *Journal of the american statistical association*, 81(393):82–86, 1986.

Cédric Villani. *Optimal transport*, volume 338 of *Grundlehren der Mathematischen Wissenschaften [Fundamental Principles of Mathematical Sciences].* Springer-Verlag, Berlin, 2009. ISBN 978-3-540-71049-3. doi: 10.1007/978-3-540-71050-9. URL `https://doi.org/10.1007/978-3-540-71050-9`. Old and new.

H Wang, HJ Lin, WT Cai, LZ Ouyang, and M Zhu. Tuning kinetics and thermodynamics of hydrogen storage in light metal element based systems–a review of recent progress. *Journal of Alloys and Compounds*, 658:280–300, 2016.

Max Welling and Yee W Teh. Bayesian learning via stochastic gradient Langevin dynamics. In *Proceedings of the 28th international conference on machine learning (ICML-11)*, pages 681–688. Citeseer, 2011.

Hassler Whitney. Analytic extensions of differentiable functions defined in closed sets. *Transactions of the American Mathematical Society*, 36(1):63–89, 1934.

R. Wong. *Asymptotic approximations of integrals*, volume 34 of *Classics in Applied Mathematics*. Society for Industrial and Applied Mathematics (SIAM), Philadelphia, PA, 2001. ISBN 0-89871-497-4. doi: 10.1137/1.9780898719260. URL `https://doi.org/10.1137/1.9780898719260`. Corrected reprint of the 1989 original.

Pan Xu, Jinghui Chen, Difan Zou, and Quanquan Gu. Global convergence of Langevin dynamics based algorithms for nonconvex optimization. *arXiv preprint arXiv:1707.06618*, 2017.

R. L. Yang. Convergence of the simulated annealing algorithm for continuous global optimization. *J. Optim. Theory Appl.*, 104(3):691–716, 2000. ISSN 0022-3239. doi: 10.1023/A:1004697811243. URL `https://doi.org/10.1023/A:1004697811243`.

Nanyang Ye, Zhanxing Zhu, and Rafal K Mantiuk. Langevin dynamics with continuous tempering for training deep neural networks. *arXiv preprint arXiv:1703.04379*, 2017.

Ying Zhang, Ömer Deniz Akyildiz, Theodoros Damoulas, and Sotirios Sabanis. Nonasymptotic estimates for Stochastic Gradient Langevin Dynamics under local conditions in nonconvex optimization. *arXiv preprint arXiv:1910.02008*, 2019.

Yuchen Zhang, Percy Liang, and Moses Charikar. A hitting time analysis of stochastic gradient Langevin dynamics. In *Conference on Learning Theory*, pages 1980–2022. PMLR, 2017.

Brian D. Ziebart, Andrew L. Maas, J. Andrew Bagnell, and Anind K. Dey. Maximum entropy inverse reinforcement learning. In Dieter Fox and Carla P. Gomes, editors, *Proceedings of the Twenty-Third AAAI Conference on Artificial Intelligence, AAAI 2008, Chicago, Illinois, USA, July 13-17, 2008*, pages 1433–1438. AAAI Press, 2008. URL `http://www.aaai.org/Library/AAAI/2008/aaai08-227.php`.