The Annals of Probability
2014, Vol. 42, No. 5, 1809–1884
DOI: 10.1214/14-AOP908

POINCARÉ AND LOGARITHMIC SOBOLEV INEQUALITIES BY DECOMPOSITION OF THE ENERGY LANDSCAPE

BY GEORG MENZ AND ANDRÉ SCHLICHTING[1]

Stanford University and University of Bonn

We consider a diffusion on a potential landscape which is given by a smooth Hamiltonian $H : \mathbb{R}^n \to \mathbb{R}$ in the regime of low temperature ε. We proof the Eyring–Kramers formula for the optimal constant in the Poincaré (PI) and logarithmic Sobolev inequality (LSI) for the associated generator $L = \varepsilon\Delta - \nabla H \cdot \nabla$ of the diffusion. The proof is based on a refinement of the two-scale approach introduced by Grunewald et al. [*Ann. Inst. Henri Poincaré Probab. Stat.* **45** (2009) 302–351] and of the mean-difference estimate introduced by Chafaï and Malrieu [*Ann. Inst. Henri Poincaré Probab. Stat.* **46** (2010) 72–96]. The Eyring–Kramers formula follows as a simple corollary from two main ingredients: The first one shows that the PI and LSI constant of the diffusion restricted to metastable regions corresponding to the local minima scales well in ε. This mimics the fast convergence of the diffusion to metastable states. The second ingredient is the estimation of a mean-difference by a weighted transport distance. It contains the main contribution to the PI and LSI constant, resulting from exponentially long waiting times of jumps between metastable states of the diffusion.

1. Introduction. Let us consider a diffusion on a potential landscape which is given by a sufficiently smooth *Hamiltonian function* $H : \mathbb{R}^n \to \mathbb{R}$. We are interested in the regime of low temperature $\varepsilon > 0$. The *generator* of the diffusion has the following form:

$$L := \varepsilon\Delta - \nabla H \cdot \nabla. \tag{1.1}$$

The associated *Dirichlet form* is given for a test function $f \in H^1(\mu)$ by

$$\mathcal{E}(f) := \int (-Lf) f \,\mathrm{d}\mu = \varepsilon \int |\nabla f|^2 \,\mathrm{d}\mu.$$

The corresponding diffusion ξ_t satisfies the stochastic differential equation

$$\mathrm{d}\xi_t = -\nabla H(\xi_t)\,\mathrm{d}t + \sqrt{2\varepsilon}\,\mathrm{d}B_t, \tag{1.2}$$

where B_t is the *Brownian motion* on $\mathbb{R}^n$. Equation (1.2) is also called *over-damped Langevin equation* (cf., e.g., [32]). Under some growth assumption on H, there

Received July 2012; revised December 2013.

[1]Supported by DFG project FOR718, the IMPRS of the Max-Planck-Institute and the University of Leipzig.

MSC2010 subject classifications. Primary 60J60; secondary 35P15, 49R05.

Key words and phrases. Diffusion process, Eyring–Kramers formula, Kramers law, metastability, Poincaré inequality, spectral gap, logarithmic Sobolev inequality, weighted transport distance.

exists an equilibrium measure of the according stochastic process, which is called *Gibbs measure* and is given by

$$\mu(\mathrm{d}x) = \frac{1}{Z_\mu} \exp\left(-\frac{H(x)}{\varepsilon}\right) \mathrm{d}x \qquad \text{with } Z_\mu = \int \exp\left(-\frac{H(x)}{\varepsilon}\right) \mathrm{d}x. \tag{1.3}$$

The evolution (1.2) of the stochastic process ξ_t can be translated into an evolution of the density of the process ξ_t. Namely, under the assumption that the law of the initial state ξ_0 is absolutely continuous w.r.t. the Gibbs measure μ, the density $f_t \mu$ of the process ξ_t satisfies the *Fokker–Planck equation* (cf., e.g., [37] or [44])

$$\partial_t f_t = L f_t = \varepsilon \Delta f_t - \nabla H \cdot \nabla f_t.$$

We are particularly interested in the case where H has several local minima. Then for small ε, the process shows metastable behavior in the sense that there exists a separation of scales: On the fast scale, the process converges quickly to a neighborhood of a local minimum. On the slow scale, the process stays nearby a local minimum for an exponentially long waiting time after which it eventually jumps to another local minimum.

This behavior was first described in the context of chemical reactions. The exponential waiting time follows the *Arrhenius' law* [1] meaning that the mean exit time from one local minimum of H to another one is exponentially large in the energy barrier between them. By now, the Arrhenius law is well understood even for nonreversible systems by the *Freidlin–Wentzell theory* [19], which is based on *large deviations*.

A refinement of the Arrhenius law is the *Eyring–Kramers formula* which additionally considers *pre-exponential* factors. The Eyring–Kramers formula for the Poincaré inequality (PI) goes back to Eyring [18] and Kramers [30]. Both argue that also in high-dimensional problems of chemical reactions most reactions are nearby a single trajectory called *reaction pathway*. Evaluating the Hamiltonian along this *reaction coordinate* gives the classical picture of a double well potential (cf. Figure 1) in one dimension with an *energy barrier* separating the two local minima for which explicit calculations are feasible.

However, a rigorous proof of the Eyring–Kramers formula for the multidimensional case was open for a long time. For a special case, where all the minima of the potential as well as all the lowest saddle points in-between have the same energy, Sugiura [45] defined an exponentially rescaled Markov chain on the set of minima in such a way that the preexponential factors become the transitions rates between the metastable regions of the rescaled process. For the generic case, where the local minima and saddles have different energies, the group of Bovier et al. [9, 10] obtained first-order asymptotics that are sharp in the parameter ε. They also clarified the close connection between *mean exit times*, *capacities* and the exponentially small eigenvalues of the operator L given by (1.1). The main tool of [9, 10] is potential theory. The small eigenvalues are related to the mean exit times of appropriate subsets of the state space. Further, the mean exit times are given

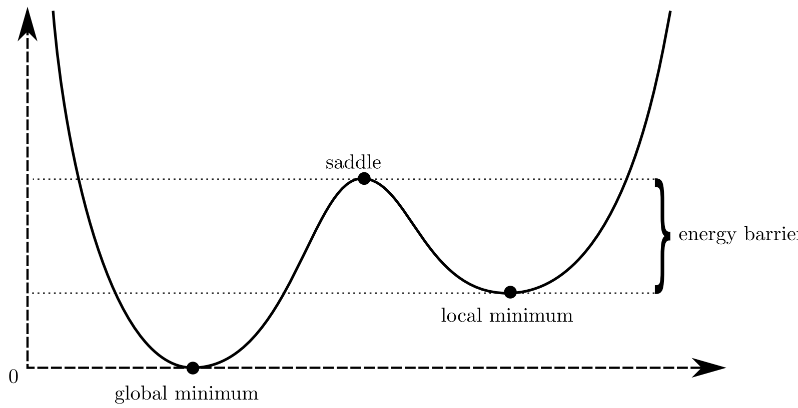

FIG. 1. *General double-well potential H on $\mathbb{R}$.*

by *Newtonian capacities* which can explicitly be calculated in the regime of low temperature ε.

Shortly after, Helffer, Klein and Nier [23–25] also deduced the Eyring–Kramers formula using the connection of the spectral gap estimate of the Fokker–Planck operator L given by (1.1) to the one of the *Witten Laplacian*. This approach makes it possible to get quantitative results with the help of semiclassical analysis. They deduced sharp asymptotics of the exponentially small eigenvalues of L and gave an explicit expansion in ε to theoretically any order. An overview on the Eyring–Kramers formula can be found in the review article of Berglund [6].

In this work, we provide a new proof of the Eyring–Kramers formula for the first eigenvalue of the operator L, that is, its *spectral gap*. The advantage of this new approach is that it extends to the logarithmic Sobolev inequality (LSI), which was not investigated before. The LSI was introduced by [21] and is stronger than the PI. Therefore, the LSI is usually harder to deduce than the PI due to its nonlinear structure.

By deducing the Eyring–Kramers formula for the LSI, we encounter a surprising effect: In the generic situation of having two local minima with different energies, the Eyring–Kramers formula for the LSI differs from the Eyring–Kramers formula for the PI by a term of inverse order in ε. However, in the symmetric situation of having local minima with the same energy, the Eyring–Kramers formula for the LSI coincides with the corresponding formula for the PI (cf. Corollary 2.18).

We conclude the Introduction with an overview of the article:

In Section 1.1, we introduce PI and LSI.
In Section 1.2, we discuss the setting and the assumptions on the Hamiltonian H.
In Section 2, we outline the new approach and state the main results of this work.
In Section 3 and Section 4, we proof the main ingredients of our new approach. Namely, in Section 3, we deduce a local PI and a local LSI with optimal scaling

in ε, whereas in Section 4 we estimate a mean-difference by using a weighted transport distance.
In the Appendices, we provide for the convenience of the reader some basic but nonstandard facts that are used in our arguments.

1.1. *Poincaré and logarithmic Sobolev inequality.*

DEFINITION 1.1 [PI(ϱ) and LSI(α)]. Let X be an Euclidean space. A Borel probability measure μ on X satisfies the *Poincaré inequality* with constant $\varrho > 0$, if for all test functions $f \in H^1(\mu)$

$$\text{var}_\mu(f) := \int \left(f - \int f \, \mathrm{d}\mu \right)^2 \mathrm{d}\mu \le \frac{1}{\varrho} \int |\nabla f|^2 \, \mathrm{d}\mu. \tag{PI(ϱ)}$$

In a similar way, the probability measure μ satisfies the *logarithmic Sobolev inequality* with constant $\alpha > 0$, if for all test function $f : X \to \mathbb{R}^+$ with $I(f\mu|\mu) < \infty$ holds

$$\text{Ent}_\mu(f) := \int f \log \frac{f}{\int f \, \mathrm{d}\mu} \, \mathrm{d}\mu \le \frac{1}{\alpha} \int \frac{|\nabla f|^2}{2f} \, \mathrm{d}\mu =: I(f\mu|\mu), \tag{LSI(α)}$$

where $I(f\mu|\mu)$ is called *Fisher information*. The gradient ∇ is determined by the Euclidean structure of X. Test functions are those functions for which the gradient exists and the right-hand side in PI(ϱ) and LSI(α) is well defined.

REMARK 1.2 [Relation between PI(ϱ) and LSI(α)]. Rothaus [41] observed that LSI(α) implies PI(α). This can be seen by setting $f = 1 + \eta g$ for η small and observing that

$$\text{Ent}_\mu(f^2) = 2\eta^2 \, \text{var}_\mu(g) + O(\eta^3) \quad \text{as well as} \quad \int |\nabla f|^2 \, \mathrm{d}\mu = \eta^2 \int |\nabla g|^2 \, \mathrm{d}\mu.$$

Hence, if μ satisfies LSI(α) then μ satisfies PI(α), which always implies $\alpha \le \varrho$.

1.2. *Setting and assumptions.* This article uses almost the same setting as found in [9, 10]. Before stating the precise assumptions on the Hamiltonian H, we introduce the notion of a *Morse function*.

DEFINITION 1.3 (Morse function). A smooth function $H : \mathbb{R}^n \to \mathbb{R}$ is a *Morse function*, if the Hessian $\nabla^2 H$ of H is nondegenerated on the set of critical points. More precisely, for some $1 \le C_H < \infty$ holds

$$\forall x \in \mathcal{S} := \{ x \in \mathbb{R}^n : \nabla H = 0 \} : \frac{|\xi|^2}{C_H} \le \langle \xi, \nabla^2 H(x) \xi \rangle \le C_H |\xi|^2. \tag{1.4}$$

We make the following growth assumption on the Hamiltonian H sufficient to ensure the existence of PI and LSI. Hereby, we have to assume stronger properties for H if we want to proof the LSI.

ASSUMPTION 1.4 (PI). $H \in C^3(\mathbb{R}^n, \mathbb{R})$ is a nonnegative Morse function, such that for some constants $C_H > 0$ and $K_H \geq 0$ holds

$$\liminf_{|x|\to\infty} |\nabla H| \geq C_H, \tag{A1$_{\mathtt{PI}}$}$$

$$\liminf_{|x|\to\infty} \big(|\nabla H|^2 - \Delta H\big) \geq -K_H. \tag{A2$_{\mathtt{PI}}$}$$

ASSUMPTION 1.5 (LSI). $H \in C^3(\mathbb{R}^n, \mathbb{R})$ is a nonnegative Morse function, such that for some constants $C_H > 0$ and $K_H \geq 0$ holds

$$\liminf_{|x|\to\infty} \frac{|\nabla H(x)|^2 - \Delta H(x)}{|x|^2} \geq C_H, \tag{A1$_{\mathtt{LSI}}$}$$

$$\inf_x \nabla^2 H(x) \geq -K_H. \tag{A2$_{\mathtt{LSI}}$}$$

REMARK 1.6 (Discussion of assumptions). The Assumption 1.4 yields the following consequences for the Hamiltonian H:

- The condition (A1$_{\mathtt{PI}}$) and $H(x) \geq 0$ ensures that e^{-H} is integrable and can be normalized to a probability measure on $\mathbb{R}^n$ (see Lemma 3.14). Hence, the Gibbs measure μ given by (1.3) is well defined.
- The Morse Assumption (1.4) together with the growth condition (A1$_{\mathtt{PI}}$) ensures that the set $\mathcal{S}$ of critical points is discrete and finite. In particular, it follows that the set of local minima $\mathcal{M} = \{m_1, \ldots, m_M\}$ is also finite, that is, $M := \#\mathcal{M} < \infty$.
- The *Lyapunov-type* condition (A2$_{\mathtt{PI}}$) allows to recover the Poincaré constant of the full Gibbs measure μ from the Poincaré constant of the Gibbs measure μ_U restricted to some bounded domain U (cf. Section 3). Because Gibbs measures with finite support and smooth Hamiltonian always satisfy a Poincaré inequality with some unspecified constant, we get that the Gibbs measure μ also satisfies a Poincaré inequality. Equivalently, this means that there exists a spectral gap for the operator L given by (1.1).

Similarly the Assumption 1.5 has the following consequences for the Hamiltonian H:

- One difference between the Assumptions 1.4 and 1.5 is that (A1$_{\mathtt{PI}}$) yields linear growth at infinity for H, whereas a combination of condition (A1$_{\mathtt{LSI}}$) and (A2$_{\mathtt{LSI}}$) yields quadratic growth; that is,

$$\liminf_{|x|\to\infty} \frac{|\nabla H(x)|}{|x|} \geq C_H. \tag{A0$_{\mathtt{LSI}}$}$$

  Note that quadratic growth at infinity is a necessary condition to obtain LSI(α) with $\alpha > 0$ (cf. [42], Theorem 3.1.21).

- In addition, ($\mathrm{A1}_{\mathtt{LSI}}$) and ($\mathrm{A2}_{\mathtt{LSI}}$) imply ($\mathrm{A1}_{\mathtt{PI}}$) and ($\mathrm{A2}_{\mathtt{PI}}$), which is only an indication that LSI(α) is stronger than PI(ϱ) in the sense of Remark 1.2.
- The condition ($\mathrm{A1}_{\mathtt{LSI}}$) is again a *Lyapunov type* condition. To enforce it to a LSI, additionally the condition ($\mathrm{A2}_{\mathtt{LSI}}$) has to be assumed (cf. Section 3).

To keep the presentation feasible and clear, we additionally assume a nondegeneracy assumption, even if it is not really needed for the proof of the Eyring–Kramers formula. The saddle height $\widehat{H}(m_i, m_j)$ between two local minima m_i, m_j is defined by

$$\widehat{H}(m_i, m_j) := \inf\Big\{ \max_{s\in[0,1]} H\big(\gamma(s)\big) : \gamma \in C\big([0,1], \mathbb{R}^n\big), \gamma(0) = m_i, \gamma(1) = m_j \Big\}.$$

ASSUMPTION 1.7 (Nondegeneracy). There exists $\delta > 0$ such that:

(i) The saddle height between two local minima m_i, m_j is attained at a unique critical point $s_{i,j} \in \mathcal{S}$ of index one, that is, it holds $H(s_{i,j}) = \widehat{H}(m_i, m_j)$ and if $\{\lambda_1, \dots, \lambda_n\}$ denote the eigenvalues of $\nabla^2 H(s_{i,j})$, then it holds $\lambda_1 < 0$ and $\lambda_i > 0$ for $i = 2, \dots, n$. The point $s_{i,j}$ is called *communicating saddle* between the minima m_i and m_j.

(ii) The set of local minima $\mathcal{M} = \{m_1, \dots, m_M\}$ is ordered such that m_1 is a global minimum and for all $i \in \{3, \dots, M\}$ yields

$$H(s_{1,2}) - H(m_2) \geq H(s_{1,i}) - H(m_i) + \delta.$$

REMARK 1.8. The fact, that $s_{i,j}$ is indeed a critical point is explained in [29], Proposition 6.2.1. Since H is a Morse function after Assumption 1.4 the critical point $s_{i,j}$ is nondegenerate. Moreover, an indirect perturbation argument implies that $s_{i,j}$ is a saddle point of index one, which shows that except for uniqueness, Assumption 1.7(i) is already implied by Assumption 1.4. This fact is known as Murrell–Laidler theorem in the chemical literature [47].

2. Outline of the new approach and main results. In this section, we present the new approach to the Eyring–Kramers formula and formulate the main results of this article. Because the strategy is the same for the PI and LSI, we consider both cases simultaneously. The approach uses ideas of the two-scale approach for LSI [22, 33, 39] and the method by [14] to deduce PI and LSI estimates for mixtures of measures. However, the heuristics outlined in the Introduction provide a good orientation for our proceeding. Remember that we have a splitting into two time-scales:

- the fast scale describes the *fast relaxation* to a local minima of H and
- the slow scale describes the *exponentially long transitions* between local equilibrium states.

Motivated by these two time scales, we specify in Section 2.1 a splitting of the measure μ into *local measures* living on a *metastable regions* around the local minima of H. This splitting is lifted from the level of the measure to the level of the variance and entropy. In this way, we obtain *local variances and entropies*, which heuristically should correspond to the fast relaxation, and *coarse-grained variances and entropies*, which should correspond to the exponentially long transitions.

Now, we handle each contribution separately. The local variances and entropies are estimated by local PI (cf. Theorem 2.9) and local LSI, respectively (cf. Theorem 2.10). The heuristics suggest that this contribution should be of higher order because this step only relies on the fast scale.

Before we estimate the coarse-grained variances and entropies, we bring them in the form of mean-differences. This is automatically the case for the variances. However, for the coarse-grained entropies one has to apply a new weighted discrete LSI (cf. Section 2.2), which causes the difference between the PI and LSI in the Eyring–Kramers formula. The main contribution to the Eyring–Kramers formula (cf. Corollary 2.15 and Corollary 2.17) results from the estimation of the mean-difference, which is stated in Theorem 2.12.

At this point, let us shortly summarize the main results of this article:

- We provide good estimates for the local variances and entropies (cf. Section 2.3.1) and
- We provide sharp estimates for the mean-differences (cf. Section 2.3.2).
- From these main ingredients, the Eyring–Kramers formulas follow as simple corollaries (cf. Section 2.3.3).

We close this chapter with a discussion on the optimality of the Eyring–Kramers formula for the LSI in one dimension (cf. Section 2.4).

Notational remark: Almost all of the following definitions and quantities will depend on ε, for lucidity this dependence is not expressed in the notation. The arguments and main results hold for $\varepsilon > 0$ fixed and small.

2.1. *Partition of the state space*. The inspiration to partition the state space comes from the work [28] for discrete Markov chains. In order to get sharp results, the partition of the state space $\mathbb{R}^n$ cannot be arbitrarily but has to satisfy certain conditions.

DEFINITION 2.1 (Admissible partition). The family $\mathcal{P}_{\mathcal{M}} = \{\Omega_i\}_{i=1}^M$ with Ω_i open and connected is called an admissible partition for μ if the following conditions hold:

(i) For each local minimum $m_i \in \mathcal{M}$ exists $\Omega_i \in \mathcal{P}_{\mathcal{M}}$ with $m_i \in \Omega_i$ for $i = 1, \dots, M$.

(ii) $\{\Omega_i\}_{i=1}^M$ is a partition of $\mathbb{R}^n$ up to sets of Lebesgue measure zero, which is denoted by $\mathbb{R}^n = \biguplus_{i=1}^M \Omega_i$.

(iii) The partition sum of each element Ω_i of $\mathcal{P}_{\mathcal{M}}$ is approximately Gaussian, that is, for $i = 1, \dots, M$

$$\mu(\Omega_i) Z_\mu = \frac{(2\pi\varepsilon)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\varepsilon}\right)\left(1 + O\left(\sqrt{\varepsilon}|\log\varepsilon|^{3/2}\right)\right). \tag{2.1}$$

REMARK 2.2. A canonical way to obtain an admissible partition for μ would be to associate to every local minimum $m_i \in \mathcal{M}$ for $i = 1, \dots, M$ its *basin of attraction* Ω_i *w.r.t.* H defined by

$$\Omega_i := \left\{ y \in \mathbb{R}^n : \lim_{t\to\infty} y_t = m_i, \dot{y}_t = -\nabla H(y_t), y_0 = y \right\}.$$

Unfortunately, this choice would lead to technical difficulties later on. We get rid of these technical problems by choosing the partition Ω_i in a slightly different way. For details, we refer the reader to Section 3.

Using an admissible partition of the state space, one can decompose the Gibbs measure μ into a mixture of *local* Gibbs measures μ_i.

DEFINITION 2.3 (Mixture representation of μ). Let $\mathcal{P}_{\mathcal{M}} = \{\Omega_i\}_{i=1}^M$ be an admissible partition for μ. The *local Gibbs measures* μ_i are defined as the restriction of μ to Ω_i

$$\mu_i(\mathrm{d}x) := \frac{1}{Z_i Z_\mu} \mathbb{1}_{\Omega_i}(x) \exp\left(-\frac{H(x)}{\varepsilon}\right) \mathrm{d}x \qquad \text{where } Z_i := \mu(\Omega_i). \tag{2.2}$$

The *marginal measure* $\bar{\mu}$ is given by a sum of Dirac measures

$$\bar{\mu} := Z_1\delta_1 + \cdots + Z_M\delta_M.$$

Then the *mixture representation* of μ w.r.t. $\mathcal{P}_{\mathcal{M}}$ has the form

$$\mu := Z_1\mu_1 + \cdots + Z_M\mu_M. \tag{2.3}$$

As was shown in [14], Section 4.1, the decomposition of μ yields a decomposition of the variance $\operatorname{var}_\mu(f)$ and entropy $\operatorname{Ent}_\mu(f)$.

LEMMA 2.4 (Splitting of variance and entropy for partition). *For a mixture representation* (2.3) *of* μ *holds for all* $f : \mathbb{R}^n \to \mathbb{R}$

$$\operatorname{var}_\mu(f) = \sum_{i=1}^M Z_i \operatorname{var}_{\mu_i}(f) + \sum_{i=1}^M \sum_{j>i} Z_i Z_j \left(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\right)^2, \tag{2.4}$$

$$\operatorname{Ent}_\mu(f) = \sum_{i=1}^M Z_i \operatorname{Ent}_{\mu_i}(f) + \operatorname{Ent}_{\bar{\mu}}(\bar{f}). \tag{2.5}$$

We call the terms $\operatorname{var}_{\mu_i}(f)$ *and* $\operatorname{Ent}_{\mu_i}(f)$ local variance *and* local entropy. *The term* $(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f))^2$ *is called* mean-difference. *The term* $\operatorname{Ent}_{\bar{\mu}}(\bar{f})$ *is called* coarse-grained entropy *and is given by*

$$\operatorname{Ent}_{\bar{\mu}}(\bar{f}) := \sum_{i=1}^{M} Z_i \bar{f}_i \log \frac{\bar{f}_i}{\sum_{j=1}^{M} Z_j \bar{f}_j}, \tag{2.6}$$

where $\bar{f}_i := \mathbb{E}_{\mu_i}(f)$.

We skip the proof of Lemma 2.4 because it only consists of a straightforward substitution of the mixture representation (2.3). The formula (2.4) for estimating the variance $\operatorname{var}_\mu(f)$ is already in its final form. For the relative entropy $\operatorname{Ent}_\mu(f)$, we still have to do some work. The aim is to get an estimate that only involves the local terms like $\operatorname{var}_\mu(f)$ and $\operatorname{Ent}_{\mu_i}(f)$ and a mean difference $(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f))^2$. This is achieved in the next subsection [cf. Corollary 2.8 and (2.13)].

2.2. *Discrete logarithmic Sobolev type inequalities*. Starting with the identity (2.5), we have to estimate the *coarse-grained entropy* $\operatorname{Ent}_{\bar{\mu}}(\bar{f})$. We expect that the main contribution comes from this term. If H has only two minima, we can use the following discrete LSI for a Bernoulli random variable, which was given by Higuchi and Yoshida [26] and Diaconis and Saloff-Coste [15], Theorem A.2, at the same time.

LEMMA 2.5 (Optimal logarithmic Sobolev inequality for Bernoulli measures). *A Bernoulli measure μ_p on $X = \{0, 1\}$, that is, a mixture of two Dirac measures $\mu_p = p\delta_0 + q\delta_1$ with $p + q = 1$ satisfies the discrete logarithmic Sobolev inequality*

$$\operatorname{Ent}_{\mu_p}(f^2) \le \frac{pq}{\Lambda(p,q)} \big(f(0) - f(1)\big)^2 \tag{2.7}$$

with optimal constant given by the logarithmic mean (*cf. Appendix* A)

$$\Lambda(p,q) := \frac{p-q}{\log p - \log q} \qquad \textit{for } p \neq q \quad \textit{and} \quad \Lambda(p,p) := \lim_{q\to p} \Lambda(p,q) = p.$$

We want to handle the general case with more than two minima. Therefore, we want to generalize Lemma 2.5 to discrete measures with a state space with more than two elements. An application of the modified LSI for finite Markov chains of Diaconis and Saloff-Coste [15], Theorem A.1, would not lead to an optimal results (cf. [43], Section 2.3). Even for a generic Markov chain on the 3-point space, the optimal logarithmic Sobolev constant is unknown. In this work, we use the following direct generalization of Lemma 2.5.

LEMMA 2.6 (Weighted logarithmic Sobolev inequality). *For $m \in \mathbb{N}$ let $\mu_m = \sum_{i=1}^{m} Z_i \delta_i$ be a discrete probability measure and assume that* $\min_i Z_i > 0$. *Then for a function $f : \{1, \dots, m\} \to \mathbb{R}_0^+$ holds the weighted logarithmic Sobolev in-*

equality

$$\operatorname{Ent}_{\mu_m}(f^2) \le \sum_{i=1}^{m-1} \sum_{j=i+1}^{m} \frac{Z_i Z_j}{\Lambda(Z_i, Z_j)}(f_i - f_j)^2. \tag{2.8}$$

PROOF. We conclude by induction and find that for $m = 2$ the estimate (2.8) just becomes (2.7), which shows the base case. For the inductive step, let us assume that (2.8) holds for $m \ge 2$. Then the entropy $\operatorname{Ent}_{\mu_{m+1}}(f^2)$ can be rewritten as follows:

$$\operatorname{Ent}_{\mu_{m+1}}(f^2) = (1 - Z_{m+1}) \operatorname{Ent}_{\tilde{\mu}_m}(f^2) + \operatorname{Ent}_\nu(\tilde{f}),$$

where the probability measure $\tilde{\mu}_m$ lives on $\{1, \dots, m\}$ and is given by

$$\tilde{\mu}_m := \sum_{i=1}^{m} \frac{Z_i}{1 - Z_{m+1}} \delta_i.$$

Further, ν is the Bernoulli measure given by $\nu := (1 - Z_{m+1})\delta_0 + Z_{m+1}\delta_1$ and the function $\tilde{f} : \{0, 1\} \to \mathbb{R}$ is given with values

$$\tilde{f}_0 := \sum_{i=1}^{m} \frac{Z_i f_i^2}{1 - Z_{m+1}} \quad \text{and} \quad \tilde{f}_1 := f_{m+1}^2.$$

Now, we apply the inductive hypothesis to $\operatorname{Ent}_{\tilde{\mu}_m}(f^2)$ and arrive at

$$\begin{aligned}(1 - Z_{m+1}) \operatorname{Ent}_{\tilde{\mu}_m}(f^2) &\le (1 - Z_{m+1}) \sum_{i=1}^{m} \sum_{j>i} \frac{Z_i Z_j}{(1 - Z_{m+1})^2} \frac{1 - Z_{m+1}}{\Lambda(Z_i, Z_j)} (f_i - f_j)^2 \\ &= \sum_{i=1}^{m} \sum_{j>i} \frac{Z_i Z_j}{\Lambda(Z_i, Z_j)} (f_i - f_j)^2,\end{aligned}$$

where we used $\Lambda(\cdot, \cdot)$ being homogeneous of degree one in both arguments (cf. Appendix A), that is, $\Lambda(\lambda a, \lambda b) = \lambda \Lambda(a, b)$ for $\lambda, a, b > 0$. We can apply the inductive base to the second entropy $\operatorname{Ent}_\nu(\tilde{f})$, which is nothing else but the discrete LSI for the two-point space (2.7)

$$\operatorname{Ent}_\nu(\tilde{f}) \le \frac{Z_{m+1}(1 - Z_{m+1})}{\Lambda(Z_{m+1}, 1 - Z_{m+1})} \left(\sqrt{\tilde{f}_0} - \sqrt{\tilde{f}_1}\right)^2. \tag{2.9}$$

The last step is to apply the Jensen inequality to recover the square differences $(f_i - f_{m+1})^2$ from

$$\begin{aligned}\left(\sqrt{\tilde{f}_0} - \sqrt{\tilde{f}_1}\right)^2 &= \sum_{i=1}^{m} \frac{Z_i f_i^2}{1 - Z_{m+1}} - 2 \underbrace{\sqrt{\sum_{i=1}^{m} \frac{Z_i f_i^2}{1 - Z_{m+1}}}}_{\ge \sum_{i=1}^{m} \frac{Z_i f_i}{1 - Z_{m+1}}} f_{m+1} + f_{m+1}^2 \\ &\le \sum_{i=1}^{m} \frac{Z_i}{1 - Z_{m+1}} (f_i - f_{m+1})^2.\end{aligned}$$

We obtain in combination with (2.9) the following estimate:

$$\operatorname{Ent}_{\nu}(\tilde{f}) \le \frac{Z_{m+1}}{\Lambda(Z_{m+1}, 1 - Z_{m+1})} \sum_{i=1}^{m} Z_i (f_i - f_{m+1})^2.$$

To conclude the assertion, we first note that $1 - Z_{m+1} = \sum_{j=1}^{m} Z_j \ge Z_j$ for $j = 1, \dots, m$. Further, $\Lambda(a, \cdot)$ is monotone increasing for $a > 0$, that is, $\partial_b \Lambda(a, b) > 0$ (cf. Appendix A). Both properties imply that $\Lambda(Z_{m+1}, 1 - Z_{m+1}) \ge \Lambda(Z_{m+1}, Z_j)$ for $j = 1, \dots, m$, which finally shows (2.8). □

With the help of Lemma 2.6 we estimate the coarse-grained entropy $\operatorname{Ent}_{\bar{\mu}}(\overline{f^2})$ occurring in the splitting of the entropy (2.5). This generalizes the approach of [14], Section 4.1, to the case of finite mixtures with more than two components.

LEMMA 2.7 (Estimate of the coarse-grained entropy). *The coarse-grained entropy in* (2.6) *can be estimated by*

$$\begin{aligned}
&\operatorname{Ent}_{\bar{\mu}}(\overline{f^2}) \\
&\quad \le \sum_{i=1}^{M} \biggl(\sum_{j \ne i} \frac{Z_i Z_j \operatorname{var}_{\mu_i}(f)}{\Lambda(Z_i, Z_j)} + \sum_{j > i} \frac{Z_i Z_j}{\Lambda(Z_i, Z_j)} \bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\bigr)^2 \biggr),
\end{aligned} \tag{2.10}$$

where $\overline{f^2} : \{1, \dots, M\} \to \mathbb{R}$ *is given by* $\overline{f_i^2} := \mathbb{E}_{\mu_i}(f^2)$.

PROOF. Since $\bar{\mu} = Z_1 \delta_1 + \dots + Z_M \delta_M$ is finite discrete probability measure, we can apply Lemma 2.6 to $\operatorname{Ent}_{\bar{\mu}}(\overline{f^2})$

$$\operatorname{Ent}_{\bar{\mu}}(\overline{f^2}) \le \sum_{i=1}^{m} \sum_{j > i} \frac{Z_i Z_j}{\Lambda(Z_i, Z_j)} \Bigl(\sqrt{\overline{f_i^2}} - \sqrt{\overline{f_j^2}} \Bigr)^2. \tag{2.11}$$

The square-root-mean-difference on the right-hand side of (2.11) can be estimated by using the Jensen inequality

$$\begin{aligned}
\Bigl(\sqrt{\mathbb{E}_{\mu_i}(f^2)} - \sqrt{\mathbb{E}_{\mu_j}(f^2)}\Bigr)^2 &\le \mathbb{E}_{\mu_i}(f^2) - 2 \underbrace{\sqrt{\mathbb{E}_{\mu_i}(f^2) \mathbb{E}_{\mu_j}(f^2)}}_{\ge \mathbb{E}_{\mu_i}(f) \mathbb{E}_{\mu_j}(f)} + \mathbb{E}_{\mu_j}(f^2) \\
&\le \mathbb{E}_{\mu_i}(f^2) - 2 \mathbb{E}_{\mu_i}(f) \mathbb{E}_{\mu_j}(f) + \mathbb{E}_{\mu_j}(f^2) \\
&= \operatorname{var}_{\mu_i}(f) + \operatorname{var}_{\mu_j}(f) + \bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\bigr)^2.
\end{aligned} \tag{2.12}$$

Now, we can combine (2.11) and (2.12) to arrive at the desired result (2.10). □

A combination of Lemma 2.4 and Lemma 2.7 yields the desired estimate of the entropy in terms of local variances, local entropies and mean-differences.

COROLLARY 2.8. *Let μ have a mixture representation according to Definition 2.3, then the entropy of f w.r.t. μ can be estimated by*

$$\begin{aligned}\operatorname{Ent}_{\mu}(f^2) \leq \sum_{i=1}^{M} Z_i \operatorname{Ent}_{\mu_i}(f^2) + \sum_{i=1}^{M}\sum_{j\neq i} \frac{Z_i Z_j}{\Lambda(Z_i, Z_j)} \operatorname{var}_{\mu_i}(f) \\ + \sum_{i=1}^{M}\sum_{j>i} \frac{Z_i Z_j}{\Lambda(Z_i, Z_j)} (\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f))^2.\end{aligned} \tag{2.13}$$

2.3. *Main results*. The main results of this work are good estimates of the single terms on the right-hand side of (2.4) and (2.13). In detail, we need the *local PI* and the *local LSI* provided by Theorem 2.9 and Theorem 2.10. Furthermore, we need good control of the mean-differences, which will be the content of Theorem 2.12. Finally, the Eyring–Kramers formulas of Corollary 2.15 and Corollary 2.17 are simple consequences of these representations and estimates.

2.3.1. *Local Poincaré and logarithmic Sobolev inequalities*. Let us now turn to the estimation of the local variances and entropies. From the heuristic understanding of the process ξ_t given by (1.2), we expect a good behavior of the local Poincaré and logarithmic Sobolev constant for the local Gibbs measures μ_i as it resembles the fast convergence of ξ_t to a neighborhood of the next local minimum. Therefore, the local variances and entropies should not contribute to the leading order expansion of the total Poincaré and logarithmic Sobolev constant of μ. This idea is quantified in the next two theorems.

THEOREM 2.9 (Local Poincaré inequality). *Under Assumption 1.4, there exists an admissible partition $\mathcal{P}_{\mathcal{M}} = \{\Omega_i\}_{i=1}^{M}$ for μ (cf. Definition 2.1) such that the associated local Gibbs measures $\{\mu_i\}_{i=1}^{M}$, obtained by restricting μ to Ω_i [cf. (2.2)], satisfy* PI(ϱ_i) *with*

$$\varrho_i^{-1} = O(\varepsilon).$$

THEOREM 2.10 (Local logarithmic Sobolev inequality). *Under Assumption 1.5 and for the same admissible partition $\mathcal{P}_{\mathcal{M}} = \{\Omega_i\}_{i=1}^{M}$ for μ as in Theorem 2.9, the associated local Gibbs measures $\{\mu_i\}_{i=1}^{M}$, obtained by restricting μ to Ω_i [cf. (2.2)], satisfy* LSI(α_i) *with*

$$\alpha_i^{-1} = O(1).$$

Even if Theorem 2.9 and Theorem 2.10 are very plausible, their proof is not easy. The reason is that our situation goes beyond the scope of the standard tools for PI and LSI:

- The Bakry–Émery criterion (cf. Theorem 3.1) cannot be applied because we do not have a convex Hamiltonian.
- A naive application of the Holley–Stroock perturbation principle (cf. Theorem 3.2) would yield an exponentially bad dependence on the parameter ε.
- One cannot apply a simple *Lyapunov argument*, because one cannot impose a drift condition on the boundary of all elements of the partition $\mathcal{P}_{\mathcal{M}}$, simultaneously.

For the proof we apply a subtle combination of a Lyapunov and a perturbation argument. The core of the argument is an explicit construction of a *Lyapunov function*. This Lyapunov function has to satisfy Neumann boundary conditions on the sets Ω_i. By using the canonical partition Ω_i into the basins of attraction of the gradient flow w.r.t. H (see Remark 2.2), the construction of the Lyapunov function would be technically very demanding. We avoid these difficulties by choosing another partition Ω_i such that the Lyapunov function will automatically satisfy Neumann boundary conditions on Ω_i. We outline the argument for Theorem 2.9 and Theorem 2.10 in Section 3.

REMARK 2.11 (Optimality of Theorem 2.9 and Theorem 2.10). The one-dimensional case indicates that the results of Theorem 2.9 and Theorem 2.10 are the best behavior in ε, which one can expect in general. The optimality in the one-dimensional case was investigated in [43], Section 3.3, by using the *Muckenhoupt functional* [36] and *Bobkov–Götze functional* [8].

2.3.2. *Mean-difference estimate*. Let us now turn to the estimation of the mean-difference $(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f))^2$. From the heuristics and the splitting of the variance (2.4) and entropy (2.13), we expect to see in the estimation of the mean-difference the exponential long waiting times of the jumps of the diffusion ξ_t given by (1.2) between the elements of the partition $\mathcal{P}_{\mathcal{M}}$. We have to find a good upper bound for the constant C in the inequality

$$\left(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\right)^2 \leq C \int |\nabla f|^2 \, \mathrm{d}\mu.$$

For this purpose, we introduce in Section 4.1 a *weighted transport distance* between probability measures which yields a variational bound on the constant C. By an approximation argument (cf. Section 4.2), we give an explicit construction of a transport interpolation (cf. Section 4.3), which allows for asymptotically sharp estimates of the constant C.

THEOREM 2.12 (Mean-difference estimate). *Let H satisfy Assumption 1.7 and let $\mathcal{P}_{\mathcal{M}} = \{\Omega_i\}_{i=1}^M$ be an admissible partition for μ (cf. Definition 2.1). Moreover, assume that each local Gibbs measure μ_i of the mixture representation of μ (cf. Definition 2.3) satisfy* $\mathrm{PI}(\varrho_i)$ *with $\varrho_i^{-1} = O(\varepsilon)$. Then the mean-differences between the local Gibbs measures μ_i and μ_j for $i = 1, \ldots, M-1$*

and $j = i+1, \ldots, M$ *satisfy*

$$\begin{aligned}(2.14)\quad &\left(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\right)^2 \\ &\quad \lesssim \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2 H(s_{i,j})|}}{|\lambda^-(s_{i,j})|} \exp\left(\frac{H(s_{i,j})}{\varepsilon}\right) \int |\nabla f|^2 \,\mathrm{d}\mu,\end{aligned}$$

where $\lambda^-(s_{i,j})$ *denotes the negative eigenvalue of the Hessian* $\nabla^2 H(s_{i,j})$ *at the communicating saddle* $s_{i,j}$ *defined in Assumption* 1.7. *The symbol* $\lesssim$ *means* $\le$ *up to a multiplicative error term of the form*

$$1 + O\left(\sqrt{\varepsilon}|\log \varepsilon|^{3/2}\right).$$

The proof of Theorem 2.12 is carried out in full detail in Section 4.

REMARK 2.13 (Multiple minimal saddles). In Assumption 1.7, we demand that there is exactly one minimal saddle between the local minima m_i and m_j. The technique developed in Section 4 is flexible enough to handle also cases, in which there exists more than one minimal saddle between local minima. The according adaptions and the resulting theorem can be found in [43], Section 4.5.

REMARK 2.14 (Relation to capacity). The quantity on the right-hand side of (2.14) is the inverse of the capacity of a small neighborhood around m_i w.r.t. to a small neighborhood around m_j. The capacity is the crucial ingredient of the works [9] and [10].

2.3.3. *Eyring–Kramers formulas*. Now, let us turn to the Eyring–Kramers formula. Starting from the splitting obtained in Lemma 2.4 and Corollary 2.8 a combination of Theorem 2.9, Theorem 2.10 and Theorem 2.12 immediately leads to the multidimensional Eyring–Kramers formula for the PI (cf. [10], Theorem 1.2) and LSI.

COROLLARY 2.15 (Eyring–Kramers formula for Poincaré inequality). *Under Assumptions* 1.4 *and* 1.7, *the measure* μ *satisfies* PI(ϱ) *with*

$$(2.15)\qquad \frac{1}{\varrho} \lesssim Z_1 Z_2 \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2(H(s_{1,2}))|}}{|\lambda^-(s_{1,2})|} \exp\left(\frac{H(s_{1,2})}{\varepsilon}\right),$$

where $\lambda^-(s_{1,2})$ *denotes the negative eigenvalue of the Hessian* $\nabla^2 H(s_{1,2})$ *at the communicating saddle* $s_{1,2}$. *Further, the order is given such that* $H(m_1) \le H(m_i)$ *and* $H(s_{1,2}) - H(m_2)$ *is the energy barrier of the system in the sense of Assumption* 1.7. *The prefactors* Z_i *are given by the relation*

$$(2.16)\qquad Z_i Z_\mu \approx \frac{(2\pi\varepsilon)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \exp\left(-\frac{H(m_i)}{\varepsilon}\right).$$

PROOF. Using the admissible partition $\mathcal{P}_{\mathcal{M}}$ from Theorem 2.9 we decompose the variance into local variances and mean-differences given by Lemma 2.4. An application of Theorem 2.9 and Theorem 2.12 yields the estimate

$$\begin{aligned}
\operatorname{var}_\mu(f) &\le \sum_i Z_i \operatorname{var}_{\mu_i}(f) + \sum_i \sum_{j<i} Z_i Z_j \bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\bigr)^2 \\
&\lesssim \Biggl(O(\varepsilon) + \sum_i \sum_{j>i} \frac{Z_i Z_j Z_\mu}{(2\pi\varepsilon)^{n/2}} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2 H(s_{i,j})|}}{|\lambda^-(s_{i,j})|} \exp\biggl(\frac{H(s_{i,j})}{\varepsilon}\biggr)\Biggr) \qquad (2.17)\\
&\quad \times \int |\nabla f|^2 \,\mathrm{d}\mu.
\end{aligned}$$

The final step is to observe that by Assumption 1.7 the exponential dominating term in (2.17) is given for $i = 1$ and $j = 2$. The precise form of the prefactors Z_i is obtained from (2.1) in Definition 2.1. □

In [10], Theorem 1.2, it is also shown that the upper bound of (2.15) is optimal by an approximation of the harmonic function. Therefore, in the following we can assume that (2.15) holds with $\approx$ instead of $\lesssim$.

REMARK 2.16 (Higher exponentially small eigenvalues). The main result of [10], Theorem 1.2, does not only characterize the second eigenvalue of L but also the higher exponentially small eigenvalues. In principle, these characterizations can be also obtained in the present approach: The dominating exponential modes in (2.17), that is, those obtained by setting $i = 1$, correspond to the inverse eigenvalues of L for $j = 2, \dots, M$. By using the variational characterization of the eigenvalues of the operator L, the other exponentially small eigenvalues may be obtained by restricting the class of test functions f to the orthogonal complement of the eigenspaces of smaller eigenvalues.

COROLLARY 2.17 (Eyring–Kramers formula for logarithmic Sobolev inequalities). *Under Assumptions 1.5 and 1.7, the measure μ satisfies* LSI(α) *with*

$$\begin{aligned}
\frac{2}{\alpha} &\lesssim \frac{Z_1 Z_2}{\Lambda(Z_1, Z_2)} \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2(H(s_{1,2}))|}}{|\lambda^-(s_{1,2})|} \exp\biggl(\frac{H(s_{1,2})}{\varepsilon}\biggr) \qquad (2.18)\\
&\approx \frac{1}{\Lambda(Z_1, Z_2)} \frac{1}{\varrho},
\end{aligned}$$

where the occurring constants are like in Corollary 2.15 and $\Lambda(Z_1, Z_2)$ denotes the logarithmic mean (cf. Appendix A)

$$\Lambda(Z_1, Z_2) = \frac{Z_1 - Z_2}{\log Z_1 - \log Z_2}.$$

PROOF. Using the admissible partition $\mathcal{P}_{\mathcal{M}}$ from Theorem 2.9 and Theorem 2.10, we decompose the Entropy according to Corollary 2.8. From there, we estimate the local entropies and variances as well as the mean-differences by using Theorem 2.9, Theorem 2.10 and Theorem 2.12. Overall, this yields the estimate

$$\begin{aligned}
\operatorname{Ent}_\mu(f^2) &\leq O(1)\sum_{i=1}^{M} Z_i \int |\nabla f|^2 \,\mathrm{d}\mu_i + O(\varepsilon)\sum_{i=1}^{M}\sum_{j\neq i} \frac{Z_i Z_j}{\Lambda(Z_i,Z_j)} \int |\nabla f|^2 \,\mathrm{d}\mu_i \\
&\quad + \sum_{i=1}^{M}\sum_{j>i} \frac{Z_i Z_j}{\Lambda(Z_i,Z_j)} \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2 H(s_{i,j})|}}{|\lambda^-(s_{i,j})|} \exp\biggl(\frac{H(s_{i,j})}{\varepsilon}\biggr) \qquad (2.19)\\
&\quad \times \int |\nabla f|^2 \,\mathrm{d}\mu.
\end{aligned}$$

The first term on the right-hand side of (2.19) can be rewritten as $O(1)\int |\nabla f|^2 \,\mathrm{d}\mu$. For estimating the second term in (2.19), we argue that its prefactor can be estimated as

$$\sum_{i=1}^{M}\sum_{j\neq i} \frac{Z_i Z_j}{\Lambda(Z_i,Z_j)} \overset{(2.22)}{\lesssim} M\sum_{i=1}^{M} Z_i O(\varepsilon^{-1}) = O(\varepsilon^{-1}). \qquad (2.20)$$

Indeed, using the one-homogeneity of $\Lambda(\cdot,\cdot)$ (cf. Appendix A) yields

$$\frac{Z_i Z_j}{\Lambda(Z_i,Z_j)} = Z_i \frac{\log(Z_i/Z_j)}{Z_i/Z_j - 1} = Z_i P\biggl(\frac{Z_i}{Z_j}\biggr) \qquad \text{where } P(x) := \frac{\log x}{x-1}.$$

The function $P(x)$ is decreasing and has a logarithmic singularity at 0. Therefore, using the characterization of the partitions sums Z_i from (2.16) yields the identity

$$\frac{Z_i}{Z_j} = \frac{Z_i Z_\mu}{Z_j Z_\mu} \overset{(2.16)}{\approx} \frac{\sqrt{\nabla^2 H(m_j)}}{\sqrt{\nabla^2 H(m_i)}} \exp\biggl(-\frac{H(m_i)-H(m_j)}{\varepsilon}\biggr), \qquad (2.21)$$

which becomes exponentially small provided that $H(m_i) > H(m_j)$. Hence, the logarithmic mean can be estimated as

$$\frac{Z_i Z_j}{\Lambda(Z_i,Z_j)} = Z_i P\biggl(\frac{Z_i}{Z_j}\biggr) \lesssim Z_i O(\varepsilon^{-1}) \qquad (2.22)$$

implying the desired estimate (2.20). Therefore, the second term in (2.19) can be estimated by $O(1)\int |\nabla f|^2 \,\mathrm{d}\mu$. The third term dominates the first two terms on an exponential scale. This leads to the estimate

$$\begin{aligned}
\operatorname{Ent}_\mu(f^2) &\lesssim \sum_{i=1}^{M}\sum_{j>i} \frac{Z_i Z_j}{\Lambda(Z_i,Z_j)} \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2 H(s_{i,j})|}}{|\lambda^-(s_{i,j})|} e^{H(s_{i,j})/\varepsilon} \\
&\quad \times \int |\nabla f|^2 \,\mathrm{d}\mu.
\end{aligned}$$

From Assumption 1.7 together with (2.22) follows that the exponentially leading order term is given for $i = 1$ and $j = 2$. □

The Eyring–Kramers formula for the PI and LSI stated in Corollary 2.15 and Corollary 2.17 are still implicit. To obtain an explicit formula, one still has insert the asymptotic expansion for the partition functions Z_1, Z_2, and Z_μ. The expression for Z_μ depends on the number of global minima of the Hamiltonian H. Therefore, one has to consider several cases in order to obtain the explicit Eyring–Kramer formula. In the following corollary, we look at two special cases: In the first case, there is only one unique global minimum. In the second case, there are two global minima. In both cases, the dominating term scales exponentially in the saddle height, but it is surprising that the scaling in ε of the exponential pre factor for the LSI constant changes.

COROLLARY 2.18 (Comparison of ϱ and α in special cases). *Let us state two specific cases of* (2.15) *and* (2.18). *Therefore, let* $\{\kappa_i^2\}_{i=1}^M$ *be given by*

$$\kappa_i^2 := \det \nabla^2 H(m_i). \tag{2.23}$$

On the one hand, if one has one unique global minimum, namely $H(m_1) < H(m_i)$ *for* $i \in \{2, \dots, M\}$, *it holds*

$$\frac{1}{\varrho} \approx \frac{1}{\kappa_2} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2(H(s_{1,2}))|}}{|\lambda^-(s_{1,2})|} \exp\left(\frac{H(s_{1,2}) - H(m_2)}{\varepsilon}\right), \tag{2.24}$$

$$\frac{2}{\alpha} \lesssim \left(\frac{H(m_2) - H(m_1)}{\varepsilon} + \log\left(\frac{\kappa_1}{\kappa_2}\right)\right)\frac{1}{\varrho}. \tag{2.25}$$

On the other hand, if $H(m_1) = H(m_2) < H(m_i)$ *for* $i \in \{3, \dots, M\}$, *it holds*

$$\frac{1}{\varrho} \approx \frac{1}{\kappa_1 + \kappa_2} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2(H(s_{1,2}))|}}{|\lambda^-(s_{1,2})|} \exp\left(\frac{H(s_{1,2}) - H(m_2)}{\varepsilon}\right), \tag{2.26}$$

$$\frac{2}{\alpha} \lesssim \frac{1}{\Lambda(\kappa_1, \kappa_2)} \frac{2\pi\varepsilon\sqrt{|\det \nabla^2(H(s_{1,2}))|}}{|\lambda^-(s_{1,2})|} \exp\left(\frac{H(s_{1,2}) - H(m_2)}{\varepsilon}\right). \tag{2.27}$$

PROOF. By (2.15), we still have to estimate nonexplicit factor $\frac{Z_1 Z_2 Z_\mu}{(2\pi\varepsilon)^{n/2}}$. If $H(m_1) < H(m_2)$, then it holds $Z_1 = 1 + O(e^{-(H(m_2)-H(m_1))/\varepsilon})$. The factor $Z_2 Z_\mu$ is given by (2.16) and we obtain

$$\frac{Z_1 Z_2 Z_\mu}{(2\pi\varepsilon)^{n/2}} \approx \frac{1}{\sqrt{\det \nabla^2 H(m_2)}} \exp\left(-\frac{H(m_2)}{\varepsilon}\right),$$

which leads to (2.24). For the LSI, we additionally have to evaluate the factor $\frac{1}{\Lambda(Z_i,Z_j)}$ which can be done with the help of (2.21)

$$\begin{aligned}\frac{1}{\Lambda(Z_i,Z_j)} &= \log\Bigl(\frac{Z_i}{Z_j}\Bigr)\Bigl(1+O\Bigl(\exp\Bigl(-\frac{H(m_2)-H(m_1)}{\varepsilon}\Bigr)\Bigr)\Bigr)\\ &\overset{(2.21)}{\approx} \log\Bigl(\frac{\sqrt{\nabla^2 H(m_j)}}{\sqrt{\nabla^2 H(m_i)}}\exp\Bigl(-\frac{H(m_i)-H(m_j)}{\varepsilon}\Bigr)\Bigr).\end{aligned}$$

That is already the estimate (2.25).

Let us turn now to the case $H(m_1)=H(m_2)<H(m_3)$. Then it holds $Z_1+Z_2=1+O(e^{-(H(m_2)-H(m_1))/\varepsilon})$. In particular it holds $Z_\mu \approx Z_1Z_\mu + Z_2Z_\mu$. Therewith, we can evaluate the factor $Z_1Z_2\frac{Z_\mu}{(2\pi\varepsilon)^{n/2}}$ by using (2.16)

$$\begin{aligned}Z_1Z_2\frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} &= \frac{(2\pi\varepsilon)^{n/2}}{Z_\mu}\frac{Z_1Z_\mu}{(2\pi\varepsilon)^{n/2}}\frac{Z_2Z_\mu}{(2\pi\varepsilon)^{n/2}}\\ &\approx \frac{(2\pi\varepsilon)^{n/2}}{Z_1Z_\mu+Z_2Z_\mu}\frac{Z_1Z_\mu}{(2\pi\varepsilon)^{n/2}}\frac{Z_2Z_\mu}{(2\pi\varepsilon)^{n/2}}\\ &\overset{(2.16)}{=} \frac{1}{1/\kappa_1+1/\kappa_2}\frac{1}{\kappa_1}\frac{1}{\kappa_2}=\frac{1}{\kappa_1+\kappa_2},\end{aligned}$$

which precisely leads to the expression (2.26). By using the homogeneity of $\Lambda(\cdot,\cdot)$ (cf. Appendix A) and again (2.16), it follows for the LSI

$$\frac{Z_1Z_2}{\Lambda(Z_1,Z_2)}\frac{Z_\mu}{(2\pi\varepsilon)^{n/2}}=\frac{1}{\Lambda((2\pi\varepsilon)^{n/2}/(Z_2Z_\mu),(2\pi\varepsilon)^{n/2}/(Z_1Z_\mu))}=\frac{1}{\Lambda(\kappa_2,\kappa_1)}.$$

Finally, the result (2.27) is a consequence of the symmetry of $\Lambda(\cdot,\cdot)$. □

REMARK 2.19 (Identification of α and ϱ). Remark 1.2 shows that always $\alpha\le\varrho$. We want to compare this to the case $H(m_1)=H(m_2)$. Comparing (2.26) and (2.27), we observe

$$1\le\frac{\varrho}{\alpha}\lesssim\frac{(\kappa_1+\kappa_2)/2}{\Lambda(\kappa_1,\kappa_2)}, \tag{2.28}$$

where the constant κ_1 and κ_2 are given by (2.23). The right-hand side of (2.28) consists of an quotient of the arithmetic and the logarithmic mean. The lower bound of 1 can also attained by an application of the logarithmic-arithmetic mean inequality from Lemma A.1. Moreover, equality only holds for $\kappa_1=\kappa_2$. Hence, only in the symmetric case $\varrho\approx\alpha$.

REMARK 2.20 (Relation to mixtures). If $H(m_1)<H(m_2)$, then (2.25) gives

$$\frac{\varrho}{\alpha}\lesssim\frac{1}{2}\log\Bigl(\frac{\kappa_2}{\kappa_1}e^{(H(m_2)-H(m_1))/\varepsilon}\Bigr)\approx\frac{1}{2}|\log Z_2| \qquad \text{where } Z_2=\mu(\Omega_2) \tag{2.29}$$

which shows an inverse scaling in ε. A different scaling behavior between the Poincaré and logarithmic Sobolev constant was also observed by Chafaï and Malrieu [14] in a different context. They consider mixtures of probability measures ν_0 and ν_1 satisfying PI(ϱ_i) and LSI(α_i), that is, for $p \in [0, 1]$ the measure ν_p given by

$$\nu_p = p\nu_0 + (1-p)\nu_1.$$

They deduce conditions under which also ν_p satisfies PI(ϱ_p) and LSI(α_p) and give bounds on the constants. They give one-dimensional examples where the Poincaré constant stays bounded, whereas the logarithmic Sobolev constant blows up logarithmically in the mixture parameter p going to 0 or 1. The common feature of the examples they deal with is $\nu_1 \ll \nu_2$ or $\nu_2 \ll \nu_1$. This case can be generalized to the multidimensional case, where also a different scaling of the Poincaré and logarithmic Sobolev constants is observed. The details can be found in [43], Chapter 6.

In the present case, the Gibbs measure μ has also a mixture representation (2.3). In the two-component case, it has the form

$$\mu = Z_1\mu_1 + Z_2\mu_2.$$

Let us emphasize, that $\mu_1 \perp \mu_2$. The estimate (2.29) also shows a logarithmic blow-up in the mixture parameter Z_2 for the ratio of the Poincaré and the logarithmic Sobolev constant.

2.4. *Optimality of the logarithmic Sobolev constant in one dimension.* In this section, we give a strong indication that the result of Corollary 2.17 is optimal. We explicitly construct a function attaining equality in (2.18) for the one-dimensional case. For this purpose, let μ be a probability measure on $\mathbb{R}$ having as Hamiltonian H a generic double-well (cp. Figure 2). Namely, H has two minima m_1 and m_2 with $H(m_1) \le H(m_2)$ and a saddle s in-between. Then Theorem 2.17 shows

$$\inf_{g:\int g^2\,\mathrm{d}\mu=1} \frac{\int (g')^2\,\mathrm{d}\mu}{\int g^2\log g^2\,\mathrm{d}\mu} \gtrsim \frac{\Lambda(Z_1,Z_2)}{Z_1Z_2}\frac{\sqrt{2\pi\varepsilon}}{Z_\mu}\frac{\sqrt{|H''(s)|}}{2\pi\varepsilon}e^{-H(s)/\varepsilon}. \tag{2.30}$$

We construct a function g attaining the lower bound given by (2.30). We make the following ansatz for the function g: We define g on a small δ-neighborhood around the minima m_1, m_2 and the saddle s:

$$g(x) := \begin{cases} g(m_1), & x \in B_\delta(m_1), \\ g(m_1) + \dfrac{g(m_2)-g(m_1)}{\sqrt{2\pi\varepsilon\sigma}} \displaystyle\int_{m_1}^x e^{-(y-s)^2/(2\sigma\varepsilon)}\,\mathrm{d}y, & x \in B_\delta(s), \\ g(m_2), & x \in B_\delta(m_2). \end{cases}$$

The ansatz depends on the parameters $g(m_1)$, $g(m_2)$ and σ. In between the δ-neighborhoods, the function g is smoothly extended in a monotone fashion.

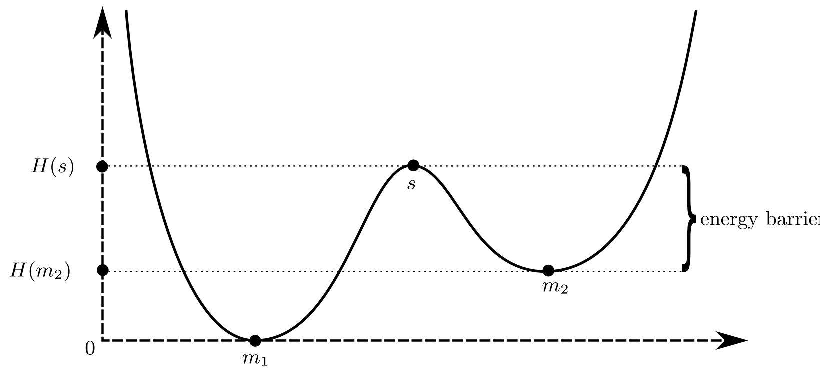

FIG. 2. *Double-well potential H on $\mathbb{R}$ (labeled).*

The measure μ is the usual Gibbs measure as in (1.3). We fix Z_μ by assuming that $H(m_1) = 0$. We represent μ as the mixture

$$\mu = Z_1\mu_1 + Z_2\mu_2 \qquad \text{where } \mu_1 := \mu \llcorner \Omega_1 \text{ and } \mu_2 := \mu \llcorner \Omega_2,$$

hereby, $\Omega_1 := (-\infty, s)$ and $\Omega_2 := (s, \infty)$ and $Z_i := \mu(\Omega_i)$ for $i = 1, 2$, which implies $Z_1 + Z_2 = 1$. Using via an asymptotic evaluation of $\int g^2 \, \mathrm{d}\mu$ one gets

$$\int g^2 \, \mathrm{d}\mu \approx Z_1 g^2(m_1) + Z_2 g^2(m_2) \stackrel{!}{=} 1.$$

This motivates the choice

$$g^2(m_1) = \frac{\tau}{Z_1} \quad \text{and} \quad g^2(m_2) = \frac{1-\tau}{Z_2} = \frac{1-\tau}{1-Z_1} \qquad \text{for some } \tau \in [0, 1].$$

Let us now calculate the denominator of (2.30)

$$\int g^2 \log g^2 \, \mathrm{d}\mu = \tau \log \frac{\tau}{Z_1} + (1-\tau) \log \frac{1-\tau}{Z_2}. \tag{2.31}$$

The final step is to evaluate the Dirichlet energy $\int (g')^2 \, \mathrm{d}\mu$. Therefore, we do a Taylor expansion of H around s. Furthermore, since s is a saddle, it holds $H''(s) < 0$

$$\begin{aligned}
\int (g')^2 \, \mathrm{d}\mu &\approx \frac{(g(m_2) - g(m_1))^2}{Z_\mu 2\pi\varepsilon\sigma} \int_{B_\delta(s)} e^{-(x-s)^2/(\sigma\varepsilon) - H(x)/\varepsilon} \, \mathrm{d}x \\
&\approx \frac{(g(m_2) - g(m_1))^2}{Z_\mu 2\pi\varepsilon\sigma} \int_{B_\delta(s)} e^{-((x-s)^2/\sigma + H(s) + H''(s)(x-s)^2/2)/\varepsilon} \, \mathrm{d}x \\
&\approx \frac{(g(m_2) - g(m_1))^2}{Z_\mu 2\pi\varepsilon\sigma} e^{-H(s)/\varepsilon} \int_{B_\delta(s)} e^{-((x-s)^2/(2\varepsilon))(2/\sigma + H''(s))} \, \mathrm{d}x \\
&\approx \left(\sqrt{\frac{\tau}{Z_1}} - \sqrt{\frac{1-\tau}{Z_2}}\right)^2 \frac{\sqrt{2\pi\varepsilon}}{Z_\mu} e^{-H(s)/\varepsilon} \frac{1}{2\pi\varepsilon} \frac{1}{\sigma\sqrt{2/\sigma + H''(s)}},
\end{aligned} \tag{2.32}$$

where we assume that σ is small enough such that $\frac{2}{\sigma} + H''(s) > 0$. The last step is to minimize the right-hand side of (2.32) in σ, which means to maximize the expression $2\sigma + \sigma^2 H''(s)$ in σ. Elementary calculus results in $\sigma = -\frac{1}{H''(s)} = \frac{1}{|H''(s)|} > 0$ and, therefore,

$$\int (g')^2 \,\mathrm{d}\mu \approx \left(\sqrt{\frac{\tau}{Z_1}} - \sqrt{\frac{1-\tau}{Z_2}}\right)^2 \frac{\sqrt{2\pi\varepsilon}}{Z_\mu} \frac{\sqrt{|H''(s)|}}{2\pi\varepsilon} e^{-H(s)/\varepsilon}. \tag{2.33}$$

Hence, we have constructed by combining (2.31) and (2.33) an upper bound for the optimization problem (2.30) given by

$$\min_{\tau\in(0,1)} \left(\frac{(\sqrt{\tau/Z_1} - \sqrt{(1-\tau)/Z_2})^2}{\tau \log(\tau/Z_1) + (1-\tau)\log((1-\tau)/Z_2)}\right) \frac{\sqrt{2\pi\varepsilon}}{Z_\mu} \frac{\sqrt{|H''(s)|}}{2\pi\varepsilon} e^{-H(s)/\varepsilon}.$$

Note that the parameter $\tau \in (0,1)$ is still free. The minimum in τ is attained at $\tau = Z_2$ according to Lemma A.3 yielding the desired statement

$$\min_{\tau\in(0,1)} \frac{(\sqrt{Z_2/Z_1} - \sqrt{Z_1/Z_2})^2}{Z_2 \log(Z_2/Z_1) + Z_1 \log(Z_1/Z_2)} = \frac{\Lambda(Z_1, Z_2)}{Z_1 Z_2}.$$

3. Local Poincaré and logarithmic Sobolev inequalities. In this section, we proof the local PI of Theorem 2.9 and the local LSI of Theorem 2.10. Even if the choice of a specific admissible partition Ω_i of the space $\mathbb{R}^n$ will be crucial, let us for the moment assume that the partition Ω_i is given by the basins of attraction of the deterministic gradient flow (cf. Remark 2.2).

There are standard criteria to deduce the PI or the LSI. Unfortunately, these criteria do not apply to our situation. Let us consider the *Bakry–Émery criterion* and the *Holley–Stroock perturbation principle*. The *Bakry–Émery criterion* connects convexity of the Hamiltonian to the validity of the PI and the LSI.

THEOREM 3.1 Bakry–Émery criterion [4], Proposition 3, Corollaire 2. *Let $H : D \to \mathbb{R}$ be a Hamiltonian with Gibbs measure*

$$\mu(\mathrm{d}x) = Z_\mu^{-1} \exp(-\varepsilon^{-1} H(x))\,\mathrm{d}x$$

on a convex domain D and assume that $\nabla^2 H(x) \geq \lambda > 0$ for all $x \in \mathbb{R}^n$. Then μ satisfies PI(ϱ) and LSI(α) with

$$\varrho \geq \frac{\lambda}{\varepsilon} \quad \textit{and} \quad \alpha \geq \frac{\lambda}{\varepsilon}.$$

One cannot apply the criterion of Bakry–Émery [4] to our situation, because H is not convex on the elements Ω of the admissible partition (cf. Definition 2.1). Moreover, the elements $\Omega \in \mathcal{P}_{\mathcal{M}}$ are not convex in general.

In nonconvex cases, the standard tool to deduce the PI and the LSI is the *Holley–Stroock perturbation principle*.

THEOREM 3.2 Holley–Stroock perturbation principle [27], p. 1184. *Let H be a Hamiltonian with Gibbs measure $\mu(\mathrm{d}x) = Z_\mu^{-1} \exp(-\varepsilon^{-1} H(x))\,\mathrm{d}x$. Further, let $\tilde{H}$ denote a bounded perturbation of H and let $\tilde{\mu}_\varepsilon$ denote the Gibbs measure associated to the Hamiltonian $\tilde{H}$. If μ satisfies* PI(ϱ) *or* LSI(α) *then also $\tilde{\mu}$ satisfy* PI($\tilde{\varrho}$) *or* LSI($\tilde{\alpha}$) *respectively, where the constants satisfy the bounds*

$$\tilde{\varrho} \geq \exp\bigl(-\varepsilon^{-1} \operatorname{osc} \psi\bigr)\varrho \quad \textit{and} \quad \tilde{\alpha} \geq \exp\bigl(-\varepsilon^{-1} \operatorname{osc}(H - \tilde{H})\bigr)\alpha,$$

where $\operatorname{osc}(H - \tilde{H}) := \sup(H - \tilde{H}) - \inf(H - \tilde{H})$.

The perturbation principle of Holley–Stroock [27] allows to deduce the PI and the LSI constants of nonconvex Hamiltonians from the PI and the LSI of an appropriately convexfied Hamiltonian. However due to its perturbative nature, a naive application Theorem (3.2) would yield an exponential dependence of the PI and the LSI constant on ε.

An important observation for our argument is that the perturbation principle of Holley–Stroock can still be useful, if applied in a careful way: Assume for a moment that the perturbed Hamiltonian $\tilde{H}_\varepsilon$ only differs slightly from the original Hamiltonian H, that is, $\operatorname{osc}(H - \tilde{H}_\varepsilon) = O(\varepsilon)$. Because the perturbation is small w.r.t. ε, the PI and LSI constants of μ and $\tilde{\mu}$ only differ up to an ε-independent factor. This observation is summarized in the following definition and subsequent Lemma 3.4.

DEFINITION 3.3 (ε-modification $\tilde{H}_\varepsilon$ of H). The family of Hamiltonians $\{\tilde{H}_\varepsilon\}_{\varepsilon>0}$ is an ε-modification of H, if there exists an ε-independent constant $C_{\tilde{H}} > 0$ such that for all ε small enough holds

$$|\tilde{H}_\varepsilon(x) - H(x)| \leq C_{\tilde{H}}\varepsilon \qquad \text{for all } x \in \Omega. \tag{$\tilde{\mathrm{H}}_\varepsilon$}$$

To each ε-modification of H we associate the family of ε-modified Gibbs measures $\tilde{\mu}_\varepsilon$ by setting

$$\tilde{\mu}_\varepsilon(\mathrm{d}x) := \frac{1}{Z_{\tilde{\mu}_\varepsilon}} \exp\bigl(-\varepsilon^{-1} \tilde{H}_\varepsilon(x)\bigr)\,\mathrm{d}x \qquad \text{with } Z_{\tilde{\mu}_\varepsilon} := \int \exp\bigl(-\varepsilon^{-1} \tilde{H}_\varepsilon(x)\bigr)\,\mathrm{d}x.$$

LEMMA 3.4 (Perturbation by an ε-modification). *If the ε-modified Gibbs measures $\tilde{\mu}_\varepsilon$ satisfy* PI($\tilde{\varrho}$) *or* LSI($\tilde{\alpha}$), *then the measure μ also satisfies* PI(ϱ) *or* LSI(α), *respectively, where the constants fulfill the estimate*

$$\varrho \geq \exp(-2C_{\tilde{H}})\tilde{\varrho} \quad \textit{and} \quad \alpha \geq \exp(-2C_{\tilde{H}})\tilde{\alpha},$$

where $C_{\tilde{H}}$ is from ($\tilde{\mathrm{H}}_\varepsilon$).

PROOF. The statement directly follows from an application of Theorem 3.2 by considering the estimate ($\tilde{\mathrm{H}}_\varepsilon$). □

Our approach to Theorem 2.9 consists of a nonstandard application of a Lyapunov argument developed by Bakry, Barthe, Cattiaux, Guillin, Wang and Wu (cf. [2, 3, 12] and [13]), which is reminiscent of the spectral gap characterization by Donsker and Varadhan [17]. Compared to these works on the Lyapunov approach, we have to explicitly elaborate the dependence of the PI and LSI constants on ε. Moreover, the theory is only established for Gibbs measure on the whole space. Therefore, the Lyapunov approach of the present work has two main ingredients:

- a Lyapunov function that has to satisfy Neumann boundary conditions on Ω and certain estimates (cf. Definition 3.7 and Theorem 3.15 below), and
- a PI for a truncated Gibbs measure (cf. Definition 3.5 and Lemma 3.6 below).

With the Lyapunov function, we are able to compare the scaling behavior of the PI constant of μ with the behavior of the PI constant of the truncated Gibbs measure $\hat{\mu}_a$ (cf. Theorem 3.8 and Theorem 3.15 below).

DEFINITION 3.5 (Truncated Gibbs measure). For a given number $a > 0$, the truncated Gibbs measures $\{\hat{\mu}_{a,i}\}_{i=1}^M$ are obtained from the Gibbs measure μ by restriction to balls of radius $a\sqrt{\varepsilon}$ around $\{m_i\}_{i=1}^M$, that is,

$$\hat{\mu}_{a,i}(\mathrm{d}x) := \frac{\mathbb{1}_{B_{a\sqrt{\varepsilon}}(m_i)}(x)}{Z_{\hat{\mu}_{a,i}}} \exp\bigl(-\varepsilon^{-1} H(x)\bigr)\,\mathrm{d}x$$

$$\text{with } Z_{\hat{\mu}_{a,i}} := \int_{B_{a\sqrt{\varepsilon}}(m_i)} \exp\bigl(-\varepsilon^{-1} H(x)\bigr)\,\mathrm{d}x.$$

Because the domain and the Hamiltonian of the truncated Gibbs measure $\hat{\mu}_{a,i}$ is convex, one can deduce the scaling behavior of the truncated Gibbs measure $\hat{\mu}_{a,i}$ from the Bakry–Émery criterion. More precisely, it holds the following.

LEMMA 3.6 (PI and LSI for truncated Gibbs measure). *For any $a > 0$ and $i = 1, \ldots, M$ the measures $\hat{\mu}_{a,i}$ satisfy* PI($\hat{\varrho}$) *and* LSI($\hat{\alpha}$) *for ε small enough, where*

$$\frac{1}{\hat{\varrho}} = O(\varepsilon) \quad \textit{and} \quad \frac{1}{\hat{\alpha}} = O(\varepsilon). \tag{3.1}$$

PROOF. In the local minimum m_i the Hessian of H is nondegenerated by Assumptions 1.4 or 1.5. Therefore, for ε small enough, H is strictly convex in $B_{a\sqrt{\varepsilon}}(m_i)$ and satisfies by the Bakry–Émery criterion (cf. Theorem 3.1) PI($\hat{\varrho}$) and LSI($\hat{\alpha}$) with $\hat{\varrho}$ and $\hat{\alpha}$ obeying the relation (3.1). $\square$

The standard ansatz $\exp(\frac{H}{2\varepsilon})$ for a Lyapunov function has the nice feature that it automatically satisfies Neumann boundary conditions on the basins of attraction w.r.t. H, which would be also a canonical choice of the partition $\mathcal{P}_{\mathcal{M}}$ (cf. Remark 2.2). Unfortunately, one cannot guarantee that the necessary estimates for $\exp(\frac{H}{2\varepsilon})$ hold because there is no control on the sign of $\Delta H(x)$ close to saddles [see (3.9) below]. We circumvent this technical problem in the following way: By the observation from above it suffices to consider an ε-modification $\tilde{H}_\varepsilon$ of H. We explicitly construct an ε-modification $\tilde{H}_\varepsilon$ on the whole space $\mathbb{R}^n$ with the property that the standard ansatz $\exp(\frac{\tilde{H}_\varepsilon}{2\varepsilon})$ satisfies the necessary estimates for being a Lyapunov function. However in general, the function $\exp(\frac{\tilde{H}_\varepsilon}{2\varepsilon})$ does not satisfy Neumann boundary conditions on the basins of attraction w.r.t. H. This problem is solved by the following two observations.

- The first one is that $\exp(\frac{\tilde{H}_\varepsilon}{2\varepsilon})$ satisfies Neumann boundary conditions on the basin of attraction w.r.t. the deterministic gradient flow defined by $\tilde{H}_\varepsilon$, that is,

$$\Omega_i := \Big\{ y \in \mathbb{R}^n : \lim_{t\to\infty} y_t = m_i,\ \dot{y}_t = -\nabla \tilde{H}_\varepsilon(y_t),\ y_0 = y \Big\}. \tag{3.2}$$

- The second observation is that this partition $\{\Omega_i\}_{i=1}^M$ of $\mathbb{R}^n$ is admissible in the sense of Definition 2.1 (see Lemma 3.10 below). This fact is intuitively clear from the fact that $\tilde{H}_\varepsilon$ is only a small perturbation of H.

Hence, we choose the partition $\mathcal{P}_\mathcal{M} := \{\Omega_i\}_{i=1}^M$ of $\mathbb{R}^n$ according to (3.2) and apply the Lyapunov approach to the local Gibbs measures $\tilde{\mu}_{\varepsilon,i}$ given by

$$\begin{aligned} \tilde{\mu}_{\varepsilon,i}(\mathrm{d}x) &:= \frac{\mathbb{1}_{\Omega_i}(x)}{Z_{\tilde{\mu}_{\varepsilon,i}}} \exp\big(-\varepsilon^{-1}\tilde{H}_\varepsilon(x)\big)\,\mathrm{d}x \\ &\qquad \text{with } Z_{\tilde{\mu}_{\varepsilon,i}} := \int_{\Omega_i} \exp\big(-\varepsilon^{-1}\tilde{H}_\varepsilon(x)\big)\,\mathrm{d}x. \end{aligned} \tag{3.3}$$

We get that the local Gibbs measures $\tilde{\mu}_{\varepsilon,i}$ satisfy a local PI and LSI with the desired scaling behavior in ε. This scaling behavior of the PI and LSI constant is then transferred to the original Gibbs measure μ restricted to the sets Ω_i by using the perturbation Lemma 3.4.

The remaining part of this section is organized in the following way.

- In Section 3.1, we present the abstract framework how the Lyapunov approach is used for deriving the local PI. We additionally motivate the perturbative nature of the construction of the Lyapunov function. Under the assumption of the existence of a Lyapunov function, we also state the proof Theorem 2.9.
- In Section 3.2, we provide the central ingredient for the Lyapunov approach, namely the existence of a Lyapunov function. We also show that the partition obtained by (3.2) is admissible.
- In Section 3.3, we present the abstract framework how the Lyapunov approach is used for deriving the local LSI. We show that one can use the same Lyapunov function for the local PI as for the local LSI. We also state the proof of Theorem 2.10 deducing the local LSI.

3.1. *Lyapunov approach for the Poincaré inequality.* We start with explaining the Lyapunov approach for deducing a PI. The central notion for the Lyapunov approach is the following definition.

DEFINITION 3.7 (Lyapunov function for Poincaré inequality). Let $H:\Omega\to\mathbb{R}$ be a Hamiltonian with Gibbs measure $\mu(\mathrm{d}x)=\mathbb{1}_\Omega(x)Z_\mu^{-1}\exp(-\varepsilon^{-1}H(x))\,\mathrm{d}x$. Then $W:\Omega\to[1,\infty)$ is a *Lyapunov function* for H provided that:

(i) There exist a domain $U\subset\Omega$ and constants $b>0$ and $\lambda>0$ such that

$$\varepsilon^{-1}LW\le-\lambda W+b\mathbb{1}_U\qquad\text{a.e. in }\Omega.\tag{3.4}$$

(ii) W satisfies Neumann boundary conditions on Ω such that the integration by parts formula holds

$$\forall f\in H^1(\mu|_\Omega):\int_\Omega f(-LW)\,\mathrm{d}\mu=\varepsilon\int_\Omega\langle\nabla f,\nabla W\rangle\,\mathrm{d}\mu.\tag{3.5}$$

Compared to the Lyapunov function of [2] the condition (ii) in Definition 3.7 is new. The reason is that we work on the domain Ω and not on the whole space $\mathbb{R}^n$. The next statement shows that a Lyapunov function and a PI for the truncated measure can be combined to get a PI for the whole measure.

THEOREM 3.8 (Lyapunov condition for PI on domains Ω). *Suppose that H has a Lyapunov functions in the sense of Definition 3.7 and that the restricted measure μ_U given by*

$$\mu_U(\mathrm{d}x):=\mu(\mathrm{d}x)\llcorner U=\frac{\mathbb{1}_U(x)}{\mu(U)}\mu(\mathrm{d}x),$$

satisfies PI(ϱ_U). *Then the associated Gibbs measure μ also satisfies* PI(ϱ) *with constant*

$$\varrho\ge\frac{\lambda}{b+\varrho_U}\varrho_U.$$

The content of the last theorem is standard (cf. [2]), except that we work on the domain Ω and not on the whole space $\mathbb{R}^n$. For the convenience of the reader, we state the short proof.

PROOF OF THEOREM 3.8. Let us rewrite the Lyapunov condition (3.4) and observe

$$1\le-\frac{LW}{\varepsilon\lambda W}+\frac{b}{\lambda}\frac{\mathbb{1}_U}{W}\le-\frac{LW}{\varepsilon\lambda W}+\frac{b}{\lambda}\mathbb{1}_U,\tag{3.6}$$

since $W \geq 1$ by Definition 3.7. By the integration by parts rule (3.5), we obtain following estimate which is due to Definition 3.7(ii). Therewith, we deduce the estimate

$$
\begin{aligned}
\int f^2 \frac{(-LW)}{\varepsilon W} \,\mathrm{d}\mu &= \int \Big\langle \nabla\Big(\frac{f^2}{W}\Big), \nabla W \Big\rangle \,\mathrm{d}\mu \\
&= 2\int \frac{f}{W} \langle \nabla f, \nabla W\rangle \,\mathrm{d}\mu - \int \frac{f^2 |\nabla W|^2}{W^2} \,\mathrm{d}\mu \\
&= \int |\nabla f|^2 \,\mathrm{d}\mu - \int \Big| \nabla f - \frac{f}{W} \nabla W \Big|^2 \,\mathrm{d}\mu \\
&\leq \int |\nabla f|^2 \,\mathrm{d}\mu.
\end{aligned} \tag{3.7}
$$

Let us now turn this estimate into one for the variance $\operatorname{var}_\mu(f)$. Due to fundamental properties of the variance, it holds $\operatorname{var}_\mu(f) \leq \int (f-m)^2 \,\mathrm{d}\mu$, for any $m \in \mathbb{R}$. Hence, applying the estimates (3.6) and (3.7) yields

$$
\begin{aligned}
\operatorname{var}_\mu(f) \ &\leq\ \int (f-m)^2 \,\mathrm{d}\mu \overset{(3.6)}{\leq} \int (f-m)^2 \frac{(-LW)}{\varepsilon\lambda W} + \frac{b}{\lambda} \int_U (f-m)^2 \,\mathrm{d}\mu \\
&\overset{(3.7)}{\leq} \frac{1}{\lambda} \int |\nabla f|^2 \,\mathrm{d}\mu + \frac{b\mu(U)}{\lambda} \int (f-m)^2 \,\mathrm{d}\mu_U.
\end{aligned} \tag{3.8}
$$

We set $m = \int f \,\mathrm{d}\mu_U$, then the last integral in the right-hand side of (3.8) becomes $\operatorname{var}_{\mu_U}(f)$, to which we apply the assumption $\mathrm{PI}(\varrho_U)$. $\square$

Considering the last theorem, it is only left to construct a Lyapunov function in the sense of Definition 3.7 in order to deduce the local PI of Theorem 2.9. An ansatz (cf. [2]) for a Lyapunov function is the function $W = \exp(\frac{1}{2\varepsilon} H)$. Why is this in general a good candidate for an Lyapunov function?

First note that because by our Assumptions 1.4 or 1.5 it holds $H \geq 0$ hence $W \geq 1$ as desired. The second reason is that this choice satisfies Neumann boundary conditions on the boundary of the basin of attraction Ω (see Theorem B.1).

The third reason is that for this choice of W the Lyapunov condition (3.4) is already almost satisfied. One only has to have a special look at critical points. To be more precise, let us consider the condition (3.4) which becomes

$$
\frac{LW}{\varepsilon W} = \frac{1}{2\varepsilon} \Delta H(x) - \frac{1}{4\varepsilon^2} \big|\nabla H(x)\big|^2 \overset{!}{\leq} -\lambda + b \mathbb{1}_U(x). \tag{3.9}
$$

We investigate under which circumstances this condition is satisfied:

- At infinity: The assumption ($\mathrm{A2}_{\mathtt{PI}}$) ensures that (3.9) is satisfied outside of a fixed large ball $B_{\tilde{R}}(0)$ [cf. (3.11) below].

- Away from critical points: The Morse assumption ensures H to be quadratic around critical points, that is, there exists a global constant $c_H > 0$ such that $|\nabla H(x)| \geq c_H \operatorname{dist}(x, \mathcal{S})$ in a neighborhoods of critical points $\mathcal{S}$. This estimate yields (3.9) for x outside of neighborhoods of order $\sqrt{\varepsilon}$ around critical points (see proof of Lemma 3.11 below).

The gradient term cannot help to establish the estimate (3.9), if one is close to critical points. More precisely, it holds:

- If x is in an $\sqrt{\varepsilon}$-neighborhood around the minimum 0, then $\Delta H(x) \approx \sum_i \lambda_i > 0$, where $\{\lambda_i\}_{i=1}^n$ are the eigenvalues of the Hessian at 0. Additionally, the gradient can be estimated as $|\nabla H(x)|^2 \gtrsim \lambda_{\min}^2 |x|^2$, where $\lambda_{\min} = \min_i \lambda_i$. Hence, one cannot compensate the positive Laplacian by the gradient of H. Therefore, one has to choose $U = B_{a\sqrt{\varepsilon}}(0)$ to guarantee the Lyapunov condition (3.9) around the minimum at 0.
- If x is close a local maximum, the Laplacian $\Delta H(x)$ is negative. Hence, the Lyapunov condition is (3.9) is satisfied in this region.
- Assume that x is in an $\sqrt{\varepsilon}$-neighborhood around a saddle, that is, a critical point $s \in \mathcal{S}$ of order $1 \leq k < n$. Again, the gradient term cannot help to establish the estimate (3.9). Hence, the condition (3.9) becomes

$$\Delta H(x) \approx \lambda_1^- + \cdots + \lambda_k^- + \lambda_{k+1}^+ + \cdots + \lambda_n^+ \stackrel{!}{\leq} -\lambda,$$

  where λ_i^- are the negative eigenvalue of the Hessian at s and λ_j^+ are the positive eigenvalues of the Hessian at s. However, for a general Hamiltonian H it may hold that

$$\lambda_1^- + \cdots + \lambda_k^- + \lambda_{k+1}^+ + \cdots + \lambda_n^+ \geq 0$$

  implying that $W = \exp(\frac{1}{2\varepsilon} H)$ is not always a Lyapunov function.

Nevertheless, these observations show that $W = \exp(\frac{1}{2\varepsilon} H)$ is a pretty good guess for a Lyapunov function: One only has to change W close to saddles of H. This leads to the following strategy (cf. Lemma 3.12 from below):

- We construct a perturbation $\tilde{H}_\varepsilon$ of the Hamiltonian H, which coincides with H except of $\sqrt{\varepsilon}$-neighborhoods around saddles.
- In a $\sqrt{\varepsilon}$-neighborhood around a saddle, the perturbation is constructed in such a way that on the one hand the Laplacian of $\tilde{H}_\varepsilon$ is strictly negative. This implies that the function $W = \exp(\frac{1}{2\varepsilon} \tilde{H}_\varepsilon)$ satisfies the estimate (3.9), which is necessary for being a Lyapunov function.
- To assure that $W = \exp(\frac{1}{2\varepsilon} \tilde{H}_\varepsilon)$ satisfies Neumann boundary condition, we choose Ω as a basin of attraction w.r.t. the gradient flow of $\tilde{H}_\varepsilon$ [cf. (3.2)].

After these considerations, let us summarize how the Lyapunov approach is used.

PROPOSITION 3.9. *Assume that an Hamiltonian $\tilde{H}_\varepsilon$ satisfies the Assumption 1.4 uniformly in ε. Let $\mathcal{M} = \{m_1, \dots, m_M\}$ denote the local minima of $\tilde{H}_\varepsilon$. Assume that there are constants $a > 0$ and $\lambda_0 > 0$ such that for all $\varepsilon > 0$ small enough holds*

$$\frac{1}{2\varepsilon}\Delta \tilde{H}_\varepsilon(x) - \frac{1}{4\varepsilon^2}\left|\nabla \tilde{H}_\varepsilon(x)\right|^2 \le -\frac{\lambda_0}{\varepsilon} \qquad \text{for all } x \notin \bigcup_{m\in\mathcal{M}} B_{a\sqrt{\varepsilon}}(m). \tag{3.10}$$

Consider the partition $\mathcal{P}_\mathcal{M} = \{\Omega_i\}_{i=1}^M$ into the basins of attraction of the gradient flow of $\tilde{H}_\varepsilon$ [*cf.* (3.2)]. *Then the associated local Gibbs measures $\{\tilde{\mu}_{\varepsilon,i}\}_{i=1}^M$ given by* (3.3) *satisfy* PI($\tilde{\varrho}_i$) *with constant*

$$\tilde{\varrho}_i^{-1} = O(\varepsilon).$$

PROOF. The function $W = \exp(\frac{1}{2\varepsilon}\tilde{H}_\varepsilon)$ satisfies Neumann boundary conditions on each domain of attraction Ω_i in the sense of (3.5) by Theorem B.1. Indeed, the gradient of W is

$$\nabla W = \frac{1}{2\varepsilon}(\nabla \tilde{H}_\varepsilon)\exp\left(\frac{1}{2\varepsilon}\tilde{H}_\varepsilon\right).$$

Hence, $\nabla W \parallel \nabla \tilde{H}_\varepsilon$ everywhere. Moreover, $\tilde{H}_\varepsilon \in C^3$ is Morse and proper by Assumption 1.4, which shows all the assumptions of Theorem B.1.

Let Ω_i be fixed. Then the estimate (3.10) is just a translation of the estimate (3.4) with constants $\lambda = \frac{\lambda_0}{\varepsilon}$ and $b = \frac{b_0}{\varepsilon}$ for some $b_0 > 0$. Moreover, we choose $U = B_{a\sqrt{\varepsilon}}(m_i)$. Therefore, the function W is a Lyapunov function in the sense of Definition 3.7 on Ω_i. Theorem 3.8 yields that the measure $\tilde{\mu}_{\varepsilon,i}$ satisfies PI($\tilde{\varrho}_i$) with

$$\tilde{\varrho}_i \ge \frac{\lambda_0 \hat{\varrho}}{b_0 + \varepsilon\hat{\varrho}},$$

where $\hat{\varrho}$ denotes the PI constant of the truncated Gibbs measure $\hat{\mu}_{a,i}$ from Definition 3.5. By Lemma 3.6 holds $\hat{\varrho}^{-1} = O(\varepsilon)$, which yields $\tilde{\varrho}_i^{-1} = O(\varepsilon)$. □

Following our strategy, the main ingredient of the proof of the local PI is the existence of an ε-modified Hamiltonian $\tilde{H}_\varepsilon$ satisfying assumption (3.10) of Proposition 3.9.

LEMMA 3.10 (Lyapunov function for PI). *There exits an ε-modification $\tilde{H}_\varepsilon$ of H in the sense of Definition 3.3 such that the Lyapunov estimate (3.10) holds for $\tilde{H}_\varepsilon$. The corresponding partition $\mathcal{P}_\mathcal{M} = \{\Omega_i\}_{i=1}^M$ into the basins of attraction of the gradient flow of $\tilde{H}_\varepsilon$* [*cf.* (3.2)] *is admissible in the sense of Definition 2.1.*

The proof of Lemma 3.10 is not complicated but a bit lengthy. It is stated in full detail in Section 3.2. Now, we only have to put together the parts in order to proof the first main result Theorem 2.9.

PROOF OF THEOREM 2.9. By a combination of Lemma 3.9 and Lemma 3.10 we know that the ε-modified Gibbs measures $\tilde{\mu}_{\varepsilon,i}$ restricted to Ω_i satisfy a PI with the desired scaling behavior $\tilde{\varrho}_i^{-1} = O(\varepsilon)$. Lemma 3.4 implies that then the unmodified Gibbs measure μ_i restricted to Ω_i also satisfies a PI with the same scaling behavior $\varrho_i^{-1} = O(\varepsilon)$. $\square$

3.2. *Construction of a Lyapunov function.* This section is devoted to the proof of Lemma 3.10. We have to construct an ε-modified Hamiltonian $\tilde{H}_\varepsilon$ that satisfies the estimate (3.10). Following the motivation of Section 3.1, we set $\tilde{H}_\varepsilon = H$ away from critical points. Therefore, we have to show that H satisfies the estimate (3.10) away from critical points, which is the content of the next statement.

LEMMA 3.11. *Assume that the Hamiltonian H satisfies the Assumption 1.4. Recall that $\mathcal{S}$ denotes the set of all critical points of H in Ω; that is,*

$$\mathcal{S} = \{y \in \Omega \mid \nabla H(y) = 0\}.$$

Then for $a > 0$ large enough exists $\lambda_0 > 0$ and $\varepsilon_0 > 0$ such that for all $\varepsilon < \varepsilon_0$

$$\frac{\Delta H(x)}{2\varepsilon} - \frac{|\nabla H(x)|^2}{4\varepsilon^2} \leq -\frac{\lambda_0}{\varepsilon} \qquad \text{for all } x \in \mathbb{R}^n \setminus \bigcup_{y \in \mathcal{S}} B_{a\sqrt{\varepsilon}}(y). \tag{3.11}$$

PROOF. The proof basically consists only of elementary calculations based on the nondegeneracy assumption on H. We consider two cases: One in which we verify (3.11) for $|x| \geq \tilde{R}$ with $\tilde{R} < \infty$ large enough. In the second case, we verify (3.11) for $|x| \leq \tilde{R}$.

Let us turn to the first case. We use the assumptions ($A1_{\mathrm{PI}}$) and ($A2_{\mathrm{PI}}$) and we define $\tilde{R}$ such that

$$\forall |x| \geq \tilde{R} : |\nabla H| \geq \frac{C_H}{2} \quad \text{and} \quad |\nabla H|^2 - \Delta H(x) \geq -2K_H. \tag{3.12}$$

Therewith, it is easy to show that for $|x| \geq \tilde{R}$

$$\begin{aligned} \frac{\Delta H(x)}{2\varepsilon} - \frac{|\nabla H(x)|^2}{4\varepsilon^2} &\overset{(3.12)}{\leq} \frac{1}{\varepsilon}\left(K_H - \frac{|\nabla H(x)|}{2}\left(\frac{\nabla H(x)}{2\varepsilon} - 1\right)\right) \\ &\leq \frac{1}{\varepsilon}\left(K_H - \frac{C_H^2}{8}\left(\frac{C_H^2}{8\varepsilon} - 1\right)\right) \leq -\frac{\lambda_0}{\varepsilon}, \end{aligned} \tag{3.13}$$

if $\varepsilon \leq \frac{C_H^2}{8}(1 + 8/C_H^2(K_H + \lambda_0))^{-1} =: \varepsilon_0$. The latter shows the desired statement in this case, with $\lambda_0 > 0$ arbitrary for $\varepsilon \leq \varepsilon_0(\lambda_0)$.

Let us consider the second case. Because $|x| \le \tilde{R}$ it holds $|\Delta H(x)| \le C_{\tilde{R}}$. Therefore, the desired estimate (3.11) follows, if we show that there is a constant $0 < c_H$ such that

$$|\nabla H(x)| \ge c_H a\sqrt{\varepsilon} \qquad \forall x \in B_{\tilde{R}}(0) \Big\backslash \bigcup_{y\in\mathcal{S}} B_{a\sqrt{\varepsilon}}(y) \text{ and } \forall a \in [0, \varepsilon^{-1/2}]. \tag{3.14}$$

Because, then it follows

$$\frac{\Delta H(x)}{2\varepsilon} - \frac{|\nabla H(x)|^2}{4\varepsilon^2} \overset{(3.14)}{\le} \frac{1}{\varepsilon}\left(\frac{C_{\tilde{R}}}{2} - \frac{c_H a}{4}\right) =: -\frac{\lambda_0}{\varepsilon},$$

with $\lambda_0 > 0$ by choosing $a > 2C_{\tilde{R}}/c_H =: a_0$. Hence, we can choose first $a > a_0$, which gives rise to some $\lambda_0(a) > 0$, by the last estimate under the assumption $a < \varepsilon_0^{-1/2} \le \varepsilon^{-1/2}$. Hence, we have to choose $\varepsilon_0 < \min\{\varepsilon_0(\lambda_0(a)), a^{-2}\}$ with $\varepsilon_0(\lambda_0(a))$ defined after (3.13).

Finally, the estimate (3.14) is a consequence of the fact that H is a Morse function (cp. Definition 1.3 and Assumption 1.4) and, therefore, nondegenerate quadratic around critical points. That means, there exists a global constant $c_H > 0$ such that $|\nabla H(x)| \ge c_H \min\{\operatorname{dist}(x, \mathcal{S}), 1\}$, which implies (3.14). $\square$

Now, we consider the ε-modification $\tilde{H}_\varepsilon$ near critical points. The verification of the following statement represents the core of the construction of the Lyapunov function.

LEMMA 3.12. *Let $\mathcal{M} = \{m_1, \dots, m_M\}$ denote the set containing the minima of H. Then there are constants $C > 0$, $a > 0$ and $\lambda_0 > 0$ such that for $\varepsilon < C$ there exists an ε-modification $\tilde{H}_\varepsilon$ of H in the sense of Definition 3.3 satisfying*

$$\tilde{H}_\varepsilon(x) = H(x) \qquad \textit{for all } x \notin \bigcup_{y\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(y)$$

and

$$\frac{\Delta \tilde{H}_\varepsilon(x)}{2\varepsilon} - \frac{|\nabla \tilde{H}_\varepsilon(x)|^2}{4\varepsilon^2} \le -\frac{\lambda_0}{\varepsilon} \qquad \textit{for all } x \in \bigcup_{y\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(y). \tag{3.15}$$

As a direct consequence of Lemma 3.11, the estimate (3.15) is satisfied for all

$$x \notin \bigcup_{m\in\mathcal{M}} B_{a\sqrt{\varepsilon}}(m).$$

PROOF. It is sufficient to construct the ε-modification $\tilde{H}_\varepsilon$ only locally on a small neighborhood of any critical point $y \in \mathcal{S} \setminus \mathcal{M}$. By translation, we may assume w.l.o.g. that $y = 0$.

Because the Hamiltonian H is a Morse function in the sense of Definition 1.3, we may assume that u_i, $i \in \{1, \dots, n\}$ are orthonormal eigenvectors w.r.t. the Hessian $\nabla^2 H(0)$. The corresponding eigenvalues are denoted by λ_i, $i \in \{1, \dots, n\}$

labeled such that $\lambda_1, \dots, \lambda_\ell < 0$ and $\lambda_{\ell+1}, \dots, \lambda_n > 0$ for some $\ell \in \{1, \dots, n\}$. If $\ell = n$, hence $\lambda_i < 0$ for $i = 1, \dots, n$, we are nearby a local maximum and set $\tilde{H}_\varepsilon(x) = H(x)$ on $B_{a\sqrt{\varepsilon}}(0)$ and the desired estimate (3.15) follows directly for $x \in B_{a\sqrt{\varepsilon}}(0)$.

Otherwise, that is, $\ell < n$, let us choose a constant $\delta > 0$ small enough such that

$$-\tilde{\delta} := (n - 2\ell)\delta + \sum_{i=1}^{\ell} \lambda_i < 0 \quad \text{and} \quad \delta \le \frac{1}{2} \min\{\lambda_i : i = \ell + 1, \dots, n\}. \tag{3.16}$$

Because $u_1, \dots, u_n$ is an orthonormal basis of $\mathbb{R}^n$, we introduce a norm $|\cdot|_\delta$ on $\mathbb{R}^n$ by

$$|x|_\delta^2 := \sum_{i=1}^{\ell} \frac{1}{2}\delta \big|\langle u_i, x\rangle\big|^2 + \sum_{i=\ell+1}^{n} \frac{1}{2}(\lambda_i - \delta)\big|\langle u_i, x\rangle\big|^2. \tag{3.17}$$

The norm $|\cdot|_\delta$ is equivalent to the standard Euclidean norm $|\cdot|$ and satisfies the estimate

$$\frac{\delta}{2}|x|^2 \le |x|_\delta^2 \le \frac{\lambda_{\max}^+ - \delta}{2}|x|^2 \le \frac{\lambda_{\max}^+}{2}|x|^2, \tag{3.18}$$

where $\lambda_{\max}^+ = \max\{\lambda_i : i = \ell+1, \dots, n\}$. The last ingredient for the construction of $\tilde{H}_\varepsilon$ is a smooth cut-off function $\xi : [0, \infty) \to \mathbb{R}$ satisfying for $a > 0$ to be specified later

$$\begin{aligned} &\xi'(r) = -1 \qquad \text{for } r \le \tfrac{1}{4}a^2\varepsilon, \qquad -1 \le \xi'(r) \le 0 \qquad \text{for } r \ge \tfrac{1}{4}a^2\varepsilon, \\ &\xi(r) = 0 \qquad \text{for } r \ge a^2\varepsilon \end{aligned} \tag{3.19}$$

and in addition for some $C_\xi > 0$,

$$0 \le \xi(r) \le C_\xi a^2\varepsilon \quad \text{and} \quad \big|\xi''(r)\big| \le \frac{C_\xi}{a^2\varepsilon}. \tag{3.20}$$

With the help of the norm $|\cdot|_\delta$ and the function ξ we define the function $\tilde{H}_\varepsilon$ by

$$\tilde{H}_\varepsilon(x) = H(x) + H_b(x) \qquad \text{where } H_b(x) := \xi\big(|x|_\delta^2\big). \tag{3.21}$$

Note that by definition of H_b holds $\tilde{H}_\varepsilon(x) = H(x)$ for all $|x| \ge a\sqrt{\varepsilon}$. Because $\xi(r) = O(\varepsilon)$, it follows that $\tilde{H}_\varepsilon$ is an ε-modification of H in the sense of Definition 3.3.

Let us now turn to the verification of the estimate (3.15). It is sufficient to deduce the following two facts: The first one is the estimate

$$\Delta \tilde{H}_\varepsilon(x) \le -\frac{\tilde{\delta}}{2} \qquad \text{for all } |x|_\delta \le \frac{a}{2}\sqrt{\varepsilon}. \tag{3.22}$$

The second one is that there is a constant $\lambda_0 > 0$ such that for a large enough and ε small enough it holds

$$\frac{\Delta \tilde{H}_\varepsilon(x)}{2} - \frac{|\nabla \tilde{H}_\varepsilon(x)|^2}{4\varepsilon} \le -\lambda_0 \qquad \text{for all } \frac{a}{2}\sqrt{\varepsilon} \le |x|_\delta \le a\sqrt{\varepsilon}. \tag{3.23}$$

Let us first derive the estimate (3.22). Using that $\xi'(x) = -1$ for $|x|_\delta \le \frac{a}{2}\sqrt{\varepsilon}$, one obtains that $\Delta H_b(x) = -\Delta|x|_\delta^2$ for $|x|_\delta \le \frac{a}{2}\sqrt{\varepsilon}$. Hence, by Taylor expansion we get for $|x|_\delta \le \frac{a}{2}\sqrt{\varepsilon}$ that

$$\begin{aligned} \Delta \tilde{H}(x) = \Delta H(0) - \Delta|x|_\delta^2 + O(\sqrt{\varepsilon}) &\le \sum_{i=1}^n \lambda_i - \sum_{i=\ell+1}^n \lambda_i + (n-2\ell)\delta + O(\sqrt{\varepsilon}) \\ &= \sum_{i=1}^{\ell} \lambda_i + (n-2\ell)\delta + O(\sqrt{\varepsilon}) \overset{(3.16)}{\le} -\tilde{\delta} + O(\sqrt{\varepsilon}) \le -\frac{\tilde{\delta}}{2}, \end{aligned}$$

for ε small enough, which yields the desired statement (3.22).

Let us turn to the verification of (3.23). We need that there exists a constant $0 < C_\Delta < \infty$ independent of ε and a such that

$$\Delta \tilde{H}(x) \le C_\Delta \qquad \text{for all } \frac{a}{2}\sqrt{\varepsilon} < |x|_\delta < a\sqrt{\varepsilon}. \tag{3.24}$$

Indeed, observe that

$$\begin{aligned} \Delta \tilde{H}_\varepsilon(x) &= \Delta H(x) + \xi''\big(|x|_\delta^2\big)\big|\nabla|x|_\delta^2\big|^2 + \underbrace{\xi'\big(|x|_\delta^2\big)}_{\le 0}\underbrace{\Delta|x|_\delta^2}_{\ge 0} \\ &\overset{(3.20)}{\le} \Delta H(x) + \frac{C_\xi}{a^2\varepsilon}\left|\sum_{i=1}^{\ell} \delta\langle u_i, x\rangle u_i + \sum_{i=\ell+1}^n (\lambda_i - \delta)\langle u_i, x\rangle u_i\right|^2 \\ &\le \Delta H(x) + \frac{C_\xi}{a^2\varepsilon}\left(\sum_{i=1}^{\ell} \delta^2\big|\langle u_i, x\rangle\big|^2 + \sum_{i=\ell+1}^n (\lambda_i-\delta)^2\big|\langle u_i, x\rangle\big|^2\right) \\ &\overset{(3.17)}{\le} \Delta H(x) + \frac{C_\xi}{a^2\varepsilon} 2\lambda_{\max}^+ |x|_\delta^2 \le C_H + 2C_\xi\lambda_{\max}^+ =: C_\Delta, \end{aligned}$$

where C_Δ is independent of ε and a, which yields (3.24).

Additionally, we need that there is a constant $0 < c_\nabla < \infty$ such that

$$\big|\nabla \tilde{H}_\varepsilon(x)\big|^2 \ge c_\nabla a^2 \varepsilon \qquad \text{for all } \frac{a}{2}\sqrt{\varepsilon} < |x|_\delta < \tilde{a}\sqrt{\varepsilon}. \tag{3.25}$$

Before deducing (3.25), we want to show that the observations (3.24) and (3.25) already yield the desired statement (3.23): For $a^2 \ge 4\frac{C_\Delta}{c_\nabla}$, one gets

$$\frac{\Delta \tilde{H}_\varepsilon(x)}{2\varepsilon} - \frac{|\nabla \tilde{H}_\varepsilon(x)|^2}{4\varepsilon^2} \le \frac{C_\Delta}{2\varepsilon} - \frac{c_\nabla a^2}{4\varepsilon} \le -\frac{C_\Delta}{2\varepsilon} \qquad \text{for all } \frac{a}{2}\sqrt{\varepsilon} < |x|_\delta < a\sqrt{\varepsilon},$$

which is the desired statement (3.23). Therefore, it is only left to deduce the estimate (3.25). By the definition of $\tilde{H}_\varepsilon$ from above, we can write

$$|\nabla \tilde{H}_\varepsilon(x)|^2 = |\nabla H(x)|^2 + |\nabla H_b(x)|^2 + 2\langle \nabla H(x), \nabla H_b(x)\rangle. \tag{3.26}$$

Let us have a closer look at each term on the right-hand side of the last identity and let us start with the first term. By applying Taylor's formula to $\nabla H(x)$, we obtain

$$|\nabla H(x) - \nabla^2 H(0)x| \le \tilde{C}_\nabla |x| \overset{(3.18)}{\le} C_\nabla |x|_\delta \tag{3.27}$$

for some $\tilde{C}_\nabla, C_\nabla > 0$. Therefore, we can estimate

$$|\nabla H(x)|^2 \ge |\nabla^2 H(0)x|^2 - C_\nabla^2 a^4 \varepsilon^2 \qquad \text{for } |x|_\delta \le a\sqrt{\varepsilon}. \tag{3.28}$$

By the definition of $\lambda_1, \dots, \lambda_n$, we also know

$$|\nabla^2 H(0)x|^2 = \sum_{i=1}^n \lambda_i^2 |\langle u_i, x\rangle|^2. \tag{3.29}$$

Let us have a closer look at the second term in (3.26), namely $|\nabla H_b(x)|^2$. From the definition (3.21) of $|\nabla H_b(x)|^2$ follows

$$\begin{aligned} |\nabla H_b(x)|^2 &= |\xi'(|x|_\delta^2)|^2 \left(\sum_{i=1}^{\ell} \delta^2 |\langle u_i, x\rangle|^2 + \sum_{i=\ell+1}^n (\lambda_i - \delta)^2 |\langle u_i, x\rangle|^2 \right) \\ &\overset{(3.17)}{\le} 2\lambda_{\max}^+ |x|_\delta^2. \end{aligned} \tag{3.30}$$

Now, we turn the analysis of the last term, namely $2\langle \nabla H(x), \nabla H_b(x)\rangle$. By using the estimates (3.27) and (3.30), we get for $|x|_\delta \le a\sqrt{\varepsilon}$.

$$\begin{aligned} \langle \nabla H(x), \nabla H_b(x)\rangle &= \langle \nabla^2 H(0)x, \nabla H_b(x)\rangle + \langle \nabla H(x) - \nabla^2 H(0)x, \nabla H_b(x)\rangle \\ &\underset{(3.30)}{\overset{(3.27)}{\ge}} \langle \nabla^2 H(0)x, \nabla H_b(x)\rangle - 2C_\nabla \lambda_{\max} |x|_\delta^3 \\ &\ge -\sum_{i=1}^{\ell} \lambda_i \delta |\xi'(|x|_\delta^2)| |\langle u_i, x\rangle|^2 \\ &\quad - \sum_{i=\ell+1}^n \lambda_i(\lambda_i - \delta) |\xi'(|x|_\delta^2)| |\langle u_i, x\rangle|^2 - O(\varepsilon^{3/2}). \end{aligned} \tag{3.31}$$

Combining now the estimates and identities (3.26), (3.28), (3.29), (3.30) and (3.31), we arrive for $|x|_\delta \le a\sqrt{\varepsilon}$ at

$$\begin{aligned} |\nabla \tilde{H}_\varepsilon(x)|^2 \ge &\sum_{i=1}^{\ell} (\lambda_i - \delta|\xi'(|x|_\delta^2)|)^2 |\langle u_i, x\rangle|^2 \\ &+ \sum_{i=\ell+1}^n (\lambda_i - (\lambda_i - \delta)|\xi'(|x|_\delta^2)|)^2 |\langle u_i, x\rangle|^2 - O(\varepsilon^{3/2}). \end{aligned}$$

By (3.19) holds $|\xi'(|x|_\delta^2)| \le 1$, which applied to the last inequality yields

$$\left|\nabla \tilde{H}_\varepsilon(x)\right|^2 \ge \delta^2 \sum_{i=1}^n \left|\langle u_i, x\rangle\right|^2 - O\left(\varepsilon^{3/2}\right).$$

Because $u_1, \dots, u_n$ is an orthonormal basis of $\mathbb{R}^n$, the desired statement (3.25) follows for $\frac{a\sqrt{\varepsilon}}{2} \le |x|_\delta \le a\sqrt{\varepsilon}$ from

$$\begin{aligned}\left|\nabla \tilde{H}_\varepsilon(x)\right|^2 &\ge \delta^2 |x|^2 - O\left(\varepsilon^{3/2}\right) \overset{(3.18)}{\ge} \frac{2\delta^2}{\lambda_{\max}^+}|x|_\delta^2 - O\left(\varepsilon^{3/2}\right) \\ &\ge \frac{\delta^2}{2\lambda_{\max}^+} a^2\varepsilon - O\left(\varepsilon^{3/2}\right) \ge c_\nabla a^2 \varepsilon\end{aligned}$$

for some $c_\nabla < \frac{\delta^2}{2\lambda_{\max}^+}$ and ε small enough. □

Considering the statement of Lemma 3.12, there is only one thing to show in order to verify Lemma 3.10.

LEMMA 3.13. *Let $\mathcal{P}_\mathcal{M} = \{\Omega_i\}_{i=1}^M$ be the partition obtained from the $\tilde{H}_\varepsilon$ from Lemma 3.12 by considering the basins of attraction Ω_i from (3.2). Then $\mathcal{P}_\mathcal{M}$ is an admissible partition in the sense of Definition 2.1.*

Before we turn to the proof of Lemma 3.13, we show the following auxiliary statement.

LEMMA 3.14. *If an Hamiltonian $H : \mathbb{R}^n \to \mathbb{R}$ satisfies the Assumption 1.4, then there exist numbers $R > 0$ and $c_H > 0$ such that*

$$H(x) \ge \min_{|z|=R} H(z) + c_H\left(|x| - R\right).$$

Because $H \ge 0$ by Assumption 1.4, a direct consequence is $\int \exp(-H(x))\,\mathrm{d}x < \infty$.

PROOF. By the assumption ($\mathrm{A1}_{\mathtt{PI}}$), we can choose $R > 0$ large enough such that

$$\left|\nabla H(x)\right| \ge \frac{C_H}{2} \qquad \text{for all } |x| \ge R. \tag{3.32}$$

In particular, this implies that for all critical points $s \in \mathcal{S}$ holds $|s| \le R$. Now, let us we consider the following evolution:

$$\dot{x}_t = -\frac{\nabla H(x_t)}{|\nabla H(x_t)|}, \qquad x_0 = x, 0 \le t < \infty$$

with starting point x, $|x| > R$. Because by Lemma B.12

$$\mathbb{R}^n = \biguplus_{s\in\mathcal{S}} \Big\{ y \in \mathbb{R}^n : \lim_{t\to\infty} y_t = s,\ \dot{y}_t = -\nabla H(y_t),\ y_0 = y \Big\}$$

and for all critical points $s \in \mathcal{S}$ of H it holds $|s| \leq R$, the gradient line $\{x_t\}$ has to hit the ball $B_R(0)$ after some time $t > 0$ at some point x_t for the first time. It follows

$$\begin{aligned} H(x_t) - H(x_0) &= \int_0^t \frac{\mathrm{d}}{\mathrm{d}s} H(x_s)\,\mathrm{d}s \\ &= -\int_0^t \nabla H(x_s) \cdot \frac{\nabla H(x_s)}{|\nabla H(x_s)|}\,\mathrm{d}s = -\int_0^t \big|\nabla H(x_s)\big|\,\mathrm{d}s. \end{aligned}$$

Using the lower bound (3.32) on $|\nabla H(x_t)|$, we get that

$$H(x) = H(x_t) + \int_0^t \big|\nabla H(x_s)\big|\,\mathrm{d}s \geq \inf_{|z|=R} H(z) + t\frac{c_H}{2}.$$

Because the evolution x_t moves at speed 1, we know that t is the length of the gradient-flow line connecting the points x and x_t. However, this length cannot be shorter than $t \geq |x| - R$, which yields the desired statement. $\square$

PROOF OF LEMMA 3.13. We start with showing that $\tilde{H}_\varepsilon$ has the same local minima $\mathcal{M} = \{m_1, \dots, m_M\}$ as the original Hamiltonian H. Because

$$\tilde{H}_\varepsilon(x) = H(x) \qquad \text{for all } x \notin \bigcup_{y\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(y),$$

it suffices to show that $\tilde{H}_\varepsilon$ has no local minima in the set

$$\bigcup_{y\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(y).$$

However, this statement follows directly from the estimate (3.15), that is,

$$\frac{\Delta\tilde{H}_\varepsilon(x)}{2\varepsilon} - \frac{|\nabla\tilde{H}_\varepsilon(x)|^2}{4\varepsilon^2} \leq -\frac{\lambda_0}{\varepsilon} \qquad \text{for all } x \in \bigcup_{y\in\mathcal{S}\setminus\mathcal{M}} B_{a\sqrt{\varepsilon}}(y).$$

Indeed, the last estimate shows that either $|\nabla\tilde{H}_\varepsilon(x)| \neq 0$ or $\Delta\tilde{H}_\varepsilon(x) < 0$.

The fact that $\tilde{H}_\varepsilon$ has the same local minima as H allows us to apply Lemma B.12 showing

$$\mathbb{R}^n = \biguplus_{i=1}^M \Omega_i = \biguplus_{i=1}^M \Big\{ y \in \mathbb{R}^n : \lim_{t\to\infty} y_t = m_i,\ \dot{y}_t = -\nabla\tilde{H}_\varepsilon(y_t),\ y_0 = m_i \Big\},$$

which is already (ii) of Definition 2.1.

The last step in the proof is to show that $\mu(\Omega_i)Z_\mu$ satisfies the asymptotic expansion given by (2.1). Let us consider one local minimum $m_i \in \mathcal{M}$. W.l.o.g. we assume $\tilde{H}_\varepsilon(m_i) = H(m_i) = 0$. We introduce $\Sigma_i := (\nabla^2 H(m_i))^{-1}$ and define for $r_0 > 0$ specified later the ellipsoid

$$E_i := \{x \in \mathbb{R}^n : |\Sigma_i^{-1/2}(x - m_i)| \le \sqrt{2r_0\varepsilon|\log\varepsilon|}\},$$

where the square root of Σ_i^{-1} is uniquely defined in the set of positive symmetric matrices. Note that for small enough ε it holds $E_i \subset \Omega_i$ and $\tilde{H}_\varepsilon(x) = H(x)$ for $x \in E_i$. The covariance matrix Σ_i is nondegenerate because of H being a Morse function. Therefore, there is a constant $c_H < 1$ such that

$$B_{\sqrt{c_H 2r_0\varepsilon|\log\varepsilon|}}(m_i) \subset E_i \subset B_{\sqrt{c_H^{-1} 2r_0\varepsilon|\log\varepsilon|}}(m_i). \tag{3.33}$$

We split the integral into

$$\mu(\Omega_i)Z_\mu = \int_{E_i} \exp\left(-\frac{\tilde{H}_\varepsilon(x)}{\varepsilon}\right) \mathrm{d}x + \int_{\Omega_i \setminus E_i} \exp\left(-\frac{\tilde{H}_\varepsilon(x)}{\varepsilon}\right) \mathrm{d}x =: I_1 + I_2.$$

The results follows from an asymptotic expansion for I_1 and an error estimate for I_2.

We start with the error estimate for I_2. Let the constant $R > 0$ be chosen as in Lemma 3.14. We split the term I_2 up into

$$\begin{aligned} I_2 &= \int_{(\Omega_i \setminus E_i)\cap B_R(0)} \exp\left(-\frac{\tilde{H}_\varepsilon(x)}{\varepsilon}\right) \mathrm{d}x + \int_{\Omega_i \setminus B_R(0)} \exp\left(-\frac{\tilde{H}_\varepsilon(x)}{\varepsilon}\right) \mathrm{d}x \\ &=: I_3 + I_4. \end{aligned}$$

Let us estimate the term I_3. On a small neighborhood around m_i it holds $H = \tilde{H}_\varepsilon$ and H is uniformly convex. Therefore, there is a constant $\delta > 0$ and $\kappa > 0$ such that for all x with $|x - m_i| \le \delta$

$$\left|(\nabla^2 H(x))^{1/2}\xi\right|^2 = \langle \xi, \nabla^2 H(x)\xi\rangle \ge \kappa|\xi|^2 \qquad \text{for all } \xi \in \mathbb{R}^n.$$

Hence, for $x \in \Omega_i \setminus E_i$ we have the lower bound by additionally considering (3.33)

$$\tilde{H}_\varepsilon(x) \ge \inf_{z\in\partial E_i} \tilde{H}_\varepsilon(z) \ge \frac{\kappa}{2} \inf_{z\in\partial E_i} |z - m_i|^2 \ge \kappa c_H r_0 \varepsilon|\log\varepsilon|.$$

Now, we can estimate I_3 as

$$I_3 \le \exp(-\kappa c_H r_0|\log\varepsilon|)|B_R(0)|.$$

Let us turn to the estimation of I_4. An application of Lemma 3.14 yields

$$\begin{aligned} I_4 &\le \exp\left(-\varepsilon^{-1} \min_{|z|=R} H(z)\right) \int_{\Omega_i \setminus B_R(0)} \exp\left(-c_H \frac{|x| - R}{\varepsilon}\right) \mathrm{d}x \\ &\le C_H \exp(-\kappa c_h r_0|\log\varepsilon|). \end{aligned}$$

So overall, we have estimated the term I_2 as

$$I_2 \le C_H \exp\big(-\kappa c_h r_0 |\log \varepsilon|\big) = C_H \varepsilon^{\kappa c_H r_0} = O\big(\varepsilon^{\alpha}\big) \qquad \text{for } r_0 > \frac{\alpha}{\kappa c_H} \text{ and } \alpha > 0. \tag{3.34}$$

Hence, I_2 becomes smaller than every power of ε for r_0 large enough.

Now, we turn to the asymptotic approximation of the term I_1. The Taylor expansion of H on E_i yields for $x \in E_i$

$$H(x) = \tfrac{1}{2}\langle x, \nabla^2 H(m_i)x\rangle + O\big((\varepsilon|\log \varepsilon|)^{3/2}\big).$$

In particular, this implies

$$\exp\left(-\frac{H(x)}{\varepsilon}\right) = \exp\left(-\frac{1}{2\varepsilon}\langle x, \nabla^2 H(m_i)x\rangle\right)\exp\big(O\big(\sqrt{\varepsilon}|\log \varepsilon|^{3/2}\big)\big).$$

For ε small enough, it holds $\exp(O(\sqrt{\varepsilon}|\log \varepsilon|^{3/2})) = 1 + O(\sqrt{\varepsilon}|\log \varepsilon|^{3/2})$. Therewith, we get the following expression for I_1:

$$\begin{aligned}
I_1 &= \int_{E_i} \exp\left(-\frac{1}{2\varepsilon}\langle x, \nabla^2 H(m_i)x\rangle\right) \mathrm{d}x \big(1 + O\big(\sqrt{\varepsilon}|\log \varepsilon|^{3/2}\big)\big) \\
&= \frac{(2\pi\varepsilon)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}} \\
&\quad \times \left(1 - \frac{\sqrt{\det \nabla^2 H(m_i)}}{(2\pi\varepsilon)^{n/2}} \int_{\mathbb{R}^n \setminus E_i} \exp\left(-\frac{1}{2\varepsilon}\langle x, \nabla^2 H(m_i)x\rangle\right) \mathrm{d}x\right) \\
&\quad \times \big(1 + O\big(\sqrt{\varepsilon}|\log \varepsilon|^{3/2}\big)\big).
\end{aligned}$$

Now, we apply the following tail estimate for a Gaussian, which we will proofed for the convenience of the reader below:

$$\frac{\sqrt{\det \nabla^2 H(m_i)}}{(2\pi\varepsilon)^{n/2}} \int_{\mathbb{R}^n \setminus E_i} \exp\left(-\frac{1}{2\varepsilon}\langle x, \nabla^2 H(m_i)x\rangle\right) \mathrm{d}x = O(\sqrt{\varepsilon}). \tag{3.35}$$

The latter yields the asymptotic expansion

$$I_1 = \frac{(2\pi\varepsilon)^{n/2}}{\sqrt{\det \nabla^2 H(m_i)}}\big(1 + O\big(\sqrt{\varepsilon}|\log \varepsilon|^{3/2}\big)\big). \tag{3.36}$$

Now, the desired asymptotic expansion (2.1) for $\mu(\Omega_i)Z_\mu$ follows form a combination of the expansion for the term I_1 in (3.36) and I_2 in (3.34) with α chosen sufficiently large, that is, $\alpha > (n+1)/2$.

We close the argument by deducing the desired tail estimate (3.35). By the change of variables $x \mapsto y = (2\varepsilon\Sigma_i)^{-1/2}(x - m_i)$ and by denoting $\omega(\varepsilon) =$

$\sqrt{r_0 |\log \varepsilon|}$, we deduce

$$\begin{aligned}
&\frac{\sqrt{\det \nabla^2 H(m_i)}}{(2\pi\varepsilon)^{n/2}} \int_{\mathbb{R}^n \setminus E_i} \exp\left(-\frac{1}{2\varepsilon}\langle x, \nabla^2 H(m_i) x\rangle\right) \mathrm{d}x \\
&\quad = \frac{1}{\pi^{n/2}} \int_{\mathbb{R}^n \setminus B_{\omega(\varepsilon)}(0)} e^{-y^2} \,\mathrm{d}y \\
&\quad = \frac{n}{\Gamma(n/2+1)} \int_{\omega(\varepsilon)}^{\infty} r^{n-1} e^{-r^2} \,\mathrm{d}r = \frac{\Gamma(n/2, \omega^2(\varepsilon))}{\Gamma(n/2)},
\end{aligned}$$

where $\Gamma(\frac{n}{2}, \omega^2(\varepsilon))$ is the complementary incomplete Gamma function. It has the asymptotic expansion (cf. [38], pp. 109–112)

$$\Gamma\left(\frac{n}{2}, \omega^2(\varepsilon)\right) = O\big(e^{-\omega^2(\varepsilon)} \omega^{n-2}(\varepsilon)\big) = O\big(\varepsilon^{r_0} |r_0 \log \varepsilon|^{n/2-1}\big) = O(\sqrt{\varepsilon})$$

for r_0 large enough, which yields the desired result. □

3.3. *Lyapunov approach for the logarithmic Sobolev inequality.* The goal of this section is to prove Theorem 2.10 deducing the local LSI. We follow the same strategy as for the proof of Theorem 2.9, which we outlined in Section 3.1. Therefore, we consider the partition $\mathcal{P}_{\mathcal{M}} = \{\Omega_i\}_{i=1}^{M}$ into the basins of attraction of the gradient flow of $\tilde{H}_\varepsilon$ [cf. (3.2)].

The Lyapunov condition for proving LSI is stronger than the one for PI. Nevertheless, the construction of the ε-modified Hamiltonian $\tilde{H}_\varepsilon$ from the previous section carries over and we can use the same Lyapunov function as for the PI, but have to provide additional estimates. The Lyapunov condition for LSI goes back to the work of Cattiaux et al. [12]. We adapt [13], Theorem 1.2, to the case for domains Ω. In addition, we will work out the explicit dependence between the constants of the Lyapunov condition, the logarithmic Sobolev constant and especially their ε-dependence.

THEOREM 3.15 (Lyapunov condition for LSI). *Suppose that*:

(i) *There exists a C^2-function $W : \Omega \to [1, \infty)$ and constants $\lambda, b > 0$ such that for $L = \varepsilon\Delta - \nabla H \cdot \nabla$ holds*

$$\forall x \in \Omega : \frac{1}{\varepsilon} \frac{LW}{W} \le -\lambda |x|^2 + b. \tag{3.37}$$

(ii) *$\nabla^2 H \ge -K_H$ for some $K_H > 0$ and μ satisfies PI(ϱ).*

(iii) *W satisfies Neumann boundary conditions on Ω [cf. (3.5)].*

Then μ satisfies LSI(α) with

$$\frac{1}{\alpha} \le 2\sqrt{\frac{1}{\lambda}\left(\frac{1}{2} + \frac{b + \lambda\mu(|x|^2)}{\varrho}\right)} + \frac{K_H}{2\varepsilon\lambda} + \frac{K_H(b + \lambda\mu(|x|^2)) + 2\varepsilon\lambda}{\varrho\varepsilon\lambda}, \tag{3.38}$$

where $\mu(|x|^2)$ denotes the second moment of μ.

Before turning to the proof of Theorem (3.38), we need the following auxiliary result.

LEMMA 3.16 ([13], Lemma 3.4). *Assume that* $V : \Omega \to \mathbb{R}$ *is a nonnegative locally Lipschitz function such that*:

(i) *For some lower bounded function* ϕ

$$\frac{Le^V}{e^V} = LV + \varepsilon |\nabla V|^2 \le -\varepsilon \phi \tag{3.39}$$

in the distributional sense.

(ii) V *satisfies Neumann boundary condition on* Ω [*cf.* (3.5)].

Then for any $g \in H^1(\mu)$ *holds*

$$\int \phi g^2 \, \mathrm{d}\mu \le \int |\nabla g|^2 \, \mathrm{d}\mu.$$

PROOF. We can assume w.l.o.g. that g is smooth with bounded support and ϕ is bounded. For the verification of the desired statement, we need the symmetry of L in $L^2(\mu)$ w.r.t. to V:

$$\forall f \in H^1(\mu) : \int f(-LV) \, \mathrm{d}\mu = \varepsilon \int \nabla f \cdot \nabla V \, \mathrm{d}\mu, \tag{3.40}$$

and the Young inequality:

$$2g\nabla V \cdot \nabla g \le |\nabla V|^2 g^2 + |\nabla g|^2. \tag{3.41}$$

An application of the assumption (3.39) yields

$$\begin{aligned} \varepsilon \int \phi g^2 \, \mathrm{d}\mu &\overset{(3.39)}{\le} \int (-LV - \varepsilon |\nabla V|^2) g^2 \, \mathrm{d}\mu \\ &\overset{(3.40)}{=} \varepsilon \int (2g\nabla V \cdot \nabla g - |\nabla V|^2 g^2) \, \mathrm{d}\mu \overset{(3.41)}{\le} \varepsilon \int |\nabla g|^2 \, \mathrm{d}\mu, \end{aligned}$$

which is the desired estimate. □

The proof of Theorem 3.15 relies on an interplay of some other functional inequalities, which will not occur anywhere else.

PROOF OF THEOREM 3.15. The argument of [13] is a combination of the Lyapunov condition (3.37) leading to a defective WI inequality and the use of the HWI inequality of Otto and Villani [40]. In the following, we will use the measure ν given by $\nu(\mathrm{d}x) = h(x)\mu(\mathrm{d}x)$, where we can assume w.l.o.g. that ν is a probability measure, that is, $\int h \, \mathrm{d}\mu = 1$. The first step is to estimate the Wasserstein distance in terms of the total variation [46], Theorem 6.15

$$W_2^2(\nu, \mu) \le 2 \| | \cdot |^2 (\nu - \mu) \|_{\mathrm{TV}}. \tag{3.42}$$

For every function g with $|g| \leq \phi(x) := \lambda|x|^2$, where λ is from the Lyapunov condition (3.37) we get

$$\begin{aligned}\int g \, \mathrm{d}(\nu - \mu) &\leq \int \phi \, \mathrm{d}\nu + \int \phi \, \mathrm{d}\mu \\ &= \int \big(\lambda|x|^2 - b\big)h(x)\mu(\mathrm{d}x) + \int b \, \mathrm{d}\nu + \mu(\phi).\end{aligned} \tag{3.43}$$

We can apply to $\int (\lambda|x|^2 - b)h \, \mathrm{d}\mu$ Lemma 3.16, where the assumptions are exactly the Lyapunov condition (3.37) by choosing $V = \log W$. Moreover, the Neumann condition also translates to V since W is bounded from below by 1. Therewith, we arrive at

$$\int \big(\lambda|x|^2 - b\big)h \, \mathrm{d}\mu \leq \int |\nabla \sqrt{h}|^2 \, \mathrm{d}\mu = \int \frac{|\nabla h|^2}{4h} \, \mathrm{d}\mu = \frac{1}{2} I(\nu|\mu), \tag{3.44}$$

by the definition of the Fisher information. Taking the supremum over g in (3.43) and combining the estimate with (3.42) and (3.44) we arrive at the defective Wasserstein-information inequality

$$\frac{\lambda}{2} W_2^2(\nu, \mu) \leq \lambda \big\| |\cdot|^2(\nu - \mu)\big\|_{\mathrm{TV}} \leq \frac{1}{2} I(\nu|\mu) + b + \mu(\phi). \tag{3.45}$$

The next step is to use the HWI inequality [40], Theorem 3, which holds by the assumption $\nabla^2 H \geq -K_H$

$$\mathrm{Ent}_\mu(h) \leq W_2(\nu, \mu)\sqrt{2I(\nu|\mu)} + \frac{K_H}{2\varepsilon} W_2^2(\nu, \mu).$$

Substituting inequality (3.45) into the HWI inequality and using the Young inequality $ab \leq \frac{\tau}{2}a^2 + \frac{1}{2\tau}b^2$ for $\tau > 0$ results in

$$\begin{aligned}\mathrm{Ent}_\mu(h) &\leq \tau I(\nu|\mu) + \left(\frac{1}{2\tau} + \frac{K_H}{2\varepsilon}\right) W_2^2(\nu, \mu) \\ &\overset{(3.45)}{\leq} \left(\tau + \frac{1}{2\lambda}\left(\frac{1}{\tau} + \frac{K_H}{\varepsilon}\right)\right) I(\nu|\mu) + \frac{1}{\lambda}\left(\frac{1}{\tau} + \frac{K_H}{\varepsilon}\right)\big(b + \mu(\phi)\big).\end{aligned} \tag{3.46}$$

The last inequality is of the type $\mathrm{Ent}_\mu(h) \leq \frac{1}{\alpha_d} I(\nu|\mu) + B \int h \, \mathrm{d}\mu$ and is often called defective logarithmic Sobolev inequality $\mathrm{dLSI}(\alpha_d, B)$. It is well known that a defective logarithmic Sobolev inequality can be tightened by PI(ϱ) to LSI(α) with constant (cf. Proposition [31])

$$\frac{1}{\alpha} = \frac{1}{\alpha_d} + \frac{B+2}{\varrho}. \tag{3.47}$$

A combination of (3.46) and (3.47) reveals

$$\begin{aligned}\frac{1}{\alpha} &= \tau + \frac{1}{2\lambda}\left(\frac{1}{\tau} + \frac{K_H}{\varepsilon}\right) + \frac{1}{\varrho}\left(\frac{1}{\lambda}\left(\frac{1}{\tau} + \frac{K_H}{\varepsilon}\right)\big(b + \mu(\phi)\big) + 2\right) \\ &= \tau + \frac{1}{\tau\lambda}\left(\frac{1}{2} + \frac{b + \mu(\phi)}{\varrho}\right) + \frac{K_H}{2\varepsilon\lambda} + \frac{K_H(b + \mu(\phi)) + 2\varepsilon\lambda}{\varrho\varepsilon\lambda} =: \tau + \frac{c_1}{\tau} + c_2.\end{aligned}$$

The last step is to optimize in τ, which leads to $\tau = \sqrt{c_1}$ and, therefore, $\frac{1}{\alpha} = 2\sqrt{c_1} + c_2$. The final result (3.38) follows by recalling the definition of $\phi(x) = \lambda|x|^2$. $\square$

The crucial ingredient is a Lyapunov function satisfying the condition (3.37). We follow the ideas of Section 3.1 and Section 3.2. We use the same ε-modification $\tilde{H}_\varepsilon$ as constructed in the proof of Lemma 3.12.

LEMMA 3.17 (Lyapunov function for LSI). *We consider the ε-modification $\tilde{H}_\varepsilon$ of H constructed in Section 3.2. Then the Lyapunov function $W(x) = \exp(\frac{1}{2\varepsilon}\tilde{H}_\varepsilon(x))$ satisfies on Ω the Lyapunov condition (3.37) with constants*

$$b = \frac{b_0}{\varepsilon} \quad \textit{and} \quad \lambda \geq \frac{\lambda_0}{\varepsilon}$$

for some $b_0, \lambda_0 > 0$ and Hessian $\nabla^2 \tilde{H}_\varepsilon(x) \geq -K_{\tilde{H}_\varepsilon}$ for some $K_{\tilde{H}_\varepsilon} \geq 0$.

The proof consists of three steps, which correspond to three regions of interests. First, we will consider a neighborhood of ∞, that is, we will fix some $\tilde{R} > 0$ and only consider $|x| \geq \tilde{R}$. Then we will look at an intermediate regime for $a\sqrt{\varepsilon} \leq |x| \leq \tilde{R}$, where we will have to take special care for the neighborhoods around critical points and use the construction of Lemma 3.12. The last regime is for $|x| \leq a\sqrt{\varepsilon}$, which will be the simplest case.

Therefore, besides the construction done in the proof of Lemma 3.12, we need an analogous formulation of Lemma 3.11 under the stronger assumption ($\mathrm{A1}_{\mathtt{LSI}}$).

LEMMA 3.18. *Assume that the Hamiltonian H satisfies assumption ($\mathrm{A1}_{\mathtt{LSI}}$). Then there is a constant $0 \leq C_H < \infty$ and $0 \leq \tilde{R} < \infty$ such that $H(x) = \tilde{H}_\varepsilon(x)$ for $|x| \geq \tilde{R}$ and for ε small enough*

$$\frac{\Delta H(x)}{2\varepsilon} - \frac{|\nabla H(x)|^2}{4\varepsilon^2} \leq -\frac{C_H}{\varepsilon}|x|^2 \qquad \textit{for all } |x| \geq \tilde{R}. \tag{3.48}$$

We skip the proof of the Lemma 3.18, because it would work in the same way as for Lemma 3.11 and only consists of elementary calculations based on the nondegeneracy assumption on H. The only difference, is that we now demand the stronger statement (3.48), which is a consequence of the stronger assumption ($\mathrm{A1}_{\mathtt{LSI}}$) in comparison to assumption ($\mathrm{A2}_{\mathtt{PI}}$).

Now, we have collected the auxiliary statements and can proof Lemma 3.17.

PROOF OF LEMMA 3.17. First, let us check the lower bound on the Hessian of $\tilde{H}_\varepsilon$. Because we use the same $\tilde{H}_\varepsilon$ as constructed in Lemma 3.12, the support of $\tilde{H}_\varepsilon - H$ is compact. Additionally, $\tilde{H}_\varepsilon$ is smooth. This already implies the lower bound on the Hessian $\nabla^2 \tilde{H}_\varepsilon$ for compact domains. Outside a sufficient large domain, we know that $H = \tilde{H}_\varepsilon$. Hence, the lower bound on $\nabla^2 \tilde{H}_\varepsilon$ follows directly from assumption ($\mathrm{A2}_{\mathtt{LSI}}$).

Now, we verify the Lyapunov condition (3.37). Recall that $W = \exp(\frac{1}{2\varepsilon}\tilde{H}_\varepsilon)$. Hence, straightforward calculation reveals

$$\frac{1}{\varepsilon}\frac{LW}{W} = \frac{1}{2\varepsilon}\Delta\tilde{H}_\varepsilon + \frac{1}{4\varepsilon^2}|\nabla\tilde{H}_\varepsilon|^2 - \frac{1}{2\varepsilon^2}|\nabla\tilde{H}_\varepsilon|^2 = \frac{1}{2\varepsilon}\Delta\tilde{H}_\varepsilon - \frac{1}{4\varepsilon^2}|\nabla\tilde{H}_\varepsilon|^2.$$

If $|x| \geq \tilde{R}$ with $\tilde{R}$ given in Lemma 3.18, we apply (3.48) and have the Lyapunov condition fulfilled with constant $\lambda = \frac{C_H}{\varepsilon}$. This allows us to only consider $x \in B_{\tilde{R}} \cap \Omega$, which is a bounded domain. In this case, Lemma 3.12 yields for $a\sqrt{\varepsilon} \leq |x| \leq \tilde{R}$ the estimate

$$\frac{1}{\varepsilon}\frac{LW}{W} \leq -\frac{\lambda_0}{\varepsilon} \leq -\frac{\lambda_0}{\tilde{R}^2\varepsilon}|x|^2. \tag{3.49}$$

Let us consider the final case $|x| \leq a\sqrt{\varepsilon}$. In this case, the Hamiltonian $H = \tilde{H}_\varepsilon$. Additionally, H is smooth and strictly convex on $B_{a\sqrt{\varepsilon}}(0)$. Therefore, one easily obtains the bound

$$\frac{1}{\varepsilon}\frac{LW}{W} \leq \frac{1}{2\varepsilon}\Delta H(x) \leq \frac{b_0}{\varepsilon}. \tag{3.50}$$

A combination of (3.49) and (3.50) yields the desired estimate (3.37). $\square$

Before proceeding with the proof of Theorem 2.10, we remark, that the Lyapunov condition for the PI and in particular for the LSI imply an estimate of the second moment of μ.

LEMMA 3.19 (Second moment estimate). *If H fulfills the Lyapunov condition* (3.4) *with $U = B_R(0)$ for $R > 0$, then μ has finite second moment and it holds*

$$\int |x|^2 \mu(\mathrm{d}x) \leq \frac{1 + bR^2}{\lambda}. \tag{3.51}$$

PROOF. As it is outlined in the proof of Theorem 3.8 (cf. also [2]), the Lyapunov condition (3.4) yields the following estimate: for any function f and $m \in \mathbb{R}$ it holds

$$\int (f - m)^2 \,\mathrm{d}\mu \leq \frac{1}{\lambda}\int |\nabla f|^2 \,\mathrm{d}\mu + \frac{b}{\lambda}\int_{B_R(0)} (f - m)^2 \,\mathrm{d}\mu.$$

We set $f(x) = |x|$ and $m = 0$ to observe the estimate (3.51). $\square$

Now, we have collected all auxiliary results to proof the second main Theorem 2.10.

PROOF OF THEOREM 2.10. For the same reason as in the proof of Theorem 2.9, we omit the index i. The first step is also the same as in the proof of Theorem 2.9. By Lemma 3.4, we obtain that, whenever $\tilde{H}_\varepsilon$ is an ε-modification of μ in the sense of Definition 3.3, the logarithmic Sobolev constants α and $\tilde{\alpha}$ of μ and $\tilde{\mu}_\varepsilon$ satisfy $\alpha \geq \exp(-2C_{\tilde{H}})\tilde{\alpha}$.

The next step is to construct an explicit ε-modification $\tilde{H}$ satisfying the Lyapunov condition (3.37) of Theorem 3.15, which is provided by Lemma 3.17.

Additionally, the logarithmic Sobolev constant $\tilde{\alpha}$ depends on the second moment of $\tilde{\mu}_\varepsilon$. Since $\tilde{H}_\varepsilon$ satisfies by Lemma 3.10 in particular the Lyapunov condition for PI (3.4) with constants $\lambda \geq \frac{\lambda_0}{\varepsilon}$, $b \leq \frac{b_0}{\varepsilon}$ and $R = a\sqrt{\varepsilon}$, we can apply Lemma 3.19 and arrive at

$$\int |x|^2 \, \mathrm{d}\tilde{\mu}_\varepsilon \leq \frac{1 + R^2 b}{\lambda} \leq \frac{1 + b_0 a^2}{\lambda_0} \varepsilon = O(\varepsilon).$$

Now, we have control on all constants occurring in (3.38) and can determine the logarithmic Sobolev constant $\tilde{\alpha}$ of $\tilde{\mu}_\varepsilon$. Let us estimate term by term of (3.38) and use the fact from Theorem (2.9), that $\tilde{\mu}_\varepsilon$ satisfies PI($\tilde{\varrho}$) with $\tilde{\varrho}^{-1} = O(\varepsilon)$

$$2\sqrt{\frac{1}{\lambda}\left(\frac{1}{2} + \frac{b + \lambda \tilde{\mu}_\varepsilon(|x|^2)}{\varrho}\right)} \leq 2\sqrt{\frac{\varepsilon}{\lambda_0}\left(\frac{1}{2} + O(1)\right)} = O(\sqrt{\varepsilon}).$$

The second term evaluates to $\frac{K_H}{2\varepsilon\lambda} = O(1)$ and finally the last one

$$\frac{K_H(b + \lambda \tilde{\mu}_\varepsilon(|x|^2)) + 2\varepsilon\lambda}{\varrho\varepsilon\lambda} = O(\varepsilon)\left(K_H\left(\frac{b_0}{\varepsilon} + O(\varepsilon)\right) + O(1)\right) = O(1).$$

A combination of all the results leads to the conclusion $\tilde{\alpha}^{-1} = O(1)$ and since $\tilde{H}_\varepsilon$ is only an ε-modification of H also $\alpha^{-1} = O(1)$. $\square$

4. Mean-difference estimates—weighted transport distance. This section is devoted to the proof of Theorem 2.12. We want to estimate the mean-difference $(\mathbb{E}_{\mu_i} f - \mathbb{E}_{\mu_j} f)^2$ for i and j fixed. The proof consists of four steps:

In the first step, we introduce the *weighted transport distance* in Section 4.1. This distance depends on the transport speed similarly to the Wasserstein distance, but in addition weights the speed of a transported particle w.r.t. the reference measure μ. The weighted transport distance allows in general for a variational characterization of the constant C in the inequality

$$\left(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\right)^2 \leq C \int |\nabla f|^2 \, \mathrm{d}\mu.$$

The problem of finding good estimates of the constant C is then reduced to the problem of finding a good transport interpolation between the measures μ_i and μ_j w.r.t. to the weighted transport distance.

For measures as general as μ_i and μ_j, the construction of an explicit transport interpolation is not feasible. Therefore, the second step consists of an approximation, which is done in Section 4.2. There, the restricted measures μ_i and μ_j are replaced by *simpler* measures ν_i and ν_j, namely truncated Gaussians. We show in Lemma 4.6 that this approximation only leads to higher order error terms.

The most import step, the third one, consists of the estimation of the mean-difference w.r.t. the approximations ν_i and ν_j. Because the structure of ν_i and ν_j is very simple, we can explicitly construct a transport interpolation between ν_i and ν_j (see Lemma 4.11 in Section 4.3). The last step consists of collecting and controlling the error (cf. Section 4.4).

4.1. *Mean-difference estimates by transport.* At the moment, let us consider two arbitrary measures $\nu_0 \ll \mu$ and $\nu_1 \ll \mu$. The starting point of the estimation is a representation of the mean-difference as a transport interpolation. This idea goes back to [14]. However, they used a similar but nonoptimal estimate for our purpose. Hence, let us consider a transport interpolation $(\Phi_s : \mathbb{R}^n \to \mathbb{R}^n)_{s\in[0,1]}$ between ν_0 and ν_1, that is, the family $(\Phi_s)_{s\in[0,1]}$ satisfies

$$\Phi_0 = \mathrm{Id}, \qquad (\Phi_1)_\sharp \nu_0 = \nu_1 \quad \text{and} \quad (\Phi_s)_\sharp \nu_0 =: \nu_s.$$

The representation of the mean-difference as a transport interpolation is attained by using the fundamental theorem of calculus, that is,

$$\left(\mathbb{E}_{\nu_0}(f) - \mathbb{E}_{\nu_1}(f)\right)^2 = \left(\int_0^1 \int \langle \nabla f(\Phi_s), \dot{\Phi}_s\rangle \,\mathrm{d}\nu_0 \,\mathrm{d}s\right)^2.$$

At this point, it is tempting to apply the Cauchy–Schwarz inequality in $L^2(\mathrm{d}\nu_0 \times \mathrm{d}s)$ leading to the estimate in [14]. However, this strategy would not yield the preexponential factors in the Eyring–Kramers formula (2.15) (cf. Remark 4.2). On Stephan Luckhaus' advice, the authors realized the fact that it really matters on which integral you apply the Cauchy–Schwarz inequality. This insight lead to the following proceeding:

$$\begin{aligned}
\left(\mathbb{E}_{\nu_0}(f) - \mathbb{E}_{\nu_1}(f)\right)^2 &= \left(\int_0^1 \int \langle \nabla f, \dot{\Phi}_s \circ \Phi_s^{-1}\rangle \,\mathrm{d}\nu_s \,\mathrm{d}s\right)^2 \\
&= \left(\int \left\langle \nabla f, \int_0^1 \dot{\Phi}_s \circ \Phi_s^{-1} \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu} \,\mathrm{d}s \right\rangle \mathrm{d}\mu\right)^2 \qquad (4.1)\\
&\le \int \left| \int_0^1 \dot{\Phi}_s \circ \Phi_s^{-1} \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu} \,\mathrm{d}s \right|^2 \mathrm{d}\mu \int |\nabla f|^2 \,\mathrm{d}\mu.
\end{aligned}$$

Note that in the last step we have applied the Cauchy–Schwarz inequality only in $L^2(\mathrm{d}\mu)$ and that the desired Dirichlet integral $\int |\nabla f|^2 \,\mathrm{d}\mu$ is already recovered.

The prefactor in front of the Dirichlet energy on the right-hand side of (4.1) only depends on the transport interpolation $(\Phi_s)_{s\in[0,1]}$. Hence, we can minimize over all possible admissible transport interpolations and arrive at the following definition.

DEFINITION 4.1 (Weighted transport distance $\mathcal{T}_\mu$). Let μ be an absolutely continuous probability measure on $\mathbb{R}^n$ with connected support. Additionally, let ν_0 and ν_1 be two probability measures such that $\nu_0 \ll \mu$ and $\nu_1 \ll \mu$, then define the *weighted transport distance* by

$$\mathcal{T}_\mu^2(\nu_0, \nu_1) := \inf_{\Phi_s} \int \left| \int_0^1 \dot{\Phi}_s \circ \Phi_s^{-1} \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu} \,\mathrm{d}s \right|^2 \mathrm{d}\mu. \tag{4.2}$$

The family $(\Phi_s)_{s\in[0,1]}$ is chosen absolutely continuous in the parameter s such that $\Phi_0 = \mathrm{Id}$ on $\operatorname{supp} \nu_0$ and $(\Phi_1)_\sharp \nu_0 = \nu_1$. For a fixed family and $(\Phi_s)_{s\in[0,1]}$ and a point $x \in \operatorname{supp} \mu$, the *cost density* is defined by

$$\mathcal{A}(x) := \left| \int_0^1 \dot{\Phi}_s \circ \Phi_s^{-1}(x) \nu_s(x) \,\mathrm{d}s \right|. \tag{4.3}$$

REMARK 4.2 (Relation of $\mathcal{T}_\mu$ to [14]). The transport distance $\mathcal{T}_\mu(\nu_0, \nu_1)$ is always smaller than the constant obtained in [14], Section 4.6. Indeed, applying the Cauchy–Schwarz inequality on $L^2(\mathrm{d}s)$ in (4.2) yields

$$\begin{aligned}\mathcal{T}_\mu^2(\nu_0, \nu_1) &\le \inf_{\Phi_s} \int \int_0^1 |\dot{\Phi}_s \circ \Phi_s^{-1}|^2 \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu} \,\mathrm{d}s \int_0^1 \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu} \,\mathrm{d}s \,\mathrm{d}\mu \\ &\le \inf_{\Phi_s} \left(\sup_x \left(\int_0^1 \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu}(x) \,\mathrm{d}s \right) \int \int_0^1 |\dot{\Phi}_s|^2 \,\mathrm{d}s \,\mathrm{d}\nu_0 \right),\end{aligned}$$

where we used the assumption that $\nu_s \ll \mu$ for all $s \in [0,1]$ in the last L^1-L^∞-estimate.

REMARK 4.3 (Relation of $\mathcal{T}_\mu$ to the L^2-Wasserstein distance W_2). If the support of μ is convex, we can set the transport interpolation $(\Phi_s)_{s\in[0,1]}$ to the linear interpolation map $\Phi_s(x) = (1-s)x + sU(x)$. Assuming that U is the optimal W_2-transport map between ν_0 and ν_1, the estimate in Remark 4.2 becomes

$$\mathcal{T}_\mu^2(\nu_0, \nu_1) \le \left(\sup_x \int_0^1 \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu}(x) \,\mathrm{d}s \right) W_2^2(\nu_0, \nu_1).$$

REMARK 4.4 (Invariance under time rescaling). The cost density $\mathcal{A}$ given by (4.3) is independent of rescaling the transport interpolation in the parameter s. Indeed, we observe that

$$\mathcal{A}(x) = \left| \int_0^1 \dot{\Phi}_s \circ \Phi_s^{-1}(x) \nu_s(x) \,\mathrm{d}s \right| = \left| \int_0^T \dot{\Phi}_t^T \circ (\Phi_t^T)^{-1}(x) \nu_t^T(x) \,\mathrm{d}t \right|,$$

where $\Phi_t^T = \Phi_{t/T}$ and $\nu_t^T = \nu_{t/T}$.

REMARK 4.5 (Relation to negative Sobolev-norms). The weighted transport distance is a dynamic formulation for the homogeneous negative Sobolev norm

$\dot{H}^{-1}(\mathrm{d}\mu)$ like Benamou and Brenier did for the Wasserstein distance [5]. Precisely, for $\nu_0 = \varrho_0 \mu$ and $\nu_1 = \varrho_1 \mu$ holds

$$\mathcal{T}_\mu^2(\nu_0, \nu_1) = \|\varrho_0 - \varrho_1\|^2_{\dot{H}^{-1}(\mathrm{d}\mu)} = \inf_{h \in \dot{H}^1(\mu)} \left\{ \int |\nabla h|^2 \,\mathrm{d}\mu : Lh = \varrho_0 - \varrho_1 \right\}.$$

In fact, it is possible to define a whole class of weighted Wasserstein type distances interpolating between the negative Sobolev norm and the Wasserstein distance. Theses transports were introduced in [16].

4.2. *Approximation of the local measures μ_i.* In this subsection, we show that it is sufficient to consider only the mean-difference w.r.t. some auxiliary measures ν_i approximating μ_i for $i = 1, \dots, M$. More precisely, the next lemma shows that there are nice measures ν_i which are close to the measures μ_i in the sense of the mean-difference.

LEMMA 4.6 (Mean-difference of approximation). *For $i = 1, \dots, M$ let ν_i be a truncated Gaussian measure centered around the local minimum m_i with covariance matrix $\Sigma_i = (\nabla^2 H(m_i))^{-1}$, more precisely*

$$\nu_i(\mathrm{d}x) = \frac{1}{Z_{\nu_i}} e^{-\Sigma_i^{-1}[x - m_i]/(2\varepsilon)} \mathbb{1}_{E_i}(x)\,\mathrm{d}x \qquad \text{where } Z_{\nu_i} = \int_{E_i} e^{-\Sigma_i^{-1}[x - m_i]/(2\varepsilon)}\,\mathrm{d}x, \tag{4.4}$$

where we write $A[x] := \langle x, Ax \rangle$. The restriction E_i is given by an ellipsoid

$$E_i = \{ x \in \mathbb{R}^n : |\Sigma_i^{-1/2}(x - m_i)| \le \sqrt{2\varepsilon}\,\omega(\varepsilon) \}. \tag{4.5}$$

Additionally, assume that μ_i satisfies $\mathrm{PI}(\varrho_i)$ *with $\varrho_i^{-1} = O(\varepsilon)$.*

Then the following estimate holds:

$$\big(\mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\mu_i}(f)\big)^2 \le O\big(\varepsilon^{3/2} \omega^3(\varepsilon)\big) \int |\nabla f|^2 \,\mathrm{d}\mu, \tag{4.6}$$

where the function $\omega(\varepsilon) : \mathbb{R}^+ \to \mathbb{R}^+$ in (4.5) *and* (4.6) *is smooth and monotone satisfying*

$$\omega(\varepsilon) \ge |\log \varepsilon|^{1/2} \qquad \text{for } \varepsilon < 1.$$

The first step toward the proof of Lemma 4.6 is the following statement.

LEMMA 4.7. *Let ν_i be a probability measure satisfying $\nu_i \ll \mu_i$. Moreover, if μ_i satisfies* $\mathrm{PI}(\varrho_i)$ *for some $\varrho_i > 0$, then the following estimate holds*:

$$\big(\mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\mu_i}(f)\big)^2 \le \frac{1}{\varrho_i} \operatorname{var}_{\mu_i}\left(\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\right) \int |\nabla f|^2 \,\mathrm{d}\mu_i. \tag{4.7}$$

PROOF. The result is a consequence from the representation of the mean-difference as a covariance. Therefore, we note that $\mathrm{d}\nu_i = \frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\,\mathrm{d}\mu_i$ since $\nu_i \ll \mu_i$ and use the Cauchy–Schwarz inequality for the covariance

$$
\begin{aligned}
\big(\mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\mu_i}(f)\big)^2 &= \int f \frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\,\mathrm{d}\mu_i - \int f\,\mathrm{d}\mu_i \int \frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\,\mathrm{d}\mu_i \\
&= \operatorname{cov}^2_{\mu_i}\left(\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}, f\right) \le \operatorname{var}_{\mu_i}\left(\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\right) \operatorname{var}_{\mu_i}(f).
\end{aligned}
$$

Using the fact that μ_i satisfies a PI results in (4.7). $\square$

The above lemma tells us that we only need to construct ν_i approximating μ_i in variance for $i = 1, \dots, M$. The following lemma provides exactly this.

LEMMA 4.8 (Approximation in variance). *Let the measures ν_i be given by (4.4). Then the partition sum Z_{ν_i} satisfies for ε small enough*

$$
Z_{\nu_i} = (2\pi\varepsilon)^{n/2}\sqrt{\det \Sigma_i}\big(1 + O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big)\big). \tag{4.8}
$$

Additionally, ν_i approximates μ_i in variance, that is,

$$
\operatorname{var}_{\mu_i}\left(\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\right) = O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big). \tag{4.9}
$$

PROOF. The proof of (4.8) reduces to an estimate of a Gaussian integral on the complementary domain $\mathbb{R}^n \setminus E_i$. We deduced this estimate already in the proof of Lemma 3.13. By the same argument, we deduce

$$
Z_{\nu_i} = (2\pi\varepsilon)^{n/2}\sqrt{\det \Sigma_i}\big(1 + O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big)\big).
$$

Since μ_i comes from the restriction to an admissible partition according to Definition 2.1

$$
Z_{\mu_i} = Z_i Z_\mu \exp\left(\frac{H(m_i)}{\varepsilon}\right) \overset{(2.1)}{=} Z_{\nu_i}\big(1 + O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big)\big). \tag{4.10}
$$

The relative density of ν_i w.r.t. μ_i can be estimated by Taylor expanding H around m_i. By the definition of ν_i given in (4.4), we obtain that $\Sigma_i^{-1}[y - m_i] - H_i(y) = O(|y - m_i|^3)$. This observation together with (4.10) leads to

$$
\begin{aligned}
\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}(y) &= \frac{Z_{\mu_i}}{Z_{\nu_i}} e^{-\Sigma_i^{-1}[y-m_i]/(2\varepsilon) + H_i(y)/(2\varepsilon)} \mathbb{1}_{E_i}(y) = \frac{Z_{\mu_i}}{Z_{\nu_i}} e^{O(|y-m_i|^3)/\varepsilon} \mathbb{1}_{E_i}(y) \\
&= 1 + O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big).
\end{aligned}
$$

Now, the conclusion directly follows from the definition of the variance

$$\begin{aligned}\operatorname{var}_{\mu_i}\left(\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\right) &= \int_{E_i}\left(\frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\right)^2 \mathrm{d}\mu_i - \left(\int \frac{\mathrm{d}\nu_i}{\mathrm{d}\mu_i}\,\mathrm{d}\mu_i\right)^2 \\ &= \int_{E_i} 1 + O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big)\,\mathrm{d}\mu_i - \left(\int_{E_i}\mathrm{d}\nu_i\right)^2 \\ &\le 1 + O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big) - 1 = O\big(\sqrt{\varepsilon}\omega^3(\varepsilon)\big). \qquad \square\end{aligned}$$

PROOF OF LEMMA 4.6. A combination of Lemma 4.7 and Lemma 4.8 together with the assumption $\varrho_i^{-1} = O(\varepsilon)$ immediately reveals

$$\big(\mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\mu_i}(f)\big)^2 \overset{(4.7),(4.9)}{\le} O\big(\varepsilon^{3/2}\omega^3(\varepsilon)\big)\int |\nabla f|^2\,\mathrm{d}\mu_i. \qquad \square$$

4.3. *Affine transport interpolation*. The aim of this section is to estimate $(\mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\nu_j}(f))^2$ with the help of the weighted transport distance $\mathcal{T}_\mu(\nu_i, \nu_j)$ introduced in Section 4.1 and is formulated in Lemma 4.11. For the proof of Lemma 4.11, we construct an explicit transport interpolation between ν_i and ν_j w.r.t. the measure μ. We start with a class of possible transport interpolations and optimize the weighted transport cost in this class.

Let us state the main idea of this optimization procedure. Therefore, we recall that the measures ν_i and ν_j are truncated Gaussians by the approximation we have done in the previous Section 4.2. Hence, the measures ν_i and ν_j are characterized by their mean and covariance matrix. We will choose the transport interpolation (cf. Section 4.3.1) such that the push forward measures $\nu_s := (\Phi_s)_\sharp \nu_0$ are again truncated Gaussians. Hence, it is sufficient to optimize among all paths γ connecting the minima m_i and m_j and all covariance matrices interpolating between Σ_i and Σ_j.

4.3.1. *Definition of regular affine transport interpolations*. Let us state in this section the class of transport interpolation among we want to optimize the weighted transport cost.

DEFINITION 4.9 (Affine transport interpolations). Assume that the measures ν_i and ν_j are given by Lemma 4.6. In detail, $\nu_i = \mathcal{N}(m_i, \varepsilon^{-1}\Sigma_i)\llcorner E_i$ and $\nu_j = \mathcal{N}(m_j, \varepsilon^{-1}\Sigma_j)\llcorner E_j$ are truncated Gaussians centered in m_i and m_j with covariance matrices $\varepsilon^{-1}\Sigma_i$ and $\varepsilon^{-1}\Sigma_j$. The restriction E_i and E_j are given for $l = 1, \dots, M$ by the ellipsoids

$$E_l := \big\{x \in \mathbb{R}^n : \big|\Sigma_l^{-1/2}(x - m_l)\big| \le \sqrt{2\varepsilon}\omega(\varepsilon)\big\} \qquad \text{where } \omega(\varepsilon) \ge |\log\varepsilon|^{1/2}.$$

A transport interpolation Φ_s between ν_i and ν_j is called *affine transport interpolation* if there exists:

- an interpolation path $(\gamma_s)_{s\in[0,T]}$ between $m_i = \gamma_0$ and $m_j = \gamma_T$ satisfying

$$\gamma = (\gamma_s)_{s\in[0,T]} \in C^2([0,T],\mathbb{R}^n) \quad \text{and} \quad \forall s \in [0,T] : \dot\gamma_s \in S^{n-1}, \tag{4.11}$$

- an interpolation path $(\Sigma_s)_{s\in[0,T]}$ of covariance matrices between Σ_i and Σ_j satisfying

$$\Sigma = (\Sigma_s)_{s\in[0,T]} \in C^2([0,T],\mathbb{R}^{n\times n}_{\mathrm{sym},+}), \qquad \Sigma_0 = \Sigma_i \text{ and } \Sigma_T = \Sigma_j,$$

such that the transport interpolation $(\Phi_s)_{s\in[0,T]}$ is given by

$$\Phi_s(x) = \Sigma_s^{1/2}\Sigma_0^{-1/2}(x - m_0) + \gamma_s. \tag{4.12}$$

Since the cost density $\mathcal{A}$ given by (4.3) is invariant under rescaling of time (cf. Remark 4.4), one can always assume that the interpolation path γ_s is parameterized by arc-length. Hence, the condition $\dot\gamma_s \in S^{n-1}$ [cf. (4.11)] is not restricting.

We want to emphasize that for an affine transport interpolation $(\Phi_s)_{s\in[0,T]}$ the push forward measure $(\Phi_s)_\sharp\nu_0 = \nu_s$ is again a truncated Gaussian $\mathcal{N}(\gamma_s, \varepsilon^{-1}\Sigma_s)\llcorner E_s$, where E_s is the support of ν_s being again an ellipsoid in $\mathbb{R}^n$ given by

$$E_s = \{x \in \mathbb{R}^n : |\Sigma_s^{-1/2}(x-\gamma_s)| \le \sqrt{2\varepsilon}\omega(\varepsilon)\}. \tag{4.13}$$

Therewith, the partition sum of ν_s is given by [cf. (4.8)]

$$Z_{\nu_s} = (2\pi\varepsilon)^{n/2}\sqrt{\det \Sigma_s}\big(1 + O(\sqrt{\varepsilon})\big). \tag{4.14}$$

By denoting $\sigma_s = \Sigma_s^{1/2}$ and using the definition (4.12) of the affine transport interpolation $(\Phi_s)_{s\in[0,T]}$, we arrive at the relations

$$\begin{aligned}
\dot\Phi_s(x) &= \dot\sigma_s\sigma_0^{-1}(x - m_0) + \dot\gamma_s,\\
\Phi_s^{-1}(y) &= \sigma_0\sigma_s^{-1}(y-\gamma_s) + m_0,\\
\dot\Phi_s \circ \Phi_s^{-1}(y) &= \dot\sigma_s\sigma_s^{-1}(y-\gamma_s) + \dot\gamma_s.
\end{aligned}$$

Among all possible affine transport interpolations, we are considering only those satisfying the following regularity assumption.

ASSUMPTION 4.10 (Regular affine transport interpolations). An affine transport interpolation $(\gamma_s, \Sigma_s)_{s\in[0,T]}$ belongs to the class of *regular affine transport interpolations* if the length $T < T^*$ is bounded by some uniform $T^* > 0$ large enough. Further, for a uniform constant $c_\gamma > 0$ holds

$$\inf\{r(x,y,z) : x,y,z \in \gamma, x \ne y \ne z \ne x\} \ge c_\gamma, \tag{4.15}$$

where $r(x,y,z)$ denotes the radius of the unique circle through the three distinct points x, y and z. Furthermore, there exists a uniform constant $C_\Sigma \ge 1$ for which

$$C_\Sigma^{-1}\,\mathrm{Id} \le \Sigma_s \le C_\Sigma\,\mathrm{Id} \quad \text{and} \quad \|\dot\Sigma_s\| \le C_\Sigma. \tag{4.16}$$

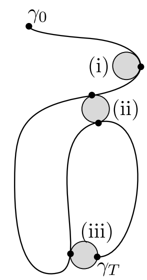

FIG. 3. *Global radius of curvature.*

The infimum in condition (4.15) is called *global radius of curvature* (cf. [20]). It ensures that a small neighborhood of size $\frac{c_\gamma}{2}$ around γ is not self-intersecting, since the infimum can only be attained for the following three cases (cp. Figure 3):

(i) All three points in a minimizing sequence of (4.15) coalesce to a point at which the radius of curvature is minimal.

(ii) Two points coalesce to a single point and the third converges to another point, such that the both points are a pair of closest approach.

(iii) Two points coalesce to a single point and the third converges to the starting or ending point of γ.

In the following calculations, there often occurs a multiplicative error of the form $1 + O(\sqrt{\varepsilon}\omega^3(\varepsilon))$. Therefore, let us introduce for convenience the notation "$\approx$" meaning "$=$" up to the multiplicative error $1 + O(\sqrt{\varepsilon}\omega^3(\varepsilon))$. The symbols "$\lesssim$" and "$\gtrsim$" have the analogous meaning.

Now, we can formulate the key ingredient for the proof of Theorem 2.12, namely the estimation of the weighted transport distance $\mathcal{T}_\mu(\nu_i, \nu_j)$.

LEMMA 4.11. *Assume that ν_i and ν_j are given by Lemma 4.6. Then the weighted transport distance $\mathcal{T}_\mu(\nu_i, \nu_j)$ can be estimated as*

$$
\begin{aligned}
\mathcal{T}_\mu^2(\nu_i, \nu_j) &= \inf_{\Phi_s} \int \left(\int_0^1 |\dot{\Phi}_s \circ \Phi_s^{-1}| \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu} \,\mathrm{d}s \right)^2 \mathrm{d}\mu \\
&\le \inf_{\Psi_s} \int \left(\int_0^1 |\dot{\Psi}_s \circ \Psi_s^{-1}| \frac{\mathrm{d}\nu_s}{\mathrm{d}\mu} \,\mathrm{d}s \right)^2 \mathrm{d}\mu \\
&\lesssim \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} 2\pi\varepsilon \left(\frac{\sqrt{|\det(\nabla^2 H(s_{i,j})|}}{|\lambda^-(s_{i,j})|} + \frac{T(C_\Sigma)^{(n-1)/2}}{\sqrt{2\pi\varepsilon}} e^{-\omega^2(\varepsilon)} \right) \\
&\quad \times e^{H(s_{i,j})/\varepsilon},
\end{aligned}
\tag{4.17}
$$

where the infimum over Ψ_s only considers regular affine transport interpolations Ψ_s in the sense of Assumption 4.10.

In particular, if we choose $\omega(\varepsilon) \geq |\log \varepsilon|^{1/2}$, which is enforced by Lemma 4.6, we get the estimate

$$\mathcal{T}_\mu^2(\nu_i, \nu_j) \leq \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} \frac{2\pi\varepsilon\sqrt{|\det(\nabla^2 H(s_{i,j})|}}{|\lambda^-(s_{i,j})|} e^{H(s_{i,j})/\varepsilon} \times \bigl(1 + O\bigl(\sqrt{\varepsilon}\omega^3(\varepsilon)\bigr)\bigr). \tag{4.18}$$

Before turning to the proof of Lemma 4.11, we want to anticipate the structure of the affine transport interpolation (γ, Σ) which realizes the desired estimate (4.18): Having a closer look at the structure of the weighted transport distance $\mathcal{T}_\mu^2(\nu_i, \nu_j)$, it becomes heuristically clear that the mass should be transported from E_i to E_j over the saddle point $s_{i,j}$ into the direction of the eigenvector to the negative eigenvalue $\lambda^-(s_{i,j})$ of $\nabla^2 H(s_{i,j})$. There, only the region around the saddle gives the main contribution to the estimate (4.18). Then we only have one more free parameter to choose for our affine transport interpolation (γ, Σ): It is the covariance structure Σ_{τ^*} of the interpolating truncated Gaussian measure ν_{τ^*} at the passage time τ^* at the saddle point $s_{i,j}$. In the proof of Lemma 4.11 below, we will see by an optimization procedure that the best Σ_{τ^*} is given by $\Sigma_{\tau^*}^{-1} = \nabla^2 H(s_{i,j})$, restricted to the stable subspace $\nabla^2 H(s_{i,j})$.

The proof of Lemma 4.11 presents the core of the proof of the Eyring–Kramers formulas and consists of three steps carried out in the following sections:

- In Section 4.3.2, we carry out some preparatory work: We introduce tube coordinates on the support of the transport cost $\mathcal{A}$ given by (4.3) (cf. Lemma 4.12), we deduce a pointwise estimate on the transport cost $\mathcal{A}$ and we give a rough a priori estimate on the transport cost $\mathcal{A}$.
- In Section 4.3.3, we split the transport cost into a transport cost around the saddle and the complement. We also estimate the transport cost of the complement yielding the second summand in the desired estimate (4.17).
- In Section 4.3.4, we finally deduce a sharp estimate of the transport cost around the saddle yielding the first summand in the desired estimate (4.17).

4.3.2. *Preparations and auxiliary estimates.* The main reason for making the regularity Assumption 4.10 on affine transport interpolations is that we can introduce tube coordinates around the path γ as illustrated in Figure 4. In these coordinates, the calculation of the cost density $\mathcal{A}$ given by (4.3) becomes a lot handier.

We start with defining the caps E_0^- and E_T^+ as

$$E_0^- := \{x \in E_0 : \langle x - \gamma_0, \dot{\gamma}_0\rangle < 0\} \quad \text{and} \quad E_T^+ := \{x \in E_T : \langle x - \gamma_T, \dot{\gamma}_T\rangle > 0\}.$$

The caps E_0^- and E_T^+ have no contribution to the total cost but unfortunately need some special treatment. Further, we define the slices V_s with $s \in [0, T]$

$$V_s = \{x \in \operatorname{span}\{\dot{\gamma}_s\}^\perp : |\Sigma_s^{-1/2} x| \leq \sqrt{2\varepsilon}\omega(\varepsilon)\}.$$

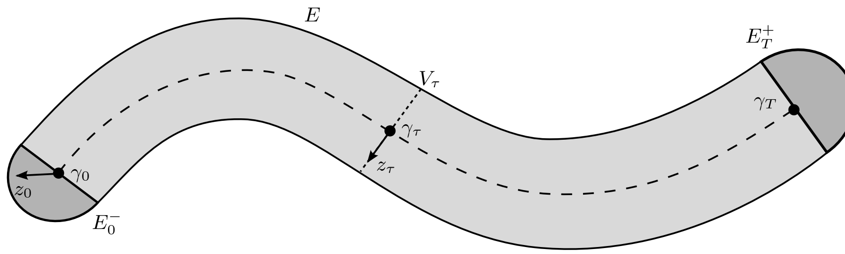

FIG. 4. *The support of $\mathcal{A}$ in tube coordinates.*

In span V_s, we can choose a basis $e_s^2, \dots, e_s^n$ smoothly depending on the parameter s. In particular, there exists a family $(Q_s)_{s\in[0,T]} \in C^2([0,T], SO(n))$ satisfying the same regularity assumption as the family $(\Sigma_\tau)_{\tau\in[0,T]}$ such that

$$Q_s e^1 = \dot{\gamma}_s, \qquad Q_s e^i = e_s^i \qquad \text{for } i = 2, \dots, n, \tag{4.19}$$

where $(e^1, \dots, e^n)$ is the canonical basis of $\mathbb{R}^n$.

Let use now define the tube E as

$$E = \bigcup_{s\in[0,T]} (\gamma_s + V_s).$$

The support of the cost density $\mathcal{A}$ given by (4.3) is now given by

$$\operatorname{supp} \mathcal{A} = E_0^- \cup E \cup E_T^+. \tag{4.20}$$

By the definition (4.13) of E_s and the uniform bound (4.16) on Σ_s holds

$$\operatorname{diam} V_s \le 2\sqrt{2\varepsilon C_\Sigma}\,\omega(\varepsilon). \tag{4.21}$$

Therewith, we find

$$\operatorname{supp} \mathcal{A} \subset B_{2\sqrt{2\varepsilon C_\Sigma}\omega(\varepsilon)}\big((\gamma_\tau)_{\tau\in[0,T]}\big) := \big\{x \in \mathbb{R}^n : |x - \gamma_\tau| \le 2\sqrt{2\varepsilon C_\Sigma}\,\omega(\varepsilon)\big\}.$$

The assumption (4.13) ensures that $B_{2\sqrt{2\varepsilon C_\Sigma}\omega(\varepsilon)}((\gamma_\tau)_{\tau\in[0,T]})$ is not self-intersecting for any ε small enough. The next lemma just states that by changing to tube coordinates in E one can asymptotically neglect the Jacobian determinant $\det J$.

LEMMA 4.12 (Change of coordinates). *The change of coordinates $(\tau, z) \mapsto x = \gamma_\tau + z_\tau$ with $z_\tau \in V_\tau$ satisfies for any function ξ on E*

$$\int_E \xi(x)\,\mathrm{d}x \approx \int_0^T \int_{V_\tau} \xi(\gamma_\tau + z_\tau)\,\mathrm{d}z_\tau\,\mathrm{d}\tau.$$

PROOF. We use the representation of the tube coordinates via (4.19). Therewith, it holds that $x = \gamma_\tau + Q_\tau z$, where $z \in \{0\} \times \mathbb{R}^{n-1}$. Then the Jacobian J of the coordinate change $x \mapsto (\tau, Q_\tau z)$ is given by

$$J = \big(\dot{\gamma}_\tau + \dot{Q}_\tau z, (Q_\tau)_2, \dots, (Q_\tau)_n\big) \in \mathbb{R}^{n\times n},$$

where $(Q_\tau)_i$ denotes the ith column of Q_τ. By the definition (4.19) of Q_τ follows $\dot\gamma_\tau = (Q_\tau)_1$. Hence, we have the representation $J = Q_\tau + \dot Q_\tau z \otimes e_1$. The determinant of J is then given by

$$\det(Q_\tau + \dot Q_\tau z \otimes e_1) = \underbrace{\det(Q_\tau)}_{=1}\det\bigl(\mathrm{Id} + (Q_\tau^\top \dot Q_\tau z)\otimes e_1\bigr) = 1 + \bigl(Q_\tau^\top \dot Q_\tau z\bigr)_1.$$

By Assumption 4.10 holds $\|\dot Q_\tau\| \le C_\Sigma$ implying $(Q_\tau^\top \dot Q_\tau z)_{1,1} = O(z)$. Since $Q_\tau z \in V_\tau$, we get $O(z) = O(\sqrt{\varepsilon}\omega(\varepsilon))$ by (4.21). Hence, we get

$$\det J = 1 + O\bigl(\sqrt{\varepsilon}\omega(\varepsilon)\bigr),$$

which concludes the proof. □

An important tool is the following auxiliary estimate.

LEMMA 4.13 (Pointwise estimate of the cost-density $\mathcal{A}$). *For $x \in \operatorname{supp}\mathcal{A}$, we define*

$$\tau = \operatorname*{arg\,min}_{s\in[0,T]} |x - \gamma_s| \quad \text{and} \quad z_\tau := x - \gamma_\tau. \tag{4.22}$$

Then the following estimate holds:

$$\begin{aligned}\mathcal{A}(x) &\lesssim (2\pi\varepsilon)^{-(n-1)/2}\sqrt{\det\nolimits_{1,1}\bigl(Q_\tau^\top \tilde\Sigma_\tau^{-1} Q_\tau\bigr)}e^{-\tilde\Sigma_\tau^{-1}[z_\tau]/(2\varepsilon)} \\ &=: P_\tau e^{-\tilde\Sigma_\tau^{-1}[z_\tau]/(2\varepsilon)},\end{aligned} \tag{4.23}$$

where Q_τ is defined in (4.19) and $\tilde\Sigma_\tau^{-1}$ is given by

$$\tilde\Sigma_\tau^{-1} = \Sigma_\tau^{-1} - \frac{1}{\Sigma_\tau^{-1}[\dot\gamma_\tau]}\Sigma_\tau^{-1}\dot\gamma_\tau \otimes \Sigma_\tau^{-1}\dot\gamma_\tau. \tag{4.24}$$

Further, $\det_{1,1} A$ *denotes the determinant of the matrix obtained from A removing the first row and column.*

REMARK 4.14. With a little bit of additionally work, one could show that (4.23) holds with "$\approx$" instead of "$\lesssim$." It follows from (4.24) that the matrix $\tilde\Sigma_\tau^{-1}$ is positive definite. Hence, $\mathcal{A}$ is an $\mathbb{R}^{n-1}$-dimensional Gaussian on the slice $\gamma_\tau + V_\tau$ up to approximation errors.

PROOF OF LEMMA 4.13. By the regularity Assumption 4.10 on the transport interpolation, we find that for all $x \in \operatorname{supp}\mathcal{A}$ holds uniformly

$$I_T(x) := \{s : E_s \ni x\} \qquad \text{satisfies } \mathcal{H}^1\bigl(I_T(x)\bigr) = O\bigl(\sqrt{\varepsilon}\omega(\varepsilon)\bigr).$$

This allows us to linearize the transport interpolation around τ given in (4.22). It holds for s such that $x \in E_s$

$$\begin{aligned}\Sigma_s^{-1}[x - \gamma_s] &= \Sigma_\tau^{-1}[\gamma_\tau + z_\tau - \gamma_s] + O\bigl(\varepsilon^{3/2}\omega^3(\varepsilon)\bigr) \\ &= \Sigma_\tau^{-1}\bigl[(\tau - s)\dot\gamma_\tau + z_\tau\bigr] + O\bigl(\varepsilon^{3/2}\omega^3(\varepsilon)\bigr).\end{aligned} \tag{4.25}$$

For similar reasons, we can linearize the determinant $\det \Sigma_s$ and have $\det \Sigma_s = \det \Sigma_\tau + O(\sqrt{\varepsilon}\omega(\varepsilon))$. Finally, we have the following bound on the transport speed:

$$
\begin{aligned}
|\dot{\Phi}_s \circ \Phi_s^{-1}(x)|\mathbb{1}_{E_s}(x) &= |\dot{\sigma}_s \sigma_s^{-1}(x-\gamma_s) + \dot{\gamma}_s|\mathbb{1}_{E_s}(x) \\
&\le (|\dot{\sigma}_s \sigma_s^{-1}(x-\gamma_s)| + |\dot{\gamma}_s|)\mathbb{1}_{E_s}(x) \qquad (4.26)\\
&\le (C_\Sigma |x-\gamma_s| + 1)\mathbb{1}_{E_s}(x) = (1 + O(\sqrt{\varepsilon}\omega(\varepsilon)))\mathbb{1}_{E_s}(x).
\end{aligned}
$$

Let us first consider the case $x \in E$. We use (4.14), (4.25) and (4.26) to arrive with $x = \gamma_\tau + z_\tau$ where $z_\tau \in V_\tau$ at

$$
\begin{aligned}
\mathcal{A}(x) &= \int_{I_T(x)} |\dot{\Phi}_s \circ \Phi_s^{-1}(x)| \frac{1}{Z_{\nu_s}} \exp\Big(-\frac{1}{2\varepsilon}\Sigma_s^{-1}[x-\gamma_s]\Big)\mathbb{1}_{E_s}(x)\,\mathrm{d}s \\
&\le \frac{1}{(2\pi\varepsilon)^{n/2}} \int_{I_T(x)} \frac{1 + O(\sqrt{\varepsilon}\omega(\varepsilon))}{\sqrt{\det \Sigma_s}} \exp\Big(-\frac{1}{2\varepsilon}\Sigma_s^{-1}[x-\gamma_s]\Big)\,\mathrm{d}s \\
&\lesssim \frac{1}{(2\pi\varepsilon)^{n/2}\sqrt{\det \Sigma_\tau}} \int_{\mathbb{R}} \exp\Big(-\frac{1}{2\varepsilon}\Sigma_\tau^{-1}\big[(\tau - s)\dot{\gamma}_\tau + z_\tau\big]\Big)\,\mathrm{d}s \\
&= \frac{\sqrt{\det \Sigma_\tau^{-1}}}{(2\pi\varepsilon)^{n/2}} \frac{\sqrt{2\pi\varepsilon}}{\sqrt{\Sigma_\tau^{-1}[\dot{\gamma}_\tau]}} \exp\Big(-\frac{1}{2\varepsilon}\tilde{\Sigma}_\tau^{-1}[z_\tau]\Big)\big(1 + O(\sqrt{\varepsilon}\omega^3(\varepsilon))\big),
\end{aligned}
$$

where the last step follows by an application of a partial Gaussian integration (cf. Lemma C.1). Finally, by using the relation (C.2), we get that

$$
\frac{\det \Sigma_\tau^{-1}}{\Sigma_\tau^{-1}[\dot{\gamma}_\tau]} = \det_{1,1}(Q_\tau^\top \tilde{\Sigma}_\tau^{-1} Q_\tau),
$$

and conclude the hypothesis for this case.

Let us now consider the case $x \in E_0^- \cup E_T^+$. For convenience, we only consider the case $x \in E_0^-$. By the definition of E_0^- holds $\tau = 0$. The integration domain $I_T(x)$ is now given by

$$
I_T(x) = [0, s^*) \qquad \text{with } s^* = O(\sqrt{\varepsilon}\omega(\varepsilon)). \tag{4.27}
$$

Therewith, we can estimate $\mathcal{A}(x)$ in the same way as for $x \in E$ and conclude the proof. $\square$

We only need one more ingredient for the proof of Lemma 4.11. It is an a priori estimate on the cost density $\mathcal{A}$.

LEMMA 4.15 (A priori estimates for the cost density $\mathcal{A}$). *For $\mathcal{A}$, it holds*:

$$
\int \mathcal{A}(x)\,\mathrm{d}x \lesssim T \quad \textit{and} \tag{4.28}
$$

$$
\mathcal{A}(x) \lesssim \Big(\frac{C_\Sigma}{2\pi\varepsilon}\Big)^{(n-1)/2} \qquad \textit{for } x \in \operatorname{supp}\mathcal{A}. \tag{4.29}
$$

PROOF. Let us first consider the estimate (4.28). It follows from the characterization (4.20) of the support of $\mathcal{A}$ that

$$\int \mathcal{A}(x)\,\mathrm{d}x = \int_E \mathcal{A}(x)\,\mathrm{d}x + \int_{E_0^- \cup E_T^+} \mathcal{A}(x)\,\mathrm{d}x. \tag{4.30}$$

Now, we estimate the first term on the right-hand side of the last identity. Using the change to tube coordinates of Lemma 4.12 and noting that the upper bound (4.23) is a $(n-1)$-dimensional Gaussian density on V_τ for $\tau \in [0, T]$, we can easily infer that

$$\int_E \mathcal{A}(x)\,\mathrm{d}x \lesssim |\gamma| = T.$$

Let us turn to the second term on the right-hand side of (4.30). For convenience, we only consider the integral w.r.t. the cap E_0^-. It follows from (4.26) and (4.27) that

$$\begin{aligned}\int_{E_0^-} \mathcal{A}(x)\,\mathrm{d}x &\lesssim \int_{E_0^-} \int_0^1 \nu_s(x)\,\mathrm{d}s\,\mathrm{d}x = \int_0^{s^*} \int_{E_0^-} \nu_s(x)\,\mathrm{d}x\,\mathrm{d}s \\ &\lesssim \int_0^{s^*} \int \nu_s(x)\,\mathrm{d}x\,\mathrm{d}s = s^* = O\big(\sqrt{\varepsilon}\omega(\varepsilon)\big),\end{aligned}$$

which yields the desired statement (4.28).

Let us now consider the estimate (4.29). Note by Remark 4.14 the matrix $\tilde{\Sigma}_\tau^{-1}$ given by (4.24) is positive definite and the matrix we subtract is also positive definite. Therefore, it holds in the sense of quadratic forms

$$0 < \tilde{\Sigma}_\tau^{-1} = \Sigma_\tau^{-1} - \frac{1}{\Sigma_\tau^{-1}[\dot{\gamma}_\tau]} \Sigma_\tau^{-1}\dot{\gamma}_\tau \otimes \Sigma_\tau^{-1}\dot{\gamma}_\tau \leq \Sigma_\tau^{-1}.$$

Now, the uniform bound (4.16) yields

$$\sqrt{\det_{1,1}\big(Q_\tau^\top \tilde{\Sigma}_\tau^{-1} Q_\tau\big)} \leq C_\Sigma^{(n-1)/2}.$$

Then the desired statement (4.29) follows directly from the estimate (4.23). □

4.3.3. *Proof of Lemma 4.11*: *Reduction to neighborhood around the saddle*. Firstly, observe that from (4.29) follows the a priori estimate

$$\frac{\mathcal{A}^2(x)}{\mu(x)} \lesssim \left(\frac{C_\Sigma}{2\pi\varepsilon}\right)^{n-1} Z_\mu e^{1/\varepsilon H(x)}. \tag{4.31}$$

Hence, on an exponential scale, the leading order contribution to the cost comes from neighborhoods of points where $H(x)$ is large. Therefore, we want to make the set, where H is comparable to its value at the optimal connecting saddle $s_{i,j}$, as small as possible. For this purpose, let us define the following set:

$$\Xi_{\gamma,\Sigma} := \{x \in \operatorname{supp}\mathcal{A} : H(x) \geq H(s_{i,j}) - \varepsilon\omega^2(\varepsilon)\}. \tag{4.32}$$

Therewith, we obtain by denoting the complement $\Xi^c_{\gamma,\Sigma} := \operatorname{supp} \mathcal{A} \setminus \Xi_{\gamma,\Sigma}$ the splitting

$$\mathcal{T}^2_\mu(\nu_i, \nu_j) \le \int_{\Xi_{\gamma,\Sigma}} \frac{\mathcal{A}^2(x)}{\mu(x)} \,\mathrm{d}x + \int_{\Xi^c_{\gamma,\Sigma}} \frac{\mathcal{A}^2(x)}{\mu(x)} \,\mathrm{d}x.$$

The integral on $\Xi^c_{\gamma,\Sigma}$ can be estimated with the a priori estimate (4.31) and Lemma 4.15 as follows:

$$\begin{aligned}
\int_{\Xi^c_{\gamma,\Sigma}} \frac{\mathcal{A}^2(x)}{\mu(x)} \,\mathrm{d}x &\overset{(4.32)}{\le} Z_\mu e^{H(s_{i,j})/\varepsilon - \omega^2(\varepsilon)} \int_{\Xi^c_{\gamma,\Sigma}} \mathcal{A}^2(x) \,\mathrm{d}x \\
&\overset{(4.29)}{\lesssim} Z_\mu e^{H(s_{i,j})/\varepsilon - \omega^2(\varepsilon)} \left(\frac{C_\Sigma}{2\pi\varepsilon}\right)^{(n-1)/2} \int \mathcal{A}(x) \,\mathrm{d}x \\
&\overset{(4.28)}{\lesssim} Z_\mu e^{H(s_{i,j})/\varepsilon - \omega^2(\varepsilon)} \left(\frac{C_\Sigma}{2\pi\varepsilon}\right)^{(n-1)/2} T.
\end{aligned} \tag{4.33}$$

We observe that estimate (4.33) is the second summand in the desired bound (4.17).

4.3.4. *Proof of Lemma* 4.11: *Cost estimate around the saddle*. The aim of this subsection is to deduce the estimate

$$\int_{\Xi_{\gamma,\Sigma}} \frac{\mathcal{A}^2(x)}{\mu(x)} \,\mathrm{d}x \lesssim \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} e^{H(s_{i,j})/\varepsilon} \frac{2\pi\varepsilon\sqrt{|\det(\nabla^2 H(s_{i,j}))|}}{|\lambda^-(s_{i,j})|}. \tag{4.34}$$

Note that this estimate would yield the missing ingredient for the verification of the desired estimate (4.17).

By the nondegeneracy Assumption 1.7, we can assume that ε is small enough such that $E^-_0 \cup E^+_T \subset \Xi^c_{\gamma,\Sigma}$. Hence, it follows that $\Xi_{\gamma,\Sigma} \subset E$. We claim that the transport interpolation Φ_s can be chosen such that there exists a connected subinterval $I_T \subset [0, T]$ satisfying

$$\Xi_{\gamma,\Sigma} \subset \bigcup_{s \in I_T} (V_s + \gamma_s) \quad \text{and} \quad \mathcal{H}^1(I_T) = O\big(\sqrt{\varepsilon}\omega(\varepsilon)\big). \tag{4.35}$$

Indeed, the level set $\{x \in \mathbb{R}^n : H(x) \le H(s_{i,j}) - \varepsilon\omega^2(\varepsilon)\}$ consists of at least two connected components M_i and M_j such that $m_i \in M_i$ and $m_j \in M_j$. Further, it holds

$$\operatorname{dist}(M_i, M_j) = \inf_{x \in M_i, y \in M_j} |x - y| = O\big(\sqrt{\varepsilon}\omega(\varepsilon)\big),$$

which follows from expanding H around $s_{i,j}$ in direction of the eigenvector corresponding to the negative eigenvalue of $\nabla^2 H(s_{i,j})$. We can choose the path γ in direction of this eigenvector in a neighborhood of size $O(\sqrt{\varepsilon}\omega(\varepsilon))$ around $s_{i,j}$, which shows (4.35).

Combining the covering (4.35) and Lemma 4.12 yields the estimate

$$\int_{\Xi_{\gamma,\Sigma}} \frac{\mathcal{A}^2(x)}{\mu(x)}\,\mathrm{d}x \lesssim \int_{I_T}\int_{V_\tau} \frac{\mathcal{A}^2(\gamma_\tau + z_\tau)}{\mu(\gamma_\tau + z_\tau)}\,\mathrm{d}z_\tau\,\mathrm{d}\tau. \tag{4.36}$$

Recalling the definition (4.19) of the family of rotations $(Q_\tau)_{\tau\in[0,T]}$, it holds that $z_\tau = Q_\tau z$ with $z \in \{0\}\times\mathbb{R}^{n-1}$. Hence, the following relation holds:

$$\begin{aligned}&\int_{I_T}\int_{V_\tau} \frac{\mathcal{A}^2(\gamma_\tau + z_\tau)}{\mu(\gamma_\tau + z_\tau)}\,\mathrm{d}z_\tau\,\mathrm{d}\tau\\ &\quad= \int_{\{0\}\times\mathbb{R}^{n-1}}\int_{I_T} \mathbb{1}_{V_\tau}(Q_\tau z)\frac{\mathcal{A}^2(\gamma_\tau + Q_\tau z)}{\mu(\gamma_\tau + Q_\tau z)}\,\mathrm{d}\tau\,\mathrm{d}z.\end{aligned} \tag{4.37}$$

The next step is to rewrite $H(\gamma_\tau + Q_\tau z)$. We assume, that γ actually passes the saddle $s_{i,j}$ at time $\tau^* \in (0,T)$. Then, by the reason that $|z_\tau| = O(\sqrt{\varepsilon}\omega(\varepsilon))$ for $z_\tau \in V_\tau$ and the global nondegeneracy assumption (1.4), we can Taylor expand $H(\gamma_\tau + z_\tau)$ around $s_{i,j} = \gamma_{\tau^*}$ for $\tau \in I_T$ and $z_\tau = Q_\tau z \in V_\tau$. More precisely, we get

$$\begin{aligned}&H(\gamma_\tau + Q_\tau z) - H(s_{i,j})\\ &\quad= \tfrac12 \nabla^2 H(s_{i,j})[\gamma_\tau + Q_\tau z - s_{i,j}] + O\big(|\gamma_\tau + Q_\tau z - s_{i,j}|^3\big)\\ &\quad= \tfrac12 \nabla^2 H(s_{i,j})[\gamma_\tau - \gamma_{\tau^*}] + \tfrac12 \nabla^2 H(s_{i,j})[Q_\tau z]\\ &\qquad + \big\langle Q_\tau z, \nabla^2 H(s_{i,j})(\gamma_\tau - \gamma_{\tau^*})\big\rangle + O\big(|\gamma_\tau + Q_\tau z - \gamma_{\tau^*}|^3\big).\end{aligned}$$

Now, further expanding γ_τ and Q_τ in τ leads to

$$\begin{aligned}\gamma_\tau &= \gamma_{\tau^*} + \dot\gamma_{\tau^*}(\tau - \tau^*) + O\big(|\tau - \tau^*|\big) \quad \text{and}\\ Q_\tau z &= Q_{\tau^*} z + O\big(|\tau - \tau^*||z|\big).\end{aligned}$$

For the expansion of H, we arrive at the identity

$$\begin{aligned}&H(\gamma_\tau + Q_\tau z) - H(s_{i,j})\\ &\quad= \tfrac12 \nabla^2 H(s_{i,j})\big[\dot\gamma_{\tau^*}(\tau - \tau^*) + O\big(|\tau - \tau^*|^2\big)\big]\\ &\qquad + \tfrac12 \nabla^2 H(s_{i,j})\big[Q_{\tau^*} z + O\big(|\tau - \tau^*||z|\big)\big]\\ &\qquad + \big\langle Q_{\tau^*} z + O\big(|\tau - \tau^*||z|\big), \nabla^2 H(s_{i,j})\big(\dot\gamma_{\tau^*}(\tau - \tau^*) + O\big(|\tau - \tau^*|^2\big)\big)\big\rangle\\ &\qquad + O\big(|\gamma_\tau + Q_\tau z - \gamma_{\tau^*}|^3\big)\\ &\quad= \tfrac12 \nabla^2 H(s_{i,j})[\dot\gamma_{\tau^*}](\tau - \tau^*)^2 + \tfrac12 \nabla^2 H(s_{i,j})[Q_{\tau^*} z]\\ &\qquad + \big\langle Q_{\tau^*} z, \nabla^2 H(s_{i,j})\dot\gamma_{\tau^*}\big\rangle(\tau - \tau^*)\\ &\qquad + O\big(|\tau - \tau^*|^3, |z||\tau - \tau^*|^2, |z|^2|\tau - \tau^*|, |z|^3\big).\end{aligned}$$

Using $|\tau - \tau^*| = O(\sqrt{\varepsilon}\omega(\varepsilon))$ and $|z| = O(\sqrt{\varepsilon}\omega(\varepsilon))$, we obtain for the error the estimate

$$\begin{aligned} &O\big(|\tau - \tau^*|^3, |z||\tau - \tau^*|^2, |z|^2|\tau - \tau^*|, |z|^3\big) \\ &\quad = O\big(\varepsilon^{3/2}\omega^3(\varepsilon)\big). \end{aligned}$$

The term $\langle Q_{\tau^*}z, \nabla^2 H(s_{i,j})\dot{\gamma}_{\tau^*}\rangle(\tau - \tau^*)$ in the expansion of H has no sign and has to vanish. This is only the case, if we choose $\dot{\gamma}_{\tau^*}$ as an eigenvector of $\nabla^2 H(s_{i,j})$ to the negative eigenvalue $\lambda^-(s_{i,j})$, because then

$$\big\langle Q_{\tau^*}z, \nabla^2 H(s_{i,j})\dot{\gamma}_{\tau^*}\big\rangle\big(\tau - \tau^*\big) = \lambda^-(s_{i,j})\langle Q_{\tau^*}z, \dot{\gamma}_{\tau^*}\rangle = 0.$$

Additionally, by this choice of $\dot{\gamma}_{\tau^*}$ the quadratic form $\nabla^2 H(s_{i,j})[\dot{\gamma}_{\tau^*}]$ evaluates to

$$\nabla^2 H(s_{i,j})[\dot{\gamma}_{\tau^*}] = \lambda^-(s_{i,j})|\dot{\gamma}_{\tau^*}|^2 = \lambda^-(s_{i,j}).$$

Therefore, we deduced the desired rewriting of $H(\gamma_\tau + Q_\tau z)$ as

$$\begin{aligned} H(\gamma_\tau + Q_\tau z) = H(s_{i,j}) - \big|\lambda^-(s_{i,j})\big|\big(\tau - \tau^*\big)^2 \\ + \tfrac{1}{2}\nabla^2 H(s_{i,j})[Q_{\tau^*}z] + O\big(\varepsilon^{3/2}\omega^3(\varepsilon)\big). \end{aligned} \tag{4.38}$$

From the regularity assumptions on the transport interpolation, we can deduce that

$$\begin{aligned} \tilde{\Sigma}_\tau^{-1}[Q_\tau z] &= \tilde{\Sigma}_{\tau^*}^{-1}[Q_\tau z] + O\big(|\tau - \tau^*||z|^2\big) \\ &= \tilde{\Sigma}_{\tau^*}^{-1}\big[Q_{\tau^*}z + O\big(|\tau - \tau^*||z|\big)\big] + O\big(|\tau - \tau^*||z|^2\big) \\ &= \tilde{\Sigma}_{\tau^*}^{-1}[Q_{\tau^*}z] + O\big(\varepsilon^{3/2}\omega^3(\varepsilon)\big). \end{aligned}$$

Then it follows easily from the definition (4.23) of P_τ that

$$P_\tau \approx P_{\tau^*} \qquad \text{for } \tau \in I_T. \tag{4.39}$$

Applying the cost estimate (4.23) of Lemma 4.13, the representation (4.38) and the identity (4.39) yields the estimate for $(\gamma_\tau + Q_\tau z) \in \Xi_{\gamma,\Sigma}$

$$\begin{aligned} &\frac{\mathcal{A}^2(\gamma_\tau + Q_\tau z)}{\mu(\gamma_\tau + Q_\tau z)} \\ &\quad \lesssim Z_\mu e^{H(s_{i,j})/\varepsilon} P_{\tau^*}^2 e^{-(2\tilde{\Sigma}_{\tau^*}^{-1} - \nabla^2 H(s_{i,j}))[Q_{\tau^*}z]/(2\varepsilon) - |\lambda^-(s_{i,j})|(\tau-\tau^*)^2/(2\varepsilon)}. \end{aligned} \tag{4.40}$$

The exponentials are densities of two Gaussian, if we put an additional constraint on the transport interpolation. Namely, we postulate

$$2\tilde{\Sigma}_{\tau^*}^{-1} - \nabla^2 H(s_{i,j}) > 0 \qquad \text{on span } V_{\tau^*}$$

in the sense of quadratic forms. It holds that $\operatorname{span} V_{\tau^*} = Q_{\tau^*}(\{0\} \times \mathbb{R}^{n-1}) = \operatorname{span}\{\dot{\gamma}_{\tau^*}\}^{\perp}$ is stable subspace of $\nabla^2 H(s_{i,j})$. With these preliminary considerations, we finally are able to estimate the right-hand side of (4.37) as follows:

$$
\begin{aligned}
&\int_{\{0\}\times\mathbb{R}^{n-1}} \int_{I_T} \mathbb{1}_{V_\tau}(Q_\tau z) \frac{\mathcal{A}^2(\gamma_\tau + Q_\tau z)}{\mu(\gamma_\tau + Q_\tau z)} \,\mathrm{d}\tau \,\mathrm{d}z \\
&\quad \overset{(4.40)}{\lesssim} Z_\mu e^{H(s_{i,j})/\varepsilon} \int_{\{0\}\times\mathbb{R}^{n-1}} \int_{I_T} P_{\tau^*}^2 \\
&\qquad\qquad \times e^{-(2\tilde{\Sigma}_{\tau^*}^{-1} - \nabla^2 H(s_{i,j}))[Q_{\tau^*} z]/(2\varepsilon) - |\lambda^-(s_{i,j})|(\tau-\tau^*)^2/(2\varepsilon)} \,\mathrm{d}\tau \,\mathrm{d}z \\
&\quad \le Z_\mu e^{H(s_{i,j})/\varepsilon} \frac{\sqrt{2\pi\varepsilon}}{\sqrt{|\lambda^-(s_{i,j})|}} \int_{\{0\}\times\mathbb{R}^{n-1}} P_{\tau^*}^2 e^{-(2\tilde{\Sigma}_{\tau^*}^{-1} - \nabla^2 H(s_{i,j}))[Q_{\tau^*} z]/(2\varepsilon)} \,\mathrm{d}z \\
&\quad = Z_\mu e^{H(s_{i,j})/\varepsilon} \frac{\sqrt{2\pi\varepsilon}}{\sqrt{|\lambda^-(s_{i,j})|}} P_{\tau^*}^2 \frac{(2\pi\varepsilon)^{(n-1)/2}}{\sqrt{\det_{1,1}(Q_{\tau^*}^\top (2\tilde{\Sigma}_{\tau^*}^{-1} - \nabla^2 H(s_{i,j})) Q_{\tau^*})}} \\
&\quad = \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} e^{H(s_{i,j})/\varepsilon} \frac{2\pi\varepsilon}{\sqrt{|\lambda^-(s_{i,j})|}} \underbrace{\frac{\det_{1,1}(Q_{\tau^*}^\top \tilde{\Sigma}_{\tau^*}^{-1} Q_{\tau^*})}{\sqrt{\det_{1,1}(Q_{\tau^*}^\top (2\tilde{\Sigma}_{\tau^*}^{-1} - \nabla^2 H(s_{i,j})) Q_{\tau^*})}}}_{\text{to optimize!}}.
\end{aligned}
\tag{4.41}
$$

The final step consists of optimizing the choice of $\tilde{\Sigma}_{\tau^*}$. Let us use the notation $A = Q_{\tau^*}^\top \tilde{\Sigma}_{\tau^*}^{-1} Q_{\tau^*}$ and $B = Q_{\tau^*}^\top H(s_{i,j}) Q_{\tau^*}$. Then the minimization problem has the structure

$$
\inf_{A\in\mathbb{R}^{n\times n}_{\mathrm{sym},+}} \left\{ \frac{\det_{1,1} A}{\sqrt{\det_{1,1}(2A - B)}} : 2A - B > 0 \text{ on } \{0\} \times \mathbb{R}^{n-1} \right\}.
\tag{4.42}
$$

In the Appendix, we show in Lemma C.2 that the optimal value of (4.42) is attained at $\tilde{\Sigma}_{\tau^*}^{-1} = \nabla^2 H(s_{i,j})$ restricted to V_{τ^*}. The optimal value is given by

$$
\frac{\det_{1,1} A}{\sqrt{\det_{1,1}(2A - B)}} = \sqrt{\det_{1,1}\big(Q_{\tau^*}^\top \nabla^2 H(s_{i,j}) Q_{\tau^*}\big)}.
$$

Because V_{τ^*} is the stable subspace of $\nabla^2 H(s_{i,j})$, it holds

$$
\det_{1,1}\big(Q_{\tau^*}^\top \nabla^2 H(s_{i,j}) Q_{\tau^*}^\top\big) = \frac{\det(\nabla^2 H(s_{i,j}))}{\lambda^-(s_{i,j})} = \frac{|\det(\nabla^2 H(s_{i,j}))|}{|\lambda^-(s_{i,j})|}.
\tag{4.43}
$$

The final step is a combination of (4.36), (4.37), (4.41) and (4.43) to obtain the desired estimate (4.34). This together with (4.33) concludes (4.17) of Lemma 4.11.

4.3.5. *Proof of Lemma* 4.11: *Total error estimate*. For the verification of Lemma 4.11, it is only left to deduce the estimate (4.18). For that purpose, we analyze the error terms in the estimate (4.17) that is,

$$\begin{aligned}&\mathcal{T}_\mu^2(\nu_i,\nu_j)\\ &\quad\lesssim \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} e^{H(s_{i,j})/\varepsilon} 2\pi\varepsilon\biggl(\underbrace{\frac{\sqrt{|\det(\nabla^2 H(s_{i,j}))|}}{|\lambda^-(s_{i,j})|}}_{=O(1)} + \underbrace{\frac{T(C_\Sigma)^{(n-1)/2}}{\sqrt{2\pi\varepsilon}} e^{-\omega^2(\varepsilon)}}_{=O(\varepsilon^{-1/2}e^{-\omega^2(\varepsilon)})}\biggr).\end{aligned}$$

By the choice of $\omega(\varepsilon) \geq |\log\varepsilon|^{1/2}$, enforced by Lemma 4.6, we see that

$$O(\varepsilon^{-1/2}e^{-\omega^2(\varepsilon)}) = O(\sqrt{\varepsilon}).$$

Recalling, that "$\lesssim$" means "$\leq$" up to a multiplicative error of order $1 + O(\sqrt{\varepsilon}\omega^3(\varepsilon))$ we get the desired estimate (4.18)

$$\mathcal{T}_\mu^2(\nu_i,\nu_j) \lesssim \frac{Z_\mu}{(2\pi\varepsilon)^{n/2}} e^{H(s_{i,j})/\varepsilon} 2\pi\varepsilon \frac{\sqrt{|\det(\nabla^2 H(s_{i,j}))|}}{|\lambda^-(s_{i,j})|}\bigl(1 + O(\sqrt{\varepsilon}\omega^3(\varepsilon))\bigr).$$

4.4. *Proof of Theorem* 2.12: *Conclusion of the mean-difference estimate*. With the help of Lemma 4.6 and Lemma 4.11 the proof of Theorem 2.12 is straightforward. We can estimate the mean-differences w.r.t. to the measure μ_i by introducing the means w.r.t. the approximations ν_i and ν_j

$$\begin{aligned}&\bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\bigr)^2\\ &\qquad= \bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\nu_i}(f) + \mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\nu_j}(f) + \mathbb{E}_{\nu_j}(f) - \mathbb{E}_{\mu_j}(f)\bigr)^2.\end{aligned}$$

We apply the Young inequality with a weight that is motivated by the final total multiplicative error term $R(\varepsilon)$ in Theorem 2.12. More precisely,

$$\begin{aligned}&\bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\bigr)^2\\ &\quad\leq \bigl(1 + \varepsilon^{1/2}\omega^3(\varepsilon)\bigr)\bigl(\mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\nu_j}(f)\bigr)^2\\ &\qquad + 2\bigl(1 + \varepsilon^{-1/2}\omega^{-3}(\varepsilon)\bigr)\bigl(\bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\nu_i}(f)\bigr)^2 + \bigl(\mathbb{E}_{\mu_j}(f) - \mathbb{E}_{\nu_j}(f)\bigr)^2\bigr).\end{aligned}$$

Then the estimate (4.6) of Lemma 4.6 yields

$$\begin{aligned}\bigl(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\bigr)^2 &\leq \bigl(1 + \sqrt{\varepsilon}\omega^3(\varepsilon)\bigr)\bigl(\mathbb{E}_{\nu_i}(f) - \mathbb{E}_{\nu_j}(f)\bigr)^2\\ &\qquad + O(\varepsilon)\int |\nabla f|^2 \,\mathrm{d}\mu,\end{aligned} \tag{4.44}$$

which justifies the statement, that the approximation only leads to higher-order error terms in ε. An application of (4.1) to the estimate (4.44) transfers the mean-difference to the Dirichlet form with the help of the weighted transport distance

$$\big(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\big)^2 \le \big((1+\sqrt{\varepsilon}\omega^3(\varepsilon))\mathcal{T}_\mu^2(\nu_i,\nu_j) + O(\varepsilon)\big) \int |\nabla f|^2 \, \mathrm{d}\mu.$$

The weighted transport distance $\mathcal{T}_\mu(\nu_i,\nu_j)$ is dominating the above estimate. Finally, we arrive at the estimate

$$\big(\mathbb{E}_{\mu_i}(f) - \mathbb{E}_{\mu_j}(f)\big)^2 \lesssim \mathcal{T}_\mu^2(\nu_i,\nu_j) \int |\nabla f|^2 \, \mathrm{d}\mu.$$

Now, the Theorem 2.12 follows directly from an application of the estimate (4.18) of Lemma 4.11 and setting $\omega(\varepsilon) = |\log \varepsilon|^{1/2}$.

APPENDIX A: PROPERTIES OF THE LOGARITHMIC MEAN Λ

In this part of the Appendix, we collect some properties of the logarithmic mean $\Lambda(\cdot,\cdot)$. A more complete study can be found in [11].

Let us first recall the definition of $\Lambda(\cdot,\cdot) : \mathbb{R}^+ \times \mathbb{R}^+ \to \mathbb{R}^+$

$$\Lambda(a,b) = \int_0^1 a^s b^{1-s} \, \mathrm{d}s = \begin{cases} \dfrac{a-b}{\log a - \log b}, & a \neq b, \\ a, & a = b. \end{cases} \tag{A.1}$$

The equation (A.1) justifies the statement, that $\Lambda(\cdot,\cdot)$ is a mean, since one immediately recovers the simple bounds $\min\{a,b\} \le \Lambda(a,b) \le \max\{a,b\}$. Moreover, two other immediate properties are:

- $\Lambda(\cdot,\cdot)$ is symmetric
- $\Lambda(\cdot,\cdot)$ is homogeneous of degree one, that is, for $\Lambda(\lambda a, \lambda b) = \lambda \Lambda(a,b)$ for $\lambda > 0$.

The derivatives of $\Lambda(\cdot,\cdot)$ are given by straight-forward calculus

$$\partial_a \Lambda(a,b) = \frac{1 - \Lambda(a,b)/a}{\log a - \log b} > 0 \quad \text{and} \quad \partial_b \Lambda(a,b) = \frac{1 - \Lambda(a,b)/b}{\log b - \log a} > 0.$$

Hence, $\Lambda(\cdot,\cdot)$ is strictly monotone increasing in both arguments.

The following result is almost classical and proven for instance in [11], Theorem 1, [35], Appendix A, and [7].

LEMMA A.1. *The logarithmic mean can be bounded below by the geometric mean and above by the arithmetic mean*

$$\sqrt{ab} \le \Lambda(a,b) \le \frac{a+b}{2}, \tag{A.2}$$

with equality if and only if $a = b$.

The bounds in (A.2) are good, if a is of the same order as b, whereas the following bound is particularly good if $\frac{a}{b}$ becomes very small or very large.

LEMMA A.2. *It holds for $p \in (0,1)$, the following bound*:

$$\frac{\Lambda(p,1-p)}{p(1-p)} < \min\left\{\frac{1}{p\log(1/p)}, \frac{1}{(1-p)\log(1/(1-p))}\right\}. \tag{A.3}$$

PROOF. Let us first consider the case $0 < p < \frac{1}{2}$. Then it is enough to show that

$$\frac{\Lambda(p,1-p)}{p(1-p)} p \log \frac{1}{p} = \frac{(1-2p)\log(1/p)}{(1-p)\log((1-p)/p)} \overset{!}{<} 1. \tag{A.4}$$

This follows easily from the following lower bound on the denominator

$$\begin{aligned}(1-p)\log\frac{1-p}{p} &= (1-2p)\log\frac{1}{p} + p\log\frac{1}{p} - (1-p)\log\frac{1}{1-p}\\ &> (1-2p)\log\frac{1}{p},\end{aligned}$$

since $p\log\frac{1}{p} - (1-p)\log\frac{1}{1-p} > 0$ for $0 < p < \frac{1}{2}$. The case $\frac{1}{2} < p < 1$ follows by symmetry under the variable change $p \mapsto 1-p$. It remains to check the case $p = \frac{1}{2}$. The left-hand side of (A.4) evaluates for $p = \frac{1}{2}$ to

$$\lim_{p\to 1/2} \frac{\Lambda(p,1-p)}{p(1-p)} p \log\frac{1}{p} = \log 2 < 1. \qquad \square$$

The logarithmic mean also occurs in the following optimization problem, which appears in the proof of the optimality of the Eyring–Kramers formula for the logarithmic Sobolev constant in one dimension (cf. Section 2.4).

LEMMA A.3. *For $p \in (0,1)$ and $t \in (0,1)$, we define the function $h_p(t)$ according to*

$$h_p(t) = \frac{(\sqrt{t/p} - \sqrt{(1-t)/(1-p)})^2}{t\log(t/p) + (1-t)\log((1-t)/(1-p))}. \tag{A.5}$$

Then it holds

$$\min_{t\in(0,1)} h_p(t) = \frac{\Lambda(p,1-p)}{p(1-p)}. \tag{A.6}$$

The minimum in (A.6) is attained for $t = 1-p$.

PROOF. Let us introduce the function $f_p : (0,1) \to \mathbb{R}$ and $g_p : (0,1) \to \mathbb{R}$ given by the nominator and denominator of h_p in (A.5), namely

$$f_p(t) := \left(\sqrt{\frac{t}{p}} - \sqrt{\frac{1-t}{1-p}}\right)^2 \quad \text{and} \quad g_p(t) := t\log\frac{t}{p} + (1-t)\log\frac{1-t}{1-p}.$$

It is easy to verify, that the following relations for the derivatives hold true:

$$f_p'(t) = \left(\sqrt{\frac{t}{p}} - \sqrt{\frac{1-t}{1-p}}\right)\left(\frac{1}{\sqrt{tp}} + \frac{1}{\sqrt{(1-p)(1-t)}}\right),$$

$$g_p'(t) = \log\frac{t}{p} - \log\frac{1-t}{1-p}, \tag{A.7}$$

$$f_p''(t) = \sqrt{\frac{(1-t)t}{(1-p)p}}\frac{1}{2(1-t)^2t^2} > 0, \qquad g_p''(t) = \frac{1}{(1-t)t} > 0.$$

Hence, both functions f_p an g_p are strictly convex and have a unique minimum for $t = p$, where they are both zero. The derivative of the quotient of f_p and g_p has the form

$$h_p'(t) := \left(\frac{f_p(t)}{g_p(t)}\right)' = \frac{1}{g_p^2(t)}\big(f_p'(t)g_p(t) - f_p(t)g_p'(t)\big). \tag{A.8}$$

The representation (A.7) for g_p' leads to

$$h_p'(t)g_p^2(t) = \big(tf_p'(t) - f_p(t)\big)\log\frac{t}{p} + \big((1-t)f_p'(t) + f_p(t)\big)\log\frac{1-t}{1-p}. \tag{A.9}$$

Now, we can use (A.7) for f_p' to find

$$\begin{aligned} &tf_p'(t) - f_p(t) \\ &\quad= \left(\sqrt{\frac{t}{p}} - \sqrt{\frac{1-t}{1-p}}\right)\left(\sqrt{\frac{t}{p}} + \frac{t}{\sqrt{(1-p)(1-t)}} - \sqrt{\frac{t}{p}} + \sqrt{\frac{1-t}{1-p}}\right) \\ &\quad= \frac{1}{\sqrt{(1-p)(1-t)}}\left(\sqrt{\frac{t}{p}} - \sqrt{\frac{1-t}{1-p}}\right) \end{aligned} \tag{A.10}$$

and likewise

$$(1-t)f_p'(t) + f_p(t) = \frac{1}{\sqrt{tp}}\left(\sqrt{\frac{t}{p}} - \sqrt{\frac{1-t}{1-p}}\right). \tag{A.11}$$

Using (A.10) and (A.11) in (A.9) leads by (A.8) to

$$h_p'(t) = \frac{1}{g_p^2(t)}\underbrace{\left(\sqrt{\frac{t}{p}} - \sqrt{\frac{1-t}{1-p}}\right)}_{=:v_p(t)}\underbrace{\left(\frac{\log(t/p)}{\sqrt{(1-p)(1-t)}} + \frac{\log((1-t)/(1-p))}{\sqrt{tp}}\right)}_{=:w_p(t)}.$$

Since $g_p(p) = g_p'(p) = 0$ and $g_p''(p) > 0$, the function $\frac{1}{g_p^2(t)}$ has a pole of order 4 in $t = p$. Moreover, the function $v_p(t)$ has a simple zero in $t = p$. We have to do some more investigations for the function $w_p(t)$. First, we observe that $w_p(t)$ can be rewritten as

$$w_p(t) = \underbrace{\frac{t-p}{\sqrt{(1-t)t(1-p)p}}}_{=:\hat{w}_p(t)} \\ \times \underbrace{\left(\frac{\sqrt{tp}\log(t/p)}{(t-p)} - \frac{\sqrt{(1-t)(1-p)}\log((1-t)/(1-p))}{(p-t)}\right)}_{:=\tilde{w}_p(t)}.$$

The function $\tilde{w}_p(t)$ can be expressed in terms of the logarithmic mean

$$\tilde{w}_p(t) = \frac{\sqrt{tp}}{\Lambda(t,p)} - \frac{\sqrt{(1-t)(1-p)}}{\Lambda(1-t,1-p)} \tag{A.12}$$

and is measuring the defect in the geometric-logarithmic mean inequality (A.2). Let us switch to exponential variables and set

$$x(t) := \log\sqrt{\frac{t}{p}} \quad \text{and} \quad y(t) := \log\sqrt{\frac{1-t}{1-p}}.$$

Note that either $x(t) \le 0 \le y(t)$ for $t \le p$ or $y(t) \le 0 \le x(t)$ for $t \ge 0$ with equality only for $t = p$. Therewith, (A.12) can be rewritten as

$$\tilde{w}_p(t) = \frac{x(t)}{\sinh(x(t))} - \frac{y(t)}{\sinh(y(t))}.$$

By making use of the fact, that the function $x \mapsto \frac{x}{\sinh x}$ is symmetric, strictly monotone decreasing in $|x|$ and has a unique maximum in 1, we can conclude that

$$\tilde{w}_p(t) = 0 \quad \text{if and only if} \quad x(t) = -y(t).$$

The solutions to the equation $x(t) = -y(t)$ are given for $t \in \{p, 1-p\}$. Let us first consider the case $t = p$, then $x(t) = y(t) = 0$ and $w_p(p)$ is a zero of order 2, since the function $x \mapsto \frac{x}{\sinh(x)}$ is strictly concave for $t = 0$. Now, we can go back to $h_p'(t)$ and argue with the representation

$$\lim_{t\to p} h_p'(t) = \lim_{t\to p} \frac{v_p(t)\hat{w}_p(t)\tilde{w}_p(t)}{g_p^2(t)} \overset{!}{\neq} 0.$$

This is a consequence of counting the zeros for $t = p$ in the nominator and denominator according to their order; for the denominator $g_p^2(p)$ is a zero of order 4. For the nominator, we have $v_p(p)$ is a zero of order 1, $\hat{w}_p(p)$ is a zero of order 1 and $\tilde{w}_p(p)$ is a zero of order 2, which leads in total again to a zero of order 4 exactly compensating the zero of the denominator.

The other case is $t = 1 - p$. Let us evaluate $h_p(1-p)$, which is given by

$$h_p(1-p) = \frac{(p-(1-p))^2/(p(1-p))}{(1-p)\log((1-p)/p) + p\log(p/(1-p))}$$
$$= \frac{1}{p(1-p)} \frac{(p-(1-p))^2}{(p-(1-p))\log(p/(1-p))} = \frac{\Lambda(p, 1-p)}{p(1-p)}.$$

Since $t = 1 - p$ is the only critical point of $h_p(t)$ inside (0, 1), it remains to check whether the boundary values are larger than $h_p(1-p)$. They are given by

$$\lim_{t\to 0} h_p(t) = \frac{1}{(1-p)\log(1/(1-p))} \quad \text{and} \quad \lim_{t\to 1} h_p(t) = \frac{1}{p\log(1/p)}.$$

We observe that the demanded inequality to be in a global minimum

$$h_p(1-p) = \frac{\Lambda(p, 1-p)}{p(1-p)} \overset{!}{<} \min\left\{\frac{1}{p\log(1/p)}, \frac{1}{(1-p)\log(1/(1-p))}\right\}$$

is just (A.3) of Lemma A.2. □

APPENDIX B: INTEGRATION BY PARTS ON BASINS OF ATTRACTION

The goal of this Appendix is to proof the integration by parts formula, which is an ingredient of the Lyapunov approach in Section 3.

THEOREM B.1 (Integration by parts). *Let $H \in C^3(\mathbb{R}^n, \mathbb{R})$ be a Morse function (cf. Definition 1.3) with compact sublevel sets and let Ω be the basin of attraction associated to a local minimum of H (cf. Definition B.7), then it holds*

$$\forall f, g \in H^1(\mu|_\Omega) \qquad \text{with } \nabla g \parallel \nabla H \text{ on } \partial\Omega: \int_\Omega f(-Lg)\,\mathrm{d}\mu = \varepsilon \int_\Omega \langle \nabla f, \nabla g\rangle\,\mathrm{d}\mu,$$

where $\nabla g \parallel \nabla H$ means $|\nabla g(x) \cdot \nabla H| = |\nabla g(x)||\nabla H(x)|$ for $\mathcal{H}^{n-1}$-a.e. $x \in \partial\Omega$.

REMARK B.2. The property of H possessing compact sublevel sets is called proper. This gives enough compactness, that is, the Palais–Smale condition [29], Definition 6.2.1, to apply several results from Morse theory and dynamical systems. Moreover, if H satisfies Assumption (1.4), then H is proper.

B.1. Properties of gradient flows.

DEFINITION B.3 (Gradient flow). Let $\phi_t(x)$ be the trajectory associated to the negative gradient flow of H started in x, that is,

$$\partial_t \phi_t = -\nabla H(\phi_t) \quad \text{and} \quad \phi_0(x) = x \in \mathbb{R}^n.$$

LEMMA B.4 (Properties of gradient flow trajectories).

(i) *For each* x, *the trajectory* $t \mapsto \phi_t(x)$ *has a maximal interval of definition of the form* $(-\alpha_x, \infty)$ *for* $\alpha_x \in (-\infty, 0) \cup \{-\infty\}$.

(ii) *For each* x: $\lim_{t\to\infty} \phi_t(x) =: \phi_\infty(x) \in \mathcal{S}$.

(iii) *Stability on finite time intervals*, *that is*, *for any* $T > 0$ *holds if* $x_n \to x$ *also* $\phi_T(x_n) \to \phi_T(x)$.

PROOF. Since H is locally Lipschitz, the trajectory $\phi_t(x)$ has a maximal interval of definition $0 \in (\alpha_x, \beta_x) \cup \{\pm\infty\}$ by the Picard–Lindelöf theorem. Moreover, since

$$\partial_t H(x_t) = -\left|\nabla H(x_t)\right|^2 = -|\dot{x}_t|^2 \le 0 \tag{B.1}$$

the trajectory $\{\phi_t(x)\}_{t\ge 0}$ is confined to the sublevel set $\{y : H(y) \le H(x)\}$, which is compact, since H is proper. On this sublevel set, H is globally Lipschitz and the limit $\lim_{t\to\infty} \phi_t(x) =: \phi_\infty(x)$ exists proving (i). In addition, this implies

$$\int_0^\infty \left|\nabla H(\phi_t)\right|^2 \mathrm{d}t \overset{(B.1)}{=} -\int_0^\infty \partial_t H(\phi_t)\,\mathrm{d}t = H(x) - H\big(\phi_\infty(x)\big) < \infty.$$

Therefore, it follows $\phi_\infty(x) \in \mathcal{S} := \{x \in \mathbb{R}^n : \nabla H(x) = 0\}$ is a critical point proving (ii). The stability follows from the estimate

$$\begin{aligned} \left|\phi_T(x_n) - \phi_T(x)\right| &= \left|x_n + \int_0^T \partial_t\phi_t(x_n)\,\mathrm{d}t - x - \int_0^T \partial_t\phi_t(x)\right| \\ &\le |x_n - x| + \int_0^T \left|\nabla H\big(\phi_t(x_n)\big) - \nabla H\big(\phi_t(x)\big)\right| \mathrm{d}t. \end{aligned} \tag{B.2}$$

All $\phi_t(x_n)$ are confined to a common compact set by properness of H and in particular ∇H is Lipschitz continuous in this compact set. This leads for some $K > 0$ and all $t \in (0, T)$ to the estimate

$$\left|\nabla H\big(\phi_t(x_n)\big) - \nabla H\big(\phi_t(x)\big)\right| \le K\left|\phi_t(x_n) - \phi_t(x)\right|.$$

Using this estimate in (B.2), we can apply the Gronwall inequality to obtain $|\phi_T(x_n) - \phi_T(x)| \le |x_n - x|(1 + e^{KT})$, which proves (iii). $\square$

We want to define a global flow w.r.t. ∇H. Since, ∇H can have superlinear growth and is in particular not globally Lipschitz continuous, we use the following reparameterized version for a global flow.

THEOREM B.5 (Global flow by reparameterization [34], Theorem 4.4). *A global flow of diffeomorphism* $\tilde{\phi}_t : \mathbb{R}^n \to \mathbb{R}^n$ *w.r.t.* H *is defined by*

$$\partial_t\tilde{\phi}_t(x) = F\big(\tilde{\phi}_t(x)\big) := -\frac{\nabla H(\tilde{\phi}_t(x))}{1 + |\nabla H(\tilde{\phi}_t(x))|} \quad \textit{and} \quad \tilde{\phi}_0(x) = x. \tag{B.3}$$

This flow is equivalent to the negative gradient flow of H upon a reparameterization of time. The vector field F is globally Lipschitz and bounded. It defines a global flow on $\mathbb{R}^n$, that is, $\tilde{\phi}_{t+s} = \tilde{\phi}_t \circ \tilde{\phi}_s$ for all $t, s \in \mathbb{R}$.

COROLLARY B.6. *Each point $x \in \mathbb{R}^n$ belongs to exactly one trajectory $t \to \phi_t(x)$.*

PROOF. We apply [29], Corollary 1.9.1, to the global flow $\tilde{\phi}_t$ and by Theorem B.5 translate the result back to ϕ_t. $\square$

B.2. The stable manifold.

DEFINITION B.7 (Stable manifold). To each critical point $s \in \mathcal{S}$, the *stable manifold* is defined by

$$W^{\mathrm{s}}(s) := \Big\{ x \in \mathbb{R}^n : \lim_{t\to\infty} \phi_t(x) = s \Big\}.$$

Moreover, we call the dimension $k \in \{0, \dots, n\}$ of the unstable subspace of $\nabla^2 H(s)$ the index of the saddle point s. If m is a local minimum of H, that is, a critical point of index 0, we call $W^{\mathrm{s}}(m)$ the *basin of attraction for m*.

Lemma B.4(ii) and Corollary B.6 ensure the stable manifold to be well defined and immediately provide the following.

COROLLARY B.8 (Partition of state space). *Let $\mathcal{S}$ be all critical points of H, then $\mathbb{R}^n$ is the disjoint union of all stable manifolds denoted by*

$$\mathbb{R}^n := \dot{\bigcup_{s\in\mathcal{S}}} W^{\mathrm{s}}(s).$$

THEOREM B.9 (Local stable manifold theorem [29], Theorem 6.3.1). *Let $s \in \mathcal{S}$ and $\mathcal{E}^{\mathrm{s}}(s)$ be the stable subspace of $\nabla^2 H(s)$, that is, $\nabla^2 H(s)|_{\mathcal{E}^{\mathrm{s}}}$ has a positive spectrum. Then there exists a neighborhoods $U, \tilde{U}$ of s, such that $W^{\mathrm{s}}(s) \cap U$ is a C^1-graph over $(s + \mathcal{E}^{\mathrm{s}}(s)) \cap \tilde{U}$. Especially, the dimension of $W^{\mathrm{s}}(s) \cap U$ and $\mathcal{E}^{\mathrm{s}}(s)$ are equal to $n - k$, where k is the index of s.*

The local result can be extended by the reparameterized flow to the global manifold theorem.

THEOREM B.10 (Global stable manifold theorem [29], Corollary 6.3.1). *The stable manifolds $W^{\mathrm{s}}(s)$ for $s \in \mathcal{S}$ of the flow associated to F (B.3) are immersed C^1-manifolds of dimension $n - k$, where k is the index of s.*

In the present case of a gradient flow, the result can be strengthened to the following.

THEOREM B.11 (Global stable manifold theorem for gradient systems [29], Corollary 6.4.1). *The stable manifolds $W^{\mathrm{s}}(s)$ for $s \in \mathcal{S}$ of the gradient flow associated to H are embedded C^1-submanifolds of dimension $n-k$, where k is the index of s.*

PROOF. We have to modify the proof of [29], Corollary 6.3.1, since ∇H can have superlinear growth. Instead, considering the gradient flow w.r.t. H, we consider the equivalent flow $\tilde{\phi}_t$ of Theorem B.5. We have to observe two additional facts, which we postpone to the end of the proof.

(a) The flow has no nonconstant homoclinic orbits, that is, nonconstant orbits with $\lim_{t\to-\infty}\tilde{\phi}_t(x) = \lim_{t\to\infty}\tilde{\phi}_t(x)$ (cp. [29], Lemma 6.4.3).

(b) For each x, holds $|\nabla H(\tilde{\phi}_t(x))| \to 0$ as $t \to \infty$ and either $|\nabla H(\tilde{\phi}_t(x))| \to 0$ or $H(\tilde{\phi}_t(x)) \to \infty$ as $t \to -\infty$ (cp. [29], Lemma 6.4.4).

This allows us to complete the proof by first applying Theorem B.10 to $F(x) = -\nabla H(x)/(1+|\nabla H(x)|)$. Every point $x \in \mathbb{R}^n$ is contained in a unique trajectory $\phi_t(x)$ by Corollary B.6. However, a trajectory is typical not compact. In (b) we show that limit points in $\mathbb{R}^n$ are critical points of H. The local situation around critical points is given by the local stable manifold theorem B.9, which provides a local chart around the critical point. Selfintersection of trajectory is excluded by the observation in (a). Hence, the immersion of Theorem B.10 is an embedding.

We still have to show (a) and (b):

Ad (a): The energy also decreases w.r.t. to the reparameterized flow

$$\partial_t H(\tilde{\phi}_t(x)) = -\nabla H \cdot \partial_t \tilde{\phi}_t(x) = -\frac{|\nabla H(\tilde{\phi}_t(x))|^2}{1+|\nabla H(\tilde{\phi}_t(x))|} \le 0. \tag{B.4}$$

Hence, for a trajectory either holds $|\nabla H| = 0$ or $|\nabla H| > 0$ for all t, which gives (a).

Ad (b): Integrating the identity (B.4), we obtain for $t_2 > t_1$

$$H(\tilde{\phi}_{t_1}(x)) - H(\tilde{\phi}_{t_2}(x)) = \int_{t_1}^{t_2} \frac{|\nabla H(\tilde{\phi}_t(x))|^2}{1+|\nabla H(\tilde{\phi}_t(x))|}\,\mathrm{d}t \ge \int_{t_1}^{t_2} |\nabla H(\tilde{\phi}_t(x))|\,\mathrm{d}t.$$

Since H is bounded from below, we get that $H(\tilde{\phi}_\infty(x)) > -\infty$. Hence,

$$H(\tilde{\phi}_{t_1}(x)) - H(\tilde{\phi}_{t_2}(x)) < \infty$$

for all $t_2 > t_1$ and we immediately deduce from (B.4) that $\tilde{\phi}_\infty(x) \in \mathcal{S}$ showing the first part of (b). If $H(\tilde{\phi}_{-\infty}(x)) < \infty$, then by the same argument $\phi_{-\infty}(x) \in \mathcal{S}$. Hence, we have shown the dichotomy (b). □

B.3. The boundary of the basin of attraction.

LEMMA B.12. *The set $\{W^s(m)\}_{m\in\mathcal{M}}$ is a partition of $\mathbb{R}^n$ upon Lebesgue null sets, denoted by*

$$\mathbb{R}^n = \biguplus_{m\in\mathcal{M}} W^s(m). \tag{B.5}$$

Moreover, it holds

$$\bigcup_{m\in\mathcal{M}} \partial W^s(m) = \dot{\bigcup_{y\in\mathcal{S}\setminus\mathcal{M}}} W^s(y). \tag{B.6}$$

PROOF. For (B.5), we observe that $W^s(y)$ for $y \in \mathcal{S}\setminus\mathcal{M}$ are Lebesgue null sets, since they are $(n-k)$-dimensional C^1-submanifolds for $1 \le k \le n$ (cf. Theorem B.11).

Theorem B.11 proves in particular, that for each $m \in \mathcal{M}$ the embedded submanifold $W^s(m)$ is open in $\mathbb{R}^n$, hence $\partial W^s(m) \cap W^s(m) = \varnothing$. Therewith, the second statement (B.6) follows from Corollary B.8. $\square$

THEOREM B.13 (The boundary of the basin of attraction). *Let $m \in \mathcal{M}$ be a local minimum of H. There exists a set $S_m \subset \mathcal{S}\setminus\mathcal{M}$ of k-saddles with $k \ge 1$ such that*

$$\partial W^s(m) = \dot{\bigcup_{y\in S_m}} W^s(y).$$

PROOF. We define a critical point $y \in \mathcal{S}$ to be in S_m if for each open neighborhood $U(y)$ holds $U(y) \cap W^s(m) \ne \varnothing$. From B.6 follows that $y \in S_m$ cannot be another local minimum and hence $S_m \subset \mathcal{S}\setminus\mathcal{M}$. Now, we take $x_n \to x \in \partial W^s(m)$. From (B.6) follows that $x \in W^s(y)$ for some $y \in \mathcal{S}\setminus\mathcal{M}$. We have to prove that $y \in S_m$. There exists an open neighborhood $U(x)$ such that $x_n \in U(x)$ for $n > N$. Then for any open neighborhood $U(y)$ of y exists $T > 0$ such that $\phi_T(x) \in U(y)$. By existence of the flow ϕ_t for positive time, it follows hat $\phi_T(U(x)) \cap U(y) =: U(\phi_T(x))$ is an open neighborhood of $\phi_T(x)$. By stability of the flow on finite time intervals [cf. Lemma B.4(iii)], it follows $\phi_T(x_n) \to \phi_T(x)$, hence $\phi_T(x_n) \in U(\phi_T(x))$ for n large enough, which shows that $W^s(m) \cap U(\phi_T(x)) \ne \varnothing$ and finally $y \in S_m$. $\square$

PROOF OF THEOREM B.1. Let m be a local minimum of H. By Theorem B.13 the boundary of $W^s(m)$ is the union of C^1-submanifolds. The relevant submanifolds for integration, are the $(n-1)$-dimensional ones. By Theorem B.11, these $(n-1)$-dimensional submanifolds correspond to stable manifolds of saddle points of index 1, denoted by $\mathcal{S}_1$. Hence, for $\mathcal{H}^{n-1}$-a.e. $x \in \partial W^s(m)$ exists a 1-saddle

$y \in S_m \cap \mathcal{S}_1$ such that $x \in W^{\mathrm{s}}(y)$. Therefore, the normal on $W^{\mathrm{s}}(m)$ exists $\mathcal{H}^{n-1}$-a.e., which gives enough regularity to integrate for $f, g \in H^1(\mu|_\Omega)$ by parts

$$\int_\Omega f(-Lg)\,\mathrm{d}\mu = \varepsilon \int_{W^{\mathrm{s}}(m)} \langle \nabla f, \nabla g\rangle \,\mathrm{d}\mu - \varepsilon \sum_{y \in S_m \cap \mathcal{S}_1} \int_{W^{\mathrm{s}}(y)} f \nabla g \cdot n \mathcal{H}^{n-1}(\mathrm{d}\mu).$$

By the assumption $\nabla g \parallel \nabla H$, it is enough to show that $\nabla H(x) \cdot n = 0$ for $\mathcal{H}^{n-1}$-a.e. $x \in \partial W^{\mathrm{s}}(m)$. This is proven by contradiction for $x \in \partial W^{\mathrm{s}}(m)$. Assume that $x \notin \mathcal{S}$, that is, $\nabla H(x) \neq 0$ and $\nabla H(x) \cdot n \neq 0$. Then for some $\varepsilon > 0$ there exists $t^* \in (-\varepsilon, \varepsilon)$ such that $\phi_{t^*}(x) \in W^{\mathrm{s}}(m)$. By definition of $W^{\mathrm{s}}(m)$ and global existence of the trajectory $\{\phi_t(x)\}_{t \geq t^*}$ from Lemma B.4(ii) follows $x \in W^{\mathrm{s}}(m)$, which contradicts (B.6) and Corollary B.8. □

APPENDIX C: AUXILIARY RESULTS FROM SECTION 4

C.1. Partial Gaussian integrals. This section is devoted to proof the representation for partial or incomplete Gaussian integrals. Lemma (C.1) is an ingredient to evaluate the weighted transport cost in Section 4.3.

LEMMA C.1 (Partial Gaussian integral). *Let $\Sigma^{-1} \in \mathbb{R}^{n\times n}_{\mathrm{sym},+}$ be a symmetric positive definite matrix and let $\eta \in S^{n-1}$ be a unit vector. Therewith, $\{r\eta + z^*\}_{r\in\mathbb{R}}$ is for $z^* \in \mathbb{R}^n$ with $\langle \eta, z^*\rangle = 0$ an affine subspace of $\mathbb{R}^n$. The integral of a centered Gaussian w.r.t. to this subspace evaluates to*

$$\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\Sigma^{-1}[r\eta + z^*]\right)\mathrm{d}r = \frac{\sqrt{2\pi}}{\sqrt{\Sigma^{-1}[\eta]}} \exp\left(-\tilde{\Sigma}^{-1}[z^*]\right),$$

$$\textit{with } \tilde{\Sigma}^{-1} = \Sigma^{-1} - \frac{\Sigma^{-1}\eta \otimes \Sigma^{-1}\eta}{\Sigma^{-1}[\eta]}.$$

PROOF. To evaluate this integral on an one-dimensional subspace of $\mathbb{R}^n$, we have to expand the quadratic form $\Sigma^{-1}[r\eta + z^*]$ and arrive at the relation

$$\begin{aligned}
&\int_{\mathbb{R}} \exp\left(-\frac{1}{2}\Sigma^{-1}[r\eta + z^*]\right)\mathrm{d}r \\
&\quad = \exp\left(-\frac{1}{2}\Sigma^{-1}[z^*]\right)\int_{\mathbb{R}} \exp\left(-\frac{r^2}{2}\Sigma^{-1}[\eta] + r\langle \eta, \Sigma^{-1}z^*\rangle\right)\mathrm{d}r \\
&\quad = \exp\left(-\frac{1}{2}\Sigma^{-1}[z^*]\right)\frac{\sqrt{2\pi}}{\sqrt{\Sigma^{-1}[\eta]}}\exp\left(\frac{\langle \eta, \Sigma^{-1}z^*\rangle^2}{2\Sigma^{-1}[\eta]}\right) \\
&\quad = \frac{\sqrt{2\pi}}{\sqrt{\Sigma^{-1}[\eta]}}\exp\left(-\frac{1}{2}\left(\Sigma^{-1} - \frac{\Sigma^{-1}\eta \otimes \Sigma^{-1}\eta}{\Sigma^{-1}[\eta]}\right)[z^*]\right),
\end{aligned}$$

which concludes the hypothesis. □

C.2. Subdeterminants, adjugates and inverses. Let $A \in \mathbb{R}^{n\times n}_{\mathrm{sym},+}$, then define for $\eta \in S^{n-1}$ the matrix

$$\tilde{A} := A - \frac{A\eta \otimes A\eta}{A[\eta]}. \tag{C.1}$$

The matrix $\tilde{A}$ has at least rank $n-1$, since we subtracted from the positive definite matrix A a rank-1 matrix. Further, from the representation, it is immediate that $\tilde{A}$ has rank $n-1$ if and only if η is an eigenvector of A. In this case, $\ker A = \operatorname{span} \eta$. It immediately follows $\tilde{A} > 0$ on $V := \operatorname{span}\{\eta\}^{\perp}$, which is the $(n-1)$-dimensional subspace perpendicular to η. Then for a matrix $A \in \mathbb{R}^{n\times n}_{\mathrm{sym},+}$ we want to calculate the determinant of A restricted to this subspace V. This determinant is obtained by first choosing $Q \in SO^n$ such that $Q(\{0\} \times \mathbb{R}^{n-1}) = V$ and then evaluating the determinant of the minor consisting of the $(n-1) \times (n-1)$ lower right submatrix of $Q^\top A Q$ denoted by $\det_{1,1}(Q^\top A Q)$. Hence, we have

$$\det_{1,1}(Q^\top A Q) \qquad \text{with } Q \in SO(n) : Q^\top \eta = e^1 = (1, 0, \dots, 0)^\top.$$

Since $V = \operatorname{span}\{\eta\}^{\perp}$, it follows that the first column of Q is given by η and we can decompose $Q^\top A Q$ into

$$Q^\top A Q = \begin{pmatrix} A[\eta] & \widehat{Q^\top A\eta} \\ \widehat{Q^\top A\eta}^\top & \widehat{Q^\top A Q} \end{pmatrix},$$

where for a matrix M, $\widehat{M}$ is the lower right $(n-1) \times (n-1)$ submatrix of M and for a vector v, $\widehat{v}$ the $(n-1)$ lower subvector of v. Therewith, we find a similarity transformation which applied to $Q^\top A Q$ results in

$$\begin{aligned}
\det A = \det Q^\top A Q &= \det\left(\begin{pmatrix} A[\eta] & \widehat{Q^\top A\eta} \\ \widehat{Q^\top A\eta}^\top & \widehat{Q^\top A Q} \end{pmatrix}\begin{pmatrix} 1 & -\frac{\widehat{Q^\top A\eta}}{A[\eta]} \\ 0 & \mathrm{Id}_{n-1} \end{pmatrix}\right) \\
&= \det\begin{pmatrix} A[\eta] & 0 \\ \widehat{Q^\top A\eta}^\top & \widehat{Q^\top A Q} - \frac{\widehat{A\eta} \otimes \widehat{A\eta}}{A[\eta]} \end{pmatrix} \\
&= A[\eta] \det_{1,1}\left(Q^\top A Q - \frac{Q^\top A\eta \otimes Q^\top A\eta}{A[\eta]}\right).
\end{aligned}$$

The determinant of the minor is given by

$$\det_{1,1}\left(Q^\top A Q - \frac{Q^\top A\eta \otimes Q^\top A\eta}{A[\eta]}\right) = \det_{1,1}\left(Q^\top \left(A - \frac{A\eta \otimes A\eta}{A[\eta]}\right) Q\right).$$

Hence, by the definition (C.1) of $\tilde{A}$ and the subdeterminant, we found the identity

$$\det A = A[\eta] \det_{1,1}(Q^\top \tilde{A} Q). \tag{C.2}$$

C.3. A matrix optimization.

LEMMA C.2. *Let $B \in \mathbb{R}^{n\times n}_{\mathrm{sym},+}$, then it holds*

$$\inf_{A\in\mathbb{R}^{n\times n}_{\mathrm{sym},+}} \left\{ \frac{\det A}{\sqrt{\det(2A-B)}} : 2A > B \right\} = \sqrt{\det B}$$

and for the optimal A holds $A = B$.

PROOF. We note that

$$\frac{\det A}{\sqrt{\det(2A-B)}} = \frac{1}{\sqrt{\det(A^{-1})\det(2\,\mathrm{Id} - A^{-1/2}BA^{-1/2})}}.$$

Therewith, it is enough to maximize the radical of the root. Therefore, we substitute $A^{-1/2} = CB^{-1/2}$ with $C > 0$ not necessarily symmetric and observe that $A^{-1/2} = B^{-1/2}C^{\top}$. We obtain

$$\det(A^{-1})\det(2\,\mathrm{Id} - A^{-1/2}BA^{-1/2}) = \det(B^{-1})\det(CC^{\top})\det(2\,\mathrm{Id} - CC^{\top}).$$

Note that $CC^{\top} \in \mathbb{R}^{n\times n}_{\mathrm{sym},+}$ and it is enough to calculate

$$\sup_{\tilde{C}\in\mathbb{R}^{n\times n}_{\mathrm{sym},+}} \{\det(\tilde{C})\det(2\,\mathrm{Id} - \tilde{C}) : \tilde{C} < 2\,\mathrm{Id}\}.$$

From the constraint $0 < \tilde{C} < 2\,\mathrm{Id}$, we can write $\tilde{C} = \mathrm{Id} + D$, where D is symmetric and satisfies $-\mathrm{Id} < D < \mathrm{Id}$ in the sense of quadratic forms. From here, we finally observe

$$\det(\tilde{C})\det(2\,\mathrm{Id} - \tilde{C}) = \det(\mathrm{Id} + D)\det(\mathrm{Id} - D) = \det(\mathrm{Id} - D^2).$$

Since $D^2 \geq 0$, we find the optimal $\tilde{C}$ given by Id, which yields that $A = B$. $\square$

C.4. Jacobi matrices. For a smooth function $f : \mathbb{R}^n \to \mathbb{R}^n$ denotes $Df(x)$ the *Jacobi matrix* of the partial derivatives of f in $x \in \mathbb{R}^n$ given by

$$Df(x) := \left(\frac{\mathrm{d}f_i}{\mathrm{d}x_j}(x) \right)_{i,j=1}^{n}.$$

LEMMA C.3. *Let $A, B \in \mathbb{R}^{n\times n}$ and $f : \mathbb{R}^n \to \mathbb{R}^n$ smooth, then it holds*

$$\nabla |Ax + f(Bx)| = (A + Df(x)B)^{\top} \frac{Ax + f(Bx)}{|Ax + f(Bx)|}, \tag{C.3}$$

$$D\frac{f(x)}{|f(x)|} = \frac{1}{|f(x)|}\left(\mathrm{Id} - \frac{f(x)}{|f(x)|} \otimes \frac{f(x)}{|f(x)|} \right) Df(x). \tag{C.4}$$

PROOF. Let us first check the relation (C.3) and calculate the partial derivative

$$\frac{\mathrm{d}|Ax+f(Bx)|}{\mathrm{d}x_i}=\frac{1}{2|Ax+f(Bx)|}\sum_j\frac{\mathrm{d}}{\mathrm{d}x_i}\biggl(\sum_k A_{jk}x_k+f_j(Bx)\biggr)^2. \tag{C.5}$$

The inner derivative of (C.5) evaluates to

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}x_i}\biggl(\sum_k A_{jk}x_k+f_j(Bx)\biggr)^2\\ &\qquad=2\biggl(\sum_k A_{jk}x_k+f_j(Bx)\biggr)\biggl(A_{ji}+\frac{\mathrm{d}f_j(Bx)}{\mathrm{d}x_i}\biggr).\end{aligned} \tag{C.6}$$

The derivative of $f_j(Bx)$ becomes

$$\begin{aligned}\frac{\mathrm{d}f_j(Bx)}{\mathrm{d}x_i}&=\frac{\mathrm{d}f_j(\sum_k B_{1k}x_k,\dots,\sum_k B_{nk}x_k)}{\mathrm{d}x_i}\\ &=\sum_{k=1}^n\partial_k f_j(Bx)B_{ki}=\bigl(Df(Bx)B\bigr)_{ji}.\end{aligned} \tag{C.7}$$

Hence, a combination of (C.5), (C.6) and (C.7) leads to

$$\begin{aligned}\frac{\mathrm{d}|Ax+f(Bx)|}{\mathrm{d}x_i}&=\frac{1}{|Ax+f(Bx)|}\sum_j\bigl((Ax)_j+f_j(Bx)\bigr)\bigl(A_{ji}\bigl(Df(Bx)B\bigr)_{ji}\bigr)\\ &=\sum_j\bigl(A+Df(Bx)B\bigr)_{ij}^\top\frac{(Ax+f(Bx))_j}{|Ax+f(Bx)|},\end{aligned}$$

which shows (C.3). For the equation (C.4), let us first consider the Jacobian of the function $F(x)=\frac{x}{|x|}$, which is given by

$$DF(x)=\frac{1}{|x|}\biggl(\mathrm{Id}-\frac{x}{|x|}\otimes\frac{x}{|x|}\biggr).$$

Then, by the chain rule, we observe that

$$D\frac{f(x)}{|f(x)|}=D(F\circ f)(x)=DF\bigl(f(x)\bigr)Df(x),$$

which is just (C.4). □

Acknowledgments. This work is part of the Ph.D. thesis of the second author, written under the supervision of Stephan Luckhaus at the University of Leipzig.

The authors are greatly indebted to Felix Otto for drawing their attention to the low temperature regime and for several helpful discussions.

Moreover, they want to thank the anonymous referees for pointing out the role of the Neumann boundary conditions in the Lyapunov argument.

They wish to thank the Max-Planck-Institute for Mathematics in the Sciences in Leipzig where the paper was written under great working conditions.

REFERENCES

[1] ARRHENIUS, S. (1889). On the reaction velocity of the inversion of cane sugar by acids. *Zeitschrift für Physikalische Chemie* **4** 226ff.

[2] BAKRY, D., BARTHE, F., CATTIAUX, P. and GUILLIN, A. (2008). A simple proof of the Poincaré inequality for a large class of probability measures including the log-concave case. *Electron. Commun. Probab.* **13** 60–66. MR2386063

[3] BAKRY, D., CATTIAUX, P. and GUILLIN, A. (2008). Rate of convergence for ergodic continuous Markov processes: Lyapunov versus Poincaré. *J. Funct. Anal.* **254** 727–759. MR2381160

[4] BAKRY, D. and ÉMERY, M. (1985). Diffusions hypercontractives. In *Séminaire de Probabilités*, *XIX*, 1983/84. *Lecture Notes in Math.* **1123** 177–206. Springer, Berlin. MR0889476

[5] BENAMOU, J.-D. and BRENIER, Y. (2000). A computational fluid mechanics solution to the Monge–Kantorovich mass transfer problem. *Numer. Math.* **84** 375–393. MR1738163

[6] BERGLUND, N. (2011). Kramers' law: Validity, derivations and generalisations. Available at arXiv:1106.5799.

[7] BHATIA, R. (2008). The logarithmic mean. *Resonance* **13** 583–594.

[8] BOBKOV, S. G. and GÖTZE, F. (1999). Exponential integrability and transportation cost related to logarithmic Sobolev inequalities. *J. Funct. Anal.* **163** 1–28. MR1682772

[9] BOVIER, A., ECKHOFF, M., GAYRARD, V. and KLEIN, M. (2004). Metastability in reversible diffusion processes. I. Sharp asymptotics for capacities and exit times. *J. Eur. Math. Soc.* (*JEMS*) **6** 399–424. MR2094397

[10] BOVIER, A., GAYRARD, V. and KLEIN, M. (2005). Metastability in reversible diffusion processes. II. Precise asymptotics for small eigenvalues. *J. Eur. Math. Soc.* (*JEMS*) **7** 69–99. MR2120991

[11] CARLSON, B. C. (1972). The logarithmic mean. *Amer. Math. Monthly* **79** 615–618. MR0302842

[12] CATTIAUX, P., GUILLIN, A., WANG, F.-Y. and WU, L. (2009). Lyapunov conditions for super Poincaré inequalities. *J. Funct. Anal.* **256** 1821–1841. MR2498560

[13] CATTIAUX, P., GUILLIN, A. and WU, L.-M. (2010). A note on Talagrand's transportation inequality and logarithmic Sobolev inequality. *Probab. Theory Related Fields* **148** 285–304. MR2653230

[14] CHAFAÏ, D. and MALRIEU, F. (2010). On fine properties of mixtures with respect to concentration of measure and Sobolev type inequalities. *Ann. Inst. Henri Poincaré Probab. Stat.* **46** 72–96. MR2641771

[15] DIACONIS, P. and SALOFF-COSTE, L. (1996). Logarithmic Sobolev inequalities for finite Markov chains. *Ann. Appl. Probab.* **6** 695–750. MR1410112

[16] DOLBEAULT, J., NAZARET, B. and SAVARÉ, G. (2009). A new class of transport distances between measures. *Calc. Var. Partial Differential Equations* **34** 193–231. MR2448650

[17] DONSKER, M. D. and VARADHAN, S. R. S. (1976). On the principal eigenvalue of second-order elliptic differential operators. *Comm. Pure Appl. Math.* **29** 595–621. MR0425380

[18] EYRING, H. (1935). The activated complex in chemical reactions. *J. Chem. Phys.* **3** 107.

[19] FREIDLIN, M. I. and WENTZELL, A. D. (1998). *Random Perturbations of Dynamical Systems*, 2nd ed. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **260**. Springer, New York. Translated from the 1979 Russian original by Joseph Szücs. MR1652127

[20] GONZALEZ, O., MADDOCKS, J. H., SCHURICHT, F. and VON DER MOSEL, H. (2002). Global curvature and self-contact of nonlinearly elastic curves and rods. *Calc. Var. Partial Differential Equations* **14** 29–68. MR1883599

[21] GROSS, L. (1975). Logarithmic Sobolev inequalities. *Amer. J. Math.* **97** 1061–1083. MR0420249

[22] GRUNEWALD, N., OTTO, F., VILLANI, C. and WESTDICKENBERG, M. G. (2009). A two-scale approach to logarithmic Sobolev inequalities and the hydrodynamic limit. *Ann. Inst. Henri Poincaré Probab. Stat.* **45** 302–351. MR2521405

[23] HELFFER, B., KLEIN, M. and NIER, F. (2004). Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach. *Mat. Contemp.* **26** 41–85. MR2111815

[24] HELFFER, B. and NIER, F. (2005). *Hypoelliptic Estimates and Spectral Theory for Fokker–Planck Operators and Witten Laplacians. Lecture Notes in Math.* **1862**. Springer, Berlin. MR2130405

[25] HELFFER, B. and NIER, F. (2006). Quantitative analysis of metastability in reversible diffusion processes via a Witten complex approach: The case with boundary. *Mém. Soc. Math. Fr.* (*N.S.*) **105** vi+89. MR2270650

[26] HIGUCHI, Y. and YOSHIDA, N. (1995). Analytic conditions and phase transition for ising models (in Japanese). In *Seminar on Probability, Vol. 58* 1–46. Available at http://www.math.nagoya-u.ac.jp/~noby/list.html.

[27] HOLLEY, R. and STROOCK, D. (1987). Logarithmic Sobolev inequalities and stochastic Ising models. *J. Stat. Phys.* **46** 1159–1194. MR0893137

[28] JERRUM, M., SON, J.-B., TETALI, P. and VIGODA, E. (2004). Elementary bounds on Poincaré and log-Sobolev constants for decomposable Markov chains. *Ann. Appl. Probab.* **14** 1741–1765. MR2099650

[29] JOST, J. (2008). *Riemannian Geometry and Geometric Analysis*, 5th ed. Springer, Berlin. MR2431897

[30] KRAMERS, H. A. (1940). Brownian motion in a field of force and the diffusion model of chemical reactions. *Physica* **7** 284–304. MR0002962

[31] LEDOUX, M. (1999). Concentration of measure and logarithmic Sobolev inequalities. In *Séminaire de Probabilités, XXXIII. Lecture Notes in Math.* **1709** 120–216. Springer, Berlin. MR1767995

[32] LEGOLL, F. and LELIÈVRE, T. (2010). Effective dynamics using conditional expectations. *Nonlinearity* **23** 2131–2163. MR2672640

[33] LELIÈVRE, T. (2009). A general two-scale criteria for logarithmic Sobolev inequalities. *J. Funct. Anal.* **256** 2211–2221. MR2498763

[34] MEISS, J. D. (2007). *Differential Dynamical Systems. Mathematical Modeling and Computation* **14**. SIAM, Philadelphia, PA. MR2384060

[35] MIELKE, A. (2013). Geodesic convexity of the relative entropy in reversible Markov chains. *Calc. Var. Partial Differential Equations* **48** 1–31. MR3090532

[36] MUCKENHOUPT, B. (1972). Hardy's inequality with weights. *Studia Math.* **44** 31–38. MR0311856

[37] ØKSENDAL, B. (1998). *Stochastic Differential Equations: An Introduction with Applications*, 5th ed. Springer, Berlin. MR1619188

[38] OLVER, F. W. J. (1997). *Asymptotics and Special Functions*. A K Peters Ltd., Wellesley, MA. MR1429619

[39] OTTO, F. and REZNIKOFF, M. G. (2007). A new criterion for the logarithmic Sobolev inequality and two applications. *J. Funct. Anal.* **243** 121–157. MR2291434

[40] OTTO, F. and VILLANI, C. (2000). Generalization of an inequality by Talagrand and links with the logarithmic Sobolev inequality. *J. Funct. Anal.* **173** 361–400. MR1760620

[41] ROTHAUS, O. S. (1978). Lower bounds for eigenvalues of regular Sturm-Liouville operators and the logarithmic Sobolev inequality. *Duke Math. J.* **45** 351–362. MR0481241

[42] ROYER, G. (2007). *An Initiation to Logarithmic Sobolev Inequalities. SMF/AMS Texts and Monographs* **14**. Amer. Math. Soc., Providence, RI. Translated from the 1999 French original by Donald Babbitt. MR2352327

[43] SCHLICHTING, A. (2012). The Eyring-Kramers formula for Poincaré and logarithmic Sobolev inequalities. Ph.D. thesis, Univ. Leipzig.
[44] SCHUSS, Z. (2010). *Theory and Applications of Stochastic Processes*: *An Analytical Approach*. *Applied Mathematical Sciences* **170**. Springer, New York. MR2583642
[45] SUGIURA, M. (1995). Metastable behaviors of diffusion processes with small parameter. *J. Math. Soc. Japan* **47** 755–788. MR1348758
[46] VILLANI, C. (2009). *Optimal Transport*: *Old and New*. *Grundlehren der Mathematischen Wissenschaften* [*Fundamental Principles of Mathematical Sciences*] **338**. Springer, Berlin. MR2459454
[47] WALES, D. (2003). *Energy Landscapes*: *Applications to Clusters*, *Biomolecules and Glasses*. Cambridge Univ. Press, Cambridge.

DEPARTMENT OF MATHEMATICS
STANFORD UNIVERSITY
BUILDING 380
STANFORD, CALIFORNIA 94305
USA
E-MAIL: gmenz@stanford.edu

INSTITUTE FOR APPLIED MATHEMATICS
UNIVERSITY OF BONN
ENDENICHER ALLEE 60
D-53115 BONN
GERMANY
E-MAIL: schlichting@iam.uni-bonn.de