

Concentration for Coulomb gases on compact manifolds

David García-Zelada*

Abstract

We study the non-asymptotic behavior of a Coulomb gas on a compact Riemannian manifold. This gas is a symmetric n -particle Gibbs measure associated to the two-body interaction energy given by the Green function. We encode such a particle system by using an empirical measure. Our main result is a concentration inequality in Kantorovich-Wasserstein distance inspired from the work of Chafaï, Hardy and Maïda on the Euclidean space. Their proof involves large deviation techniques together with an energy-distance comparison and a regularization procedure based on the superharmonicity of the Green function. This last ingredient is not available on a manifold. We solve this problem by using the heat kernel and its short-time asymptotic behavior.

Keywords: Gibbs measure; Green function; Coulomb gas; empirical measure; concentration of measure; interacting particle system; singular potential; heat kernel.

AMS MSC 2010: 60B05; 26D10; 35K05.

Submitted to ECP on September 12, 2018, final version accepted on January 16, 2019.

Supersedes arXiv:1809.04231v1.

1 Introduction

We shall consider the model of a Coulomb gas on a Riemannian manifold introduced in [6, Subsection 4.1] and study its non-asymptotic behavior by obtaining a concentration inequality for the empirical measure around its limit. Let us describe the model and the main theorem of this article.

Let (M, g) be a compact Riemannian manifold of volume form π . We suppose, for simplicity, that $\pi(M) = 1$ so that $\pi \in \mathcal{P}(M)$ where $\mathcal{P}(M)$ denotes the space of probability measures on M . We endow $\mathcal{P}(M)$ with the topology of weak convergence, i.e. the smallest topology such that $\mu \rightarrow \int_M f d\mu$ is continuous for every continuous function $f : M \rightarrow \mathbb{R}$. Denote by $\Delta : C^\infty(M) \rightarrow C^\infty(M)$ the Laplace-Beltrami operator on (M, g) . We shall say that

$$G : M \times M \rightarrow (-\infty, \infty]$$

is a Green function for Δ if it is a symmetric continuous function such that for every $x \in M$ the function $G_x : M \rightarrow (-\infty, \infty]$ defined by $G_x(y) = G(x, y)$ is integrable and

$$\Delta G_x = -\delta_x + 1 \tag{1.1}$$

*Université Paris-Dauphine, PSL Research university, CEREMADE, Place du Maréchal de Lattre de Tassigny, 75016 Paris, France.

E-mail: garciazelada@ceremade.dauphine.fr

in the distributional sense. It can be proven that such a G is integrable with respect to $\pi \otimes \pi$ and that if $f \in C^\infty(M)$ then $\psi : M \rightarrow \mathbb{R}$, defined by

$$\psi(x) = \int_M G(x, y) f(y) d\pi(y),$$

satisfies that

$$\psi \in C^\infty(M) \quad \text{and} \quad \Delta\psi = -f + \int_M f(y) d\pi(y). \quad (1.2)$$

In particular, $\int_M G_x d\pi$ does not depend on $x \in M$ and the Green function is unique up to an additive constant. See [1, Chapter 4] for a proof of these results. We will denote by G the Green function for Δ such that

$$\int_M G_x d\pi = 0 \quad (1.3)$$

for every $x \in M$.

For $x \in M$ the function G_x may be thought of as the potential generated by the distribution of charge $\delta_x - 1$. This would represent a unit charged particle located at $x \in M$ and a negatively charged background of unit density. The total energy of a system of n particles of charge $1/n$ (each particle coming with a negatively charged background) would be $H_n : M^n \rightarrow (-\infty, \infty]$ defined by

$$H_n(x_1, \dots, x_n) = \frac{1}{n^2} \sum_{i < j} G(x_i, x_j).$$

Take a sequence $\{\beta_n\}_{n \geq 2}$ of non-negative numbers and consider the sequence of Gibbs probability measures

$$d\mathbb{P}_n(x_1, \dots, x_n) = \frac{1}{Z_n} e^{-\beta_n H_n(x_1, \dots, x_n)} d\pi^{\otimes n}(x_1, \dots, x_n) \quad (1.4)$$

where Z_n is such that $\mathbb{P}_n(M^n) = 1$. This can be thought of as the Riemannian generalization of the usual Coulomb gas model described in [15] or [4]. In the particular case of the round two-dimensional sphere, it is known (see [9]) that if $\beta_n = 4\pi n^2$ the probability measure \mathbb{P}_n describes the eigenvalues of the quotient of two independent $n \times n$ matrices with independent Gaussian entries. Define $H : \mathcal{P}(M) \rightarrow (-\infty, \infty]$ by

$$H(\mu) = \frac{1}{2} \int_{M \times M} G(x, y) d\mu(x) d\mu(y).$$

This is a convex lower semicontinuous function. We can see [6, Subsection 4.1] for a proof of these properties and [12, Chapter 9] for a short introduction and further information in the Euclidean setting. Let $i_n : M^n \rightarrow \mathcal{P}(M)$ be defined by

$$i_n(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \delta_{x_i}.$$

If $\beta_n/n \rightarrow \infty$, the author in [6] tells us that $\{i_{n*}(\mathbb{P}_n)\}_{n \geq 2}$, the sequence of image measures of \mathbb{P}_n by i_n , satisfies a large deviation principle with speed β_n and rate function $H - \inf H$.

In particular, if F is a closed set of $\mathcal{P}(M)$ we have

$$\limsup_{n \rightarrow \infty} \frac{1}{\beta_n} \log \mathbb{P}_n(i_n^{-1}(F)) \leq - \inf_{\mu \in F} (H(\mu) - \inf H)$$

or, equivalently,

$$\mathbb{P}_n(i_n^{-1}(F)) \leq \exp \left(-\beta_n \inf_{\mu \in F} (H(\mu) - \inf H) + o(\beta_n) \right). \quad (1.5)$$

The aim of this article is to understand the $o(\beta_n)$ term for some family of closed sets F . Suppose we choose some metric d in $\mathcal{P}(M)$ that induces the topology of weak convergence. As the unique minimizer of H is $\mu_{eq} = \pi$ (see Theorem 3.1) a nice family of closed sets are the sets

$$F_r = \{\mu \in \mathcal{P}(M) : d(\mu, \mu_{eq}) \geq r\}$$

for $r > 0$. Instead of writing $\mathbb{P}_n(i_n^{-1}(F_r))$ we shall write $\mathbb{P}_n(d(i_n, \mu_{eq}) \geq r)$, in other words, when we write $\{d(i_n, \mu_{eq}) \geq r\}$ we mean the set $i_n^{-1}(F_r) = \{\vec{x} \in M^n : d(i_n(\vec{x}), \mu_{eq}) \geq r\}$. As H is lower semicontinuous we have that $\inf_{\mu \in F_r} (H(\mu) - \inf H)$ is strictly positive and the large deviation inequality is not vacuous. We would like a simple expression in terms of r for the leading term, so instead of using $\inf_{\mu \in F_r} (H(\mu) - \inf H)$ we will use a simple function of r .

Let d_g denote the Riemannian distance. The metric we shall use on $\mathcal{P}(M)$ is the function $W_1 : \mathcal{P}(M) \times \mathcal{P}(M) \rightarrow [0, \infty)$ defined by

$$W_1(\mu, \nu) = \inf \left\{ \int_{M \times M} d_g(x, y) d\Pi(x, y) : \Pi \text{ is a coupling between } \mu \text{ and } \nu \right\} \quad (1.6)$$

which is known as the Wasserstein or Kantorovich metric. See [16, Theorem 7.12] for a proof that it metrizes the topology of weak convergence. The main result of this article is the following.

Theorem 1.1 (Concentration inequality for Coulomb gases). *Let m be the dimension of M . If $m = 2$ there exists a constant $C > 0$ that does not depend on the sequence $\{\beta_n\}_{n \geq 2}$ such that for every $n \geq 2$ and $r \geq 0$*

$$\mathbb{P}_n(W_1(i_n, \pi) \geq r) \leq \exp \left(-\beta_n \frac{r^2}{4} + \frac{\beta_n \log(n)}{8\pi n} + C \frac{\beta_n}{n} \right).$$

If $m \geq 3$ there exists a constant $C > 0$ that does not depend on the sequence $\{\beta_n\}_{n \geq 2}$ such that for every $n \geq 2$ and $r \geq 0$

$$\mathbb{P}_n(W_1(i_n, \pi) \geq r) \leq \exp \left(-\beta_n \frac{r^2}{4} + C \frac{\beta_n}{n^{2/m}} \right).$$

In fact, by a slight modification we will also prove the following generalization. Denote by $D(\cdot \| \pi) : \mathcal{P}(M) \rightarrow (-\infty, \infty]$ the relative entropy of μ with respect to π , also known as the Kullback–Leibler divergence, i.e. $D(\mu \| \pi) = \int_M \rho \log \rho d\pi$ if $d\mu = \rho d\pi$ and $D(\cdot \| \pi)$ is infinity when μ is not absolutely continuous with respect to π .

Theorem 1.2 (Concentration inequality for Coulomb gases in a potential). *Take a twice continuously differentiable function $V : M \rightarrow \mathbb{R}$ and define*

$$H_n(x_1, \dots, x_n) = \frac{1}{n^2} \sum_{i < j} G(x_i, x_j) + \frac{1}{n} \sum_{i=1}^n V(x_i) \quad \text{and}$$

$$H(\mu) = \frac{1}{2} \int_{M \times M} G(x, y) d\mu(x) d\mu(y) + \int_M V(x) d\mu(x).$$

Then H has a unique minimizer that will be called μ_{eq} . Suppose \mathbb{P}_n is defined by (1.4) and let m be the dimension of M . If $m = 2$ there exists a constant $C > 0$ that does not depend on the sequence $\{\beta_n\}_{n \geq 2}$ such that for every $n \geq 2$ and $r \geq 0$

$$\mathbb{P}_n(W_1(i_n, \mu_{eq}) \geq r) \leq \exp \left(-\beta_n \frac{r^2}{4} + \frac{\beta_n \log(n)}{8\pi n} + n D(\mu_{eq} \| \pi) + C \frac{\beta_n}{n} \right).$$

If $m \geq 3$ there exists a constant $C > 0$ that does not depend on the sequence $\{\beta_n\}_{n \geq 2}$ such that for every $n \geq 2$ and $r \geq 0$

$$\mathbb{P}_n(W_1(i_n, \mu_{\text{eq}}) \geq r) \leq \exp\left(-\beta_n \frac{r^2}{4} + nD(\mu_{\text{eq}} \parallel \pi) + C \frac{\beta_n}{n^{2/m}}\right).$$

Remark 1.3 (About the sharpness). As we will see below it can be proven that

$$\mathbb{P}_n(W_1(i_n, \pi) \geq r) \leq \exp\left(-\beta_n \frac{r^2}{2} + o(\beta_n)\right)$$

and the natural question would be to find an explicit next order $o(\beta_n)$. In the two theorems above we have relaxed this inequality to

$$\mathbb{P}_n(W_1(i_n, \pi) \geq r) \leq \exp\left(-\beta_n \frac{r^2}{4} + o(\beta_n)\right)$$

and obtained a bound to $o(\beta_n)$ that does not depend on r . In this relaxed inequality and at a fixed $r > 0$ the next order terms cannot be exact. Indeed, strictly speaking we have

$$\mathbb{P}_n(W_1(i_n, \pi) \geq r) \leq \exp\left(-\beta_n \frac{r^2}{4} + \eta(\beta_n)\right)$$

where $\eta(\beta)/\beta \rightarrow -r^2/2$ as β goes to infinity. Nevertheless, the importance of our result lies on the lack of dependence on r and the explicitness of the terms.

To prove Theorem 1.1 we follow [4] in turn inspired by [13] (see also [14]). We proceed in three steps. The first part, described in Section 2, may be used in any measurable space but it demands an energy-distance comparison and a regularization procedure. The energy-distance comparison will be explained in Section 3 and it may be extended to include Green functions of some Laplace-type operators. The regularization by the heat kernel, in Section 4, will use a short time asymptotic expansion. It may apply to more general kind of energies where a short-time asymptotic expansion of their heat kernel is known. Having acquired all the tools, Section 5 will complete the proof of Theorem 1.1 and, by a slight modification, Theorem 1.2.

2 Link to an energy-distance comparison and a regularization procedure

In this section M may be any measurable space, π any probability measure on M and $H_n : M^n \rightarrow (-\infty, \infty]$ any measurable function bounded from below. Given $\beta_n > 0$ we define the Gibbs probability measure by (1.4). Let $H : \mathcal{P}(M) \rightarrow (-\infty, \infty]$ be any function that has a unique minimizer $\mu_{\text{eq}} \in \mathcal{P}(M)$. This shall be thought of as a rate function of some Laplace principle as in [6]. Consider a metric

$$d : \mathcal{P}(M) \times \mathcal{P}(M) \rightarrow [0, \infty)$$

on $\mathcal{P}(M)$ that induces the topology of weak convergence and define

$$F_r = \{\mu \in \mathcal{P}(M) : d(\mu, \mu_{\text{eq}}) \geq r\}$$

for $r > 0$. We want to understand a non-asymptotic inequality similar to (1.5) with an explicit $o(\beta_n)$ term. For this, we consider the following assumption.

Assumption 2.1. We will say that an increasing convex function $f : [0, \infty) \rightarrow [0, \infty)$ satisfies Assumption A if, for all $\mu \in \mathcal{P}(M)$,

$$f(d(\mu, \mu_{\text{eq}})) \leq H(\mu) - H(\mu_{\text{eq}}). \quad (\text{A})$$

Under Assumption A, (1.5) implies

$$\mathbb{P}_n(i_n^{-1}(F_r)) \leq \exp(-\beta_n f(r) + o(\beta_n)). \quad (2.1)$$

This $o(\beta_n)$ term may depend on r . We will prove that if we relax the inequality (2.1) to

$$\mathbb{P}_n(i_n^{-1}(F_r)) \leq \exp(-\beta_n 2f(r/2) + o(\beta_n))$$

we can find bounds of the $o(\beta_n)$ term that do not depend on r . To properly use Assumption A when μ is an empirical measure $\frac{1}{n} \sum_{i=1}^n \delta_{x_i}$ we will have to regularize μ . The reason behind this is that when μ is an empirical measure we usually obtain $H(\mu) = \infty$ by the self-interactions of the particles with themselves. In the Euclidean setting this regularization is obtained by a convolution with a radial distribution while in the Riemannian setting this will be obtained by a diffusion using the heat kernel of the Laplacian which in the Euclidean case may be seen as a convolution by a Gaussian function. The following result is the general concentration inequality we get and it is the first part of the method mentioned in Section 1.

Theorem 2.2 (General concentration inequality). *Suppose we have two real numbers a_n and b_n such that there exists a measurable function $R : M^n \rightarrow \mathcal{P}(M)$ with the following property*

- *for every $\vec{x} = (x_1, \dots, x_n) \in M^n$ we have*

$$H_n(x_1, \dots, x_n) \geq H(R(\vec{x})) - a_n, \quad \text{and} \quad d(R(\vec{x}), i_n(\vec{x})) \leq b_n.$$

Let us denote $e_n = \int_{M^n} H_n d\mu_{\text{eq}}^{\otimes n}$ and $e = H(\mu_{\text{eq}})$. If $f : [0, \infty) \rightarrow [0, \infty)$ is an increasing convex function that satisfies Assumption A then

$$\mathbb{P}_n(d(i_n, \mu_{\text{eq}}) \geq r) \leq \exp\left(-\beta_n 2f\left(\frac{r}{2}\right) + nD(\mu_{\text{eq}} \parallel \pi) + \beta_n(e_n - e) + \beta_n a_n + \beta_n f(b_n)\right).$$

Proof. We first prove the two following results. The first lemma we state is the analogue of [4, Lemma 4.1].

Lemma 2.3 (Lower bound of the partition function). *We have the following lower bound.*

$$Z_n \geq \exp(-\beta_n e_n - nD(\mu_{\text{eq}} \parallel \pi)).$$

Proof. If $d\mu_{\text{eq}} = \rho_{\text{eq}} d\pi$ we have

$$\begin{aligned} Z_n &= \int_{M^n} e^{-\beta_n H_n(x_1, \dots, x_n)} d\pi^{\otimes n}(x_1, \dots, x_n) \\ &\geq \int_{M^n} e^{-\beta_n H_n(x_1, \dots, x_n)} e^{-\sum_{i=1}^n 1_{\rho_{\text{eq}} > 0}(x_i) \log \rho_{\text{eq}}(x_i)} d\mu_{\text{eq}}^{\otimes n}(x_1, \dots, x_n) \\ &\geq \int_{M^n} e^{-\beta_n H_n(x_1, \dots, x_n) - \sum_{i=1}^n 1_{\rho_{\text{eq}} > 0}(x_i) \log \rho_{\text{eq}}(x_i)} d\mu_{\text{eq}}^{\otimes n}(x_1, \dots, x_n) \\ &\geq e^{-\int_{M^n} (\beta_n H_n(x_1, \dots, x_n) + \sum_{i=1}^n 1_{\rho_{\text{eq}} > 0}(x_i) \log \rho_{\text{eq}}(x_i)) d\mu_{\text{eq}}^{\otimes n}(x_1, \dots, x_n)} \\ &= e^{-\beta_n e_n - nD(\mu_{\text{eq}} \parallel \pi)} \end{aligned}$$

where we have used Jensen's inequality to get the last inequality. □

The second lemma will help us in the step of regularization.

Lemma 2.4 (Comparison). *Take $\vec{x} = (x_1, \dots, x_n) \in M^n$. If $d(R(\vec{x}), i_n(\vec{x})) \leq b_n$ then*

$$f(d(R(\vec{x}), \mu_{\text{eq}})) \geq 2f\left(\frac{d(i_n(\vec{x}), \mu_{\text{eq}})}{2}\right) - f(b_n).$$

Proof. As

$$d(i_n(\vec{x}), \mu_{\text{eq}}) \leq d(i_n(\vec{x}), R(\vec{x})) + d(R(\vec{x}), \mu_{\text{eq}})$$

we have that

$$\begin{aligned} f\left(\frac{d(i_n(\vec{x}), \mu_{\text{eq}})}{2}\right) &\leq f\left(\frac{1}{2}d(i_n(\vec{x}), R(\vec{x})) + \frac{1}{2}d(R(\vec{x}), \mu_{\text{eq}})\right) \\ &\leq \frac{1}{2}f(d(i_n(\vec{x}), R(\vec{x}))) + \frac{1}{2}f(d(R(\vec{x}), \mu_{\text{eq}})) \\ &\leq \frac{1}{2}f(b_n) + \frac{1}{2}f(d(R(\vec{x}), \mu_{\text{eq}})) \end{aligned}$$

where we have used that f is increasing and convex. \square

Now, define

$$A_r = i_n^{-1}(F_r) = \{\vec{x} \in M^n : d(i_n(\vec{x}), \mu_{\text{eq}}) \geq r\}.$$

Then

$$\begin{aligned} \mathbb{P}_n(A_r) &= \frac{1}{Z_n} \int_{A_r} e^{-\beta_n H_n(x_1, \dots, x_n)} d\pi^{\otimes n}(x_1, \dots, x_n) \\ &\leq e^{\beta_n e_n + nD(\mu_{\text{eq}}\|\pi)} \int_{A_r} e^{-\beta_n H(R(\vec{x})) + \beta_n a_n} d\pi^{\otimes n}(x_1, \dots, x_n) \\ &\leq e^{\beta_n e_n + \beta_n a_n + nD(\mu_{\text{eq}}\|\pi)} \int_{A_r} e^{-\beta_n H(R(\vec{x}))} d\pi^{\otimes n}(x_1, \dots, x_n) \\ &\stackrel{(*)}{\leq} e^{\beta_n(e_n - e) + \beta_n a_n + nD(\mu_{\text{eq}}\|\pi)} \int_{A_r} e^{-\beta_n f(d(R(\vec{x}), \mu_{\text{eq}}))} d\pi^{\otimes n}(x_1, \dots, x_n) \\ &\stackrel{(**)}{\leq} e^{\beta_n(e_n - e) + \beta_n a_n + nD(\mu_{\text{eq}}\|\pi)} \int_{A_r} e^{-\beta_n 2f\left(\frac{d(i_n(\vec{x}), \mu_{\text{eq}})}{2}\right) + \beta_n f(b_n)} d\pi^{\otimes n}(x_1, \dots, x_n) \\ &\stackrel{(***)}{\leq} e^{\beta_n(e_n - e) + \beta_n a_n + nD(\mu_{\text{eq}}\|\pi)} e^{-\beta_n 2f\left(\frac{r}{2}\right) + \beta_n f(b_n)} \\ &\leq e^{-\beta_n 2f\left(\frac{r}{2}\right) + nD(\mu_{\text{eq}}\|\pi) + \beta_n(e_n - e) + \beta_n a_n + \beta_n f(b_n)} \end{aligned}$$

where in $(*)$ we have used Assumption A, in $(**)$ we have used Lemma 2.4 and in $(***)$ we have used the monotonicity of f . \square

In the next section we return to the case of a compact Riemannian manifold and study a energy-distance comparison that will imply Assumption A.

3 Energy-distance comparison in compact Riemannian manifolds

We take the notation used in Section 1. The Kantorovich metric W_1 defined in (1.6) can be written as

$$W_1(\mu, \nu) = \sup \left\{ \int_M f d\mu - \int_M f d\nu : \|f\|_{\text{Lip}} \leq 1 \right\}$$

where

$$\|f\|_{\text{Lip}} = \sup_{x \neq y} \frac{|f(x) - f(y)|}{d_g(x, y)}.$$

This result is known as the Kantorovich-Rubinstein theorem (see [16, Theorem 1.14]). In the case of a Riemannian manifold, by a smooth approximation argument such as the one in [2], we can prove that

$$W_1(\mu, \nu) = \sup \left\{ \int_M f \, d\mu - \int_M f \, d\nu : f \in C^\infty(M) \text{ and } \|\nabla f\|_\infty \leq 1 \right\}.$$

The next theorem gives the energy-distance comparison required to satisfy Assumption A. This is the analogue of [13, Theorem 1.3] and [4, Lemma 3.1].

Theorem 3.1 (Comparison between distance and energy). *Suppose that $\mu_{\text{eq}} \in \mathcal{P}(M)$ is a probability measure on M such that $H(\mu_{\text{eq}}) \leq H(\mu)$ for every $\mu \in \mathcal{P}(M)$. Then*

$$\frac{1}{2} W_1(\mu, \mu_{\text{eq}})^2 \leq H(\mu) - H(\mu_{\text{eq}}) \quad (3.1)$$

for every $\mu \in \mathcal{P}(M)$. This implies, in particular, that H has a unique minimizer and that Assumption A is satisfied by $f(r) = \frac{r^2}{2}$. Furthermore, $\mu_{\text{eq}} = \pi$.

Let \mathcal{F} be the space of finite signed measures μ on M such that $\int_{M \times M} G \, d|\mu|^{\otimes 2} < \infty$. For convenience we shall define $\mathcal{E} : \mathcal{F} \rightarrow (-\infty, \infty]$ by

$$\mathcal{E}(\mu) = \int_{M \times M} G(x, y) \, d\mu(x) \, d\mu(y)$$

so that $\mathcal{E}(\mu) = 2H(\mu)$ whenever $\mu \in \mathcal{P}(M) \cap \mathcal{F}$. We can also notice that if we take two probability measures $\mu, \nu \in \mathcal{P}(M)$ such that $H(\mu)$ and $H(\nu)$ are finite then, due to the convexity of H , we have $\int_{M \times M} G(x, y) \, d\mu(x) \, d\nu(y) < \infty$, the measure $\mu - \nu$ belongs to \mathcal{F} and

$$\mathcal{E}(\mu - \nu) = \mathcal{E}(\mu) + \mathcal{E}(\nu) - 2 \int_{M \times M} G(x, y) \, d\mu(x) \, d\nu(y). \quad (3.2)$$

We begin by proving the following result that may be seen as a comparison of distances where the ‘energy distance’ between two probability measures $\mu, \nu \in \mathcal{P}(M)$ of finite energy is defined as $\sqrt{\mathcal{E}(\mu - \nu)}$. This is the analogue of [4, Theorem 1.1].

Lemma 3.2 (Comparison of distances). *Let $\mu, \nu \in \mathcal{P}(M)$ such that $H(\mu)$ and $H(\nu)$ are finite. Then $\mathcal{E}(\mu - \nu) \geq 0$ and*

$$W_1(\mu, \nu) \leq \sqrt{\mathcal{E}(\mu - \nu)}.$$

Proof. First suppose μ and ν differentiable, i.e. suppose they have a differentiable density with respect to π . Define $U : M \rightarrow \mathbb{R}$ by

$$U(x) = \int_M G(x, y) \, (d\mu(y) - d\nu(y)).$$

Then, as remarked in (1.2), we know that U is differentiable and

$$\Delta U = -(\mu - \nu).$$

Take $f \in C^\infty(M)$ such that $\|\nabla f\|_\infty \leq 1$. We can see that

$$\int_M f \, (d\mu - d\nu) = - \int_M f \, \Delta U = \int_M \langle \nabla f, \nabla U \rangle \, d\pi \leq \|\nabla f\|_2 \|\nabla U\|_2 \leq \|\nabla f\|_\infty \|\nabla U\|_2.$$

We also know that

$$(\|\nabla U\|_2)^2 = \int_M \langle \nabla U, \nabla U \rangle \, d\pi = - \int_M U \, \Delta U = \int_M U \, (d\mu - d\nu) = \mathcal{E}(\mu - \nu).$$

Then,

$$\int_M f(d\mu - d\nu) \leq \|\nabla f\|_\infty \|\nabla U\|_2 \leq \|\nabla f\|_\infty \sqrt{\mathcal{E}(\mu - \nu)}.$$

This implies that

$$W_1(\mu, \nu) \leq \sqrt{\mathcal{E}(\mu - \nu)}.$$

In general, let $\mu, \nu \in \mathcal{P}(M)$ such that $H(\mu)$ and $H(\nu)$ are finite. Take two sequences $\{\mu_n\}_{n \in \mathbb{N}}$ and $\{\nu_n\}_{n \in \mathbb{N}}$ of differentiable probability measures that converge to μ and ν respectively and such that $\mathcal{E}(\mu_n) \rightarrow \mathcal{E}(\mu)$ and $\mathcal{E}(\nu_n) \rightarrow \mathcal{E}(\nu)$ (see [3] for a proof of their existence) and proceed by a limit argument. \square

The next step to prove Theorem 3.1 is a fact that works for general two-body interactions i.e. G is not necessarily a Green function.

Lemma 3.3 (Comparison of energies). *Suppose that μ_{eq} is a probability measure such that $H(\mu_{\text{eq}}) \leq H(\mu)$ for every $\mu \in \mathcal{P}(M)$. Then, for every $\mu \in \mathcal{P}(M)$ such that $H(\mu) < \infty$, we have*

$$\mathcal{E}(\mu - \mu_{\text{eq}}) \leq \mathcal{E}(\mu) - \mathcal{E}(\mu_{\text{eq}}).$$

Proof. As $H(\mu)$ and $H(\mu_{\text{eq}})$ are finite we use (3.2) to notice that the affirmation

$$\mathcal{E}(\mu - \mu_{\text{eq}}) \leq \mathcal{E}(\mu) - \mathcal{E}(\mu_{\text{eq}})$$

is equivalent to

$$\int_{M \times M} G(x, y) d\mu(x) d\mu_{\text{eq}}(y) \geq \mathcal{E}(\mu_{\text{eq}}).$$

But, if

$$\int_{M \times M} G(x, y) d\mu(x) d\mu_{\text{eq}}(y) < \mathcal{E}(\mu_{\text{eq}})$$

were true then, defining $\mu_t = (1-t)\mu_{\text{eq}} + t\mu = \mu_{\text{eq}} + t(\mu - \mu_{\text{eq}})$, we would see that the linear term of $\mathcal{E}(\mu_t)$ is $\int_{M \times M} G(x, y) d\mu(x) d\mu_{\text{eq}}(y) - \mathcal{E}(\mu_{\text{eq}}) < 0$. This means that $\mathcal{E}(\mu_t) < \mathcal{E}(\mu_{\text{eq}})$ for $t > 0$ small which is a contradiction. \square

Now we may complete the proof of Theorem 3.1.

Proof of Theorem 3.1. Let μ_{eq} be a minimizer of H and let $\mu \in \mathcal{P}(M)$ be a probability measure on M . If $H(\mu)$ is infinite there is nothing to prove. If it is not, by Lemma 3.2 and 3.3 we conclude (3.1).

To prove that H has a unique minimizer suppose $\tilde{\mu}_{\text{eq}}$ is another minimizer and use Inequality (3.1) with $\mu = \tilde{\mu}_{\text{eq}}$ to get $W_1(\tilde{\mu}_{\text{eq}}, \mu_{\text{eq}}) = 0$ and, thus, $\tilde{\mu}_{\text{eq}} = \mu_{\text{eq}}$.

Finally, to see that $\mu_{\text{eq}} = \pi$ we use (1.3). Then $\mathcal{E}(\mu - \pi) = \mathcal{E}(\mu) - \mathcal{E}(\pi)$ when μ has finite energy. But by Lemma 3.2 we know that $\mathcal{E}(\mu - \pi) \geq 0$ and then $\mathcal{E}(\mu) \geq \mathcal{E}(\pi)$ for every $\mu \in \mathcal{P}(M)$ of finite energy. \square

In the next section we study a way to regularize the empirical measures in the sense of the hypotheses of Theorem 2.2.

4 Heat kernel regularization of the energy

In this section the main tool is the heat kernel for Δ . A proof of the following proposition may be found in [5, Chapter VI].

Proposition 4.1 (Heat kernel). *There exists a unique differentiable function*

$$p : (0, \infty) \times M \times M \rightarrow \mathbb{R}$$

such that

$$\frac{\partial}{\partial t} p_t(x, y) = \Delta_y p_t(x, y) \quad \text{and} \quad \lim_{t \rightarrow 0} p_t(x, \cdot) = \delta_x$$

for every $x, y \in M$ and $t > 0$. Such a function will be called the heat kernel for Δ . It is non-negative, it is mass preserving, i.e.

$$\int_M p_t(x, y) d\pi(y) = 1$$

for every $x \in M$ and $t > 0$, it is symmetric, i.e.

$$p_t(x, y) = p_t(y, x)$$

for every $x, y \in M$ and $t > 0$ and it satisfies the semigroup property i.e.

$$\int_M p_t(x, y) p_s(y, z) d\pi(y) = p_{t+s}(x, z)$$

for every $x, y \in M$ and $t, s > 0$. Furthermore,

$$\lim_{t \rightarrow \infty} p_t(x, y) = 1$$

uniformly on x and y .

Let p be the heat kernel associated to Δ . For each point $x \in M$ and $t > 0$ define the probability measure $\mu_x^t \in \mathcal{P}(M)$ by

$$d\mu_x^t = p_t(x, \cdot) d\pi, \quad (4.1)$$

or, more precisely, $d\mu_x^t(y) = p_t(x, y) d\pi(y)$. Then we define $R_t : M^n \rightarrow \mathcal{P}(M)$ by

$$R_t(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \mu_{x_i}^t$$

and we want to find a_n and b_n of the hypotheses of Theorem 2.2 for $R = R_t$.

We begin by looking for b_n .

4.1 Distance to the regularized measure

Proposition 4.2 (Distance to the regularized measure). *There exists a constant $C > 0$ such that for all $t > 0$ and $\vec{x} \in M^n$*

$$W_1(R_t(\vec{x}), i_n(\vec{x})) \leq C\sqrt{t}.$$

Proof. The following arguments are very similar to those in [11] and they will be repeated for convenience of the reader. As $W_1 : \mathcal{P}(M) \times \mathcal{P}(M) \rightarrow [0, \infty)$ is the supremum of linear functions, it is convex. So

$$W_1(R_t(\vec{x}), i_n(\vec{x})) \leq \frac{1}{n} \sum_{i=1}^n W_1(\delta_{x_i}, \mu_{x_i}^t).$$

Then, we will try to find a constant $C > 0$ such that $W_1(\delta_x, \mu_x^t) \leq C\sqrt{t}$ for every $x \in M$. As the only coupling between δ_x and μ_x^t is their product we see that

$$W_1(\delta_x, \mu_x^t) = \int_M d_g(x, y) d\mu_x^t(y).$$

In fact we will study the 2-Kantorovich squared distance between δ_x and μ_x^t

$$\begin{aligned} D_t(x) &= \int_M d_g(x, y)^2 d\mu_x^t(y) \\ &= \int_M d_g(x, y)^2 p_t(x, y) d\pi(y). \end{aligned}$$

If we prove that there exists a constant $C > 0$ such that for every $x \in M$

$$D_t(x) \leq C^2 t \quad (4.2)$$

we may conclude that $W_1(\delta_x, \mu_x^t) \leq C\sqrt{t}$ for every $x \in M$ by Jensen's inequality. To obtain (4.2) we use the following lemma which proof may be found in [8, Section 3.4] and [8, Theorem 3.5.1].

Lemma 4.3 (Radial process representation). *Take $x \in M$. Let X be the Markov process with generator Δ starting at x (i.e. $X_t = B_{2t}$ where B is a Brownian motion on M starting at x). Define $r : M \rightarrow [0, \infty)$ by $r(y) = d_g(x, y)$. Then r is differentiable π -almost everywhere and there exists a non-decreasing process L and a one-dimensional Euclidean Brownian motion β such that*

$$r(X_t) = \beta_{2t} + \int_0^t \Delta r(X_s) ds - L_t$$

for every $t \geq 0$ where Δr is the π -almost everywhere defined Laplacian of r .

Applying Lemma 4.3 and Itô's formula and then taking expected values we get

$$\mathbb{E}[r(X_t)^2] = 2 \int_0^t \mathbb{E}[r(X_s) \Delta r(X_s)] ds - \mathbb{E} \left[2 \int_0^t r(X_s) dL_s \right] + 2t \leq \int_0^t 2\mathbb{E}[r(X_s) \Delta r(X_s)] ds + 2t$$

where we are using the notation of Lemma 4.3. By [8, Corollary 3.4.5] we know that $r \Delta r$ is bounded in M and as $D_t(x) = \mathbb{E}[r(X_t)^2]$ we obtain (4.2) where the constant C does not depend on x . \square

Now we will look for a_n of the hypotheses of Theorem 2.2.

4.2 Comparison between the regularized and the non-regularized energy

Theorem 4.4 (Comparison between the regularized and the non-regularized energy). *Let m be the dimension of M . If $m = 2$ there exists a constant $C > 0$ such that, for every $n \geq 2$, $t \in (0, 1]$ and $\vec{x} \in M^n$,*

$$H_n(\vec{x}) \geq H(R_t(\vec{x})) - t + \frac{1}{8\pi n} \log(t) - \frac{C}{n}.$$

If $m > 2$ there exists a constant $C > 0$ such that, for every $n \geq 2$, $t \in (0, 1]$ and $\vec{x} \in M^n$,

$$H_n(\vec{x}) \geq H(R_t(\vec{x})) - t - \frac{C}{nt^{\frac{m}{2}-1}}.$$

The terms $\frac{1}{8\pi} \log(t) - C$ and $-1/t^{m/2-1}$ come from the self-interaction of the regularized punctual charges while the term $-t$ comes from the negatively charged background. In the Euclidean setting, as there is no charged background, the $\frac{1}{8\pi} \log(t) - C$ and $-1/t^{m/2-1}$ terms arise from the self-interactions without potential and the $-t$ term arise from the regularization of the potential. The proof may be adapted to treat two-body interactions by the Green function of different Markov processes where the short-time asymptotic behavior is known.

To compare $H(R_t(\vec{x}))$ and $H_n(\vec{x})$ we will write, for $\vec{x} = (x_1, \dots, x_n) \in M^n$,

$$H(R_t(\vec{x})) = \frac{1}{n^2} \sum_{i < j} \int_{M \times M} G(\alpha, \beta) d\mu_{x_i}^t(\alpha) d\mu_{x_j}^t(\beta) + \frac{1}{2n^2} \sum_{i=1}^n \int_{M \times M} G(\alpha, \beta) d\mu_{x_i}^t(\alpha) d\mu_{x_i}^t(\beta).$$

Let us define

$$\begin{aligned} G_t(x, y) &= \int_{M \times M} G(\alpha, \beta) d\mu_x^t(\alpha) d\mu_y^t(\beta) \\ &= \int_{M \times M} G(\alpha, \beta) p_t(x, \alpha) d\pi(\alpha) p_t(y, \beta) d\pi(\beta). \end{aligned}$$

Then we may write

$$H(R_t(\vec{x})) = \frac{1}{n^2} \sum_{i < j} G_t(x_i, x_j) + \frac{1}{2n^2} \sum_{i=1}^n G_t(x_i, x_i).$$

So we want to compare G_t and G . The idea we shall use is that if G is the kernel of the operator \bar{G} and p_t is the kernel of the operator \bar{P}_t then G_t is the kernel of the operator $\bar{P}_t \bar{G} \bar{P}_t$. But using the eigenvector decomposition we can see that

$$\bar{G} = \int_0^\infty (\bar{P}_s - e_0 \otimes e_0^*) ds \quad (4.3)$$

where e_0 is the eigenvector of eigenvalue 0, i.e. the constant function equal to one. Then

$$\bar{P}_t \bar{G} \bar{P}_t = \int_0^\infty (\bar{P}_{2t+s} - e_0 \otimes e_0^*) ds \quad (4.4)$$

where we have used the semigroup property of $t \mapsto \bar{P}_t$, the fact that $\bar{P}_t e_0 = e_0$ and $\bar{P}_t^* = \bar{P}_t$. Notice that this representation can also be obtained when G is the Green function of different Markov processes.

We will prove the previous idea in a somehow different but very related way. We begin by proving the analogue of (4.3).

Proposition 4.5 (Integral representation of the Green function). *For every pair of different points $x, y \in M$ the function $t \mapsto p_t(x, y) - 1$ is integrable. For every $x \in M$ the negative part of the function $t \mapsto p_t(x, x) - 1$ is integrable. Moreover, we have the following integral representation of the Green function. For every $x, y \in M$*

$$G(x, y) = \int_0^\infty (p_t(x, y) - 1) dt.$$

Proof. To prove the integrability of $t \mapsto p_t(x, y) - 1$ we will need to know the behavior of p_t for large and short t . For the large-time behavior we have the following result.

Lemma 4.6 (Large-time behavior). *There exists $\lambda > 0$ such that for every $T > 0$, $s \geq 0$ and $x, y \in M$*

$$|p_{T+s}(x, y) - 1| \leq e^{-\lambda s} \sqrt{|p_T(x, x) - 1| |p_T(y, y) - 1|}. \quad (4.5)$$

Proof. We follow the same arguments as in the proof of [7, Corollary 3.7]. By the semigroup property, the symmetry of p_t and the Cauchy-Schwarz inequality we get

$$\begin{aligned} |p_{T+s}(x, y) - 1| &= \left| \int_M \left(p_{\frac{T+s}{2}}(x, z) - 1 \right) \left(p_{\frac{T+s}{2}}(z, y) - 1 \right) d\pi(z) \right| \\ &\leq \left\| p_{\frac{T+s}{2}}(x, \cdot) - 1 \right\|_{L^2} \left\| p_{\frac{T+s}{2}}(y, \cdot) - 1 \right\|_{L^2}. \end{aligned} \quad (4.6)$$

If λ is the first strictly positive eigenvalue of $-\Delta$ and if $f \in L^2(M)$ we get

$$\left\| \int_M (p_{\frac{s}{2}}(\cdot, z) - 1) f(z) d\pi(z) \right\|_{L^2} \leq e^{-\lambda \frac{s}{2}} \left\| f - \int_M f d\pi \right\|_{L^2}.$$

If we choose $f = p_{\frac{T}{2}}(x, \cdot) - 1$ we obtain

$$\left\| p_{\frac{T+s}{2}}(x, \cdot) - 1 \right\|_{L^2} \leq e^{-\lambda \frac{s}{2}} \left\| p_{\frac{T}{2}}(x, \cdot) - 1 \right\|_{L^2} = e^{-\lambda \frac{s}{2}} \sqrt{p_T(x, x) - 1} \quad (4.7)$$

where we have used the semigroup property for the last equality. Similarly, we get

$$\left\| p_{\frac{T+s}{2}}(y, \cdot) - 1 \right\|_{L^2} \leq e^{-\lambda \frac{s}{2}} \sqrt{p_T(y, y) - 1}. \quad (4.8)$$

By (4.6), (4.7) and (4.8) we may conclude (4.5). \square

For the short-time behavior, [8, Theorem 5.3.4] implies the following lemma.

Lemma 4.7 (Short-time behavior). *Let m be the dimension of M . Then there exist two positive constants C_1 and C_2 such that for every $t \in (0, 1)$ and $x, y \in M$ we have*

$$\frac{C_1}{t^{\frac{m}{2}}} e^{-\frac{d_g(x, y)^2}{4t}} \leq p_t(x, y) \leq \frac{C_2}{t^{m-\frac{1}{2}}} e^{-\frac{d_g(x, y)^2}{4t}}.$$

The integrability of $t \mapsto p_t(x, y) - 1$ when $x \neq y$ and the fact that $\int_0^\infty (p_t(x, x) - 1) dt = \infty$ for every $x \in M$ can be obtained from Lemma 4.7 and Lemma 4.6.

Using Lemma 4.6 and the dominated convergence theorem we obtain the continuity of the function $(x, y) \mapsto \int_1^\infty (p_t(x, y) - 1) dt$ at any $(x, y) \in M \times M$. By the dominated convergence theorem and Lemma 4.7 we obtain the continuity of the function given by $(x, y) \mapsto \int_0^1 (p_t(x, y) - 1) dt$ for $x \neq y$. Using Fatou's lemma we obtain the continuity of $(x, y) \mapsto \int_0^1 (p_t(x, y) - 1) dt$ at (x, y) such that $x = y$. So, we get that the function $K : M \times M \rightarrow (-\infty, \infty]$ defined by

$$K(x, y) = \int_0^\infty (p_t(x, y) - 1) dt$$

is continuous. The following lemma assures that $K(x, \cdot)$ is integrable for every $x \in M$.

Lemma 4.8 (Global integrability). *For every $x \in M$*

$$\int_0^\infty \int_M |p_t(x, y) - 1| d\pi(y) dt < \infty.$$

Proof. Take $T > 0$. By Lemma 4.6 we obtain that

$$\int_T^\infty \int_M |p_t(x, y) - 1| d\pi(y) dt < \infty.$$

On the other hand we have

$$\int_0^T \int_M |p_t(x, y) - 1| d\pi(y) dt \leq \int_0^T \int_M (p_t(x, y) + 1) d\pi(y) dt = 2T < \infty. \quad \square$$

Let $0 = \lambda_0 < \lambda_1 \leq \lambda_2 \leq \dots$ be the sequence of eigenvalues of $-\Delta$ and e_0, e_1, e_2, \dots the sequence of respective eigenfunctions. Then, for every $\psi \in C^\infty(M)$

$$\sum_{n=0}^\infty \exp(-\lambda_n t) |\langle e_n, \psi \rangle|^2 = \langle \psi, e^{t\Delta} \psi \rangle = \int_{M \times M} \psi(x) p_t(x, y) \psi(y) d\pi(x) d\pi(y).$$

Equivalently, we have

$$\sum_{n=1}^{\infty} \exp(-\lambda_n t) |\langle e_n, \psi \rangle|^2 = \int_{M \times M} \psi(x) (p_t(x, y) - 1) \psi(y) d\pi(x) d\pi(y)$$

and integrating in t from zero to infinity we obtain

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} |\langle e_n, \psi \rangle|^2 = \int_{M \times M} \psi(x) K(x, y) \psi(y) d\pi(x) d\pi(y).$$

By a polarization identity we have that, for every $\phi, \psi \in C^\infty(M)$,

$$\sum_{n=1}^{\infty} \frac{1}{\lambda_n} \langle \psi, e_n \rangle \langle e_n, \phi \rangle = \int_{M \times M} \psi(x) K(x, y) \phi(y) d\pi(x) d\pi(y).$$

Taking $\phi = \Delta \alpha$ we get

$$\langle \psi, \alpha \rangle - \int_M \psi d\pi \int_M \alpha d\pi = \sum_{n=1}^{\infty} \langle \psi, e_n \rangle \langle e_n, \alpha \rangle = \int_{M \times M} \psi(x) K(x, y) \Delta \alpha(y) d\pi(x) d\pi(y).$$

By definition of the Green function we know that $\int_M G(x, y) \Delta \alpha(y) d\pi(y) = -\alpha(x) + \int_M \alpha d\pi$ and thus

$$\int_{M \times M} \psi(x) G(x, y) \Delta \alpha(y) d\pi(x) d\pi(y) = \int_{M \times M} \psi(x) K(x, y) \Delta \alpha(y) d\pi(x) d\pi(y).$$

As $\int_M K(x, y) d\pi(y) = 0 = \int_M G(x, y) d\pi(y)$ and by the continuity of K and G we obtain $G(x, y) = K(x, y)$ for every $x, y \in M$. \square

Now we will state and prove (4.4).

Proposition 4.9 (Integral representation of the regularized Green function). *For every $t > 0$ and $x, y \in M$*

$$G_t(x, y) = \int_{2t}^{\infty} (p_s(x, y) - 1) ds.$$

Proof. Take the time (i.e. with respect to t) derivative (denoted by a dot above the function)

$$\begin{aligned} \dot{G}_t(x, y) &= \int_{M \times M} \dot{p}_t(x, \alpha) G(\alpha, \beta) p_t(y, \beta) d\pi(\alpha) d\pi(\beta) \\ &\quad + \int_{M \times M} p_t(x, \alpha) G(\alpha, \beta) \dot{p}_t(y, \beta) d\pi(\alpha) d\pi(\beta). \end{aligned}$$

We will study the first term of the sum (the second being analogous).

$$\begin{aligned} &\int_{M \times M} \dot{p}_t(x, \alpha) G(\alpha, \beta) p_t(y, \beta) d\pi(\alpha) d\pi(\beta) \\ &= \int_{M \times M} \Delta_\alpha p_t(x, \alpha) G(\alpha, \beta) p_t(y, \beta) d\pi(\alpha) d\pi(\beta) \\ &= \int_M \left(\int_M \Delta_\alpha p_t(x, \alpha) G(\alpha, \beta) d\pi(\alpha) \right) p_t(y, \beta) d\pi(\beta) \\ &= \int_M \left(\int_M p_t(x, \alpha) \Delta_\alpha G(\alpha, \beta) d\pi(\alpha) \right) p_t(y, \beta) d\pi(\beta) \\ &= \int_M \left(\int_M p_t(x, \alpha) (-\delta_\beta(\alpha) + 1) d\pi(\alpha) \right) p_t(y, \beta) d\pi(\beta) \\ &= \int_M (-p_t(x, \beta) + 1) p_t(y, \beta) d\pi(\beta) \\ &= -p_{2t}(x, y) + 1 \end{aligned}$$

where in the last line we have used the symmetry and the semigroup property of p . Using again the symmetry of p we get

$$\dot{G}_t(x, y) = -2p_{2t}(x, y) + 2,$$

and by integrating we obtain

$$G_t(x, y) - G_s(x, y) = \int_s^t (-2p_{2u}(x, y) + 2) du = \int_{2s}^{2t} (-p_s(x, y) + 1) ds$$

for every $0 < s < t < \infty$. As a consequence of the uniform convergence of Proposition 4.1 we can see that μ_x^t and μ_y^t defined in (4.1) converge to π as t goes to infinity. Fix any $T > 0$. As $G_{T+s}(x, y) = \int_{M \times M} G_T(\alpha, \beta) d\mu_x^s(\alpha) d\mu_y^s(\beta)$ for any $s > 0$ and as G_T is continuous we obtain $\lim_{t \rightarrow \infty} G_t(x, y) = \int_{M \times M} G_T(x, y) d\pi(x) d\pi(y) = 0$ and then

$$G_t(x, y) = \int_{2t}^{\infty} (p_s(x, y) - 1) ds. \quad \square$$

Using Proposition 4.5 and 4.9 we conclude the following inequality. We can find an analogous result in [10, Lemma 5.2].

Corollary 4.10 (Off-diagonal behavior). *For every $n \geq 2$, $t > 0$ and $(x_1, \dots, x_n) \in M^n$*

$$\sum_{i < j} G(x_i, x_j) \geq \sum_{i < j} G_t(x_i, x_j) - t n^2.$$

Proof. As the heat kernel is non-negative, by Proposition 4.5 and 4.9 we have that, for every $x, y \in M$,

$$G(x, y) - G_t(x, y) = \int_0^{2t} (p_s(x, y) - 1) ds \geq -2t.$$

Then, if $(x_1, \dots, x_n) \in M^n$,

$$\sum_{i < j} G(x_i, x_j) \geq \sum_{i < j} G_t(x_i, x_j) - t n(n-1) \geq \sum_{i < j} G_t(x_i, x_j) - t n^2. \quad \square$$

What is left to understand is $\sum_{i=1}^n G_t(x_i, x_i)$. This will be achieved using Proposition 4.9 and the short-time asymptotic expansion of the heat kernel. A particular case is mentioned in [10, Lemma 5.3].

Proposition 4.11 (Diagonal behavior). *Let m be the dimension of M . If $m = 2$ there exists a constant $C > 0$ such that for every $t \in (0, 1]$ and $x \in M$*

$$G_t(x, x) \leq -\frac{1}{4\pi} \log(t) + C.$$

If $m > 2$ there exists a constant $C > 0$ such that for every $t \in (0, 1]$ and $x \in M$

$$G_t(x, x) \leq \frac{C}{t^{\frac{m}{2}-1}}.$$

Proof. By the asymptotic expansion of the heat kernel (see for instance [5, Chapter VI.4]) we have that there exists a constant $\tilde{C} > 0$ (independent of x and t) such that, for $t \leq 1$,

$$\left| p_t(x, x) - \frac{1}{(4\pi t)^{\frac{m}{2}}} \right| \leq \tilde{C} t^{-\frac{m}{2}+1}.$$

Then,

$$p_t(x, x) \leq \frac{1}{(4\pi t)^{\frac{m}{2}}} + \tilde{C} t^{-\frac{m}{2}+1}. \quad (4.9)$$

We know by Proposition 4.9 that

$$\begin{aligned}
 G_t(x, x) &= \int_{2t}^{\infty} (p_s(x, x) - 1) ds \\
 &= \int_{2t}^2 (p_s(x, x) - 1) ds + \int_2^{\infty} (p_s(x, x) - 1) ds \\
 &\leq \int_{2t}^2 \left[\frac{1}{(4\pi s)^{\frac{m}{2}}} + \tilde{C} s^{-\frac{m}{2}+1} - 1 \right] ds + \int_2^{\infty} (p_s(x, x) - 1) ds \\
 &= \int_{2t}^2 \left[\frac{1}{(4\pi s)^{\frac{m}{2}}} + \tilde{C} s^{-\frac{m}{2}+1} \right] ds + G_2(x, x).
 \end{aligned}$$

In the case $m = 2$ we obtain that, for $t \in (0, 1]$,

$$G_t(x, x) \leq -\frac{1}{4\pi} \log(t) + C$$

where C is $2\tilde{C}$ plus a bound for $G_2(x, x)$ independent of x . In the case $m > 2$ we use that $s^{-m/2+1} \leq 2s^{-m/2}$ for $s \in (0, 1]$ and that $G_2(x, x)$ is bounded from above to obtain a constant C such that, for $t \in (0, 1]$,

$$G_t(x, x) \leq \frac{C}{t^{\frac{m}{2}-1}}. \quad \square$$

Knowing the diagonal and off-diagonal behavior of the regularized Green function we can proceed to prove Theorem 4.4.

Proof of Theorem 4.4. Take $\vec{x} = (x_1, \dots, x_n) \in M^n$. Then if $m = 2$ we have

$$\begin{aligned}
 H_n(\vec{x}) &\geq \frac{1}{n^2} \sum_{i < j} G_t(x_i, x_j) - t \\
 &\geq \frac{1}{n^2} \sum_{i < j} G_t(x_i, x_j) - t + \frac{1}{2n^2} \sum_{i=1}^n G_t(x_i, x_i) + \frac{1}{8\pi n} \log(t) - \frac{1}{2n} C \\
 &= H(R_t(\vec{x})) - t + \frac{1}{8\pi n} \log(t) - \frac{1}{2n} C
 \end{aligned}$$

where we have used Corollary 4.10 and Proposition 4.11. If $m > 2$ we proceed in the same way to get

$$\begin{aligned}
 H_n(\vec{x}) &\geq \frac{1}{n^2} \sum_{i < j} G_t(x_i, x_j) - t \\
 &\geq \frac{1}{n^2} \sum_{i < j} G_t(x_i, x_j) - t + \frac{1}{2n^2} \sum_{i=1}^n G_t(x_i, x_i) + \frac{C}{2nt^{\frac{m}{2}-1}} \\
 &= H(R_t(\vec{x})) - t + \frac{C}{2nt^{\frac{m}{2}-1}}. \quad \square
 \end{aligned}$$

Remark 4.12 (Euclidean setting). Let us give a quick explanation of the regularization of the energy in the Euclidean case. Define the two-body interaction G by

$$G(x, y) = \begin{cases} -\log|x-y| & \text{if } m = 2 \\ |x-y|^{2-m} & \text{if } m > 2 \end{cases}.$$

Suppose μ is a radial probability measure on \mathbb{R}^m of finite energy, i.e. such that $\int_{\mathbb{R}^m \times \mathbb{R}^m} |G(x, y)| d\mu(x) d\mu(y) < \infty$. For $\varepsilon > 0$ define $S_\varepsilon : \mathbb{R}^m \rightarrow \mathbb{R}^m$ by

$$S_\varepsilon(x) = \varepsilon x$$

and for $x \in \mathbb{R}^m$ define $T_x : \mathbb{R}^m \rightarrow \mathbb{R}^m$ by

$$T_x(\alpha) = \alpha + x.$$

The regularization of the punctual charge at $x \in \mathbb{R}^m$ will be $\mu_x^\varepsilon = (T_x \circ S_\varepsilon)_* \mu$ where the subindex $*$ is used to denote the image measure. Define the two-body regularized interaction G_ε by

$$G_\varepsilon(x, y) = \int_{\mathbb{R}^m \times \mathbb{R}^m} G(\alpha, \beta) d\mu_x^\varepsilon(\alpha) d\mu_y^\varepsilon(\beta).$$

The analogue of Corollary 4.10 would be

$$\sum_{i < j} G(x_i, x_j) \geq \sum_{i < j} G_\varepsilon(x_i, x_j)$$

which is a consequence of the superharmonicity of $G(x, \cdot)$. The analogue of Proposition 4.11 would be

$$G_\varepsilon(x, x) = -\log \varepsilon - \int_{\mathbb{R}^m \times \mathbb{R}^m} \log |\alpha - \beta| d\mu(\alpha) d\mu(\beta)$$

when $m = 2$ and

$$G_\varepsilon(x, x) = \varepsilon^{2-m} \int_{\mathbb{R}^m \times \mathbb{R}^m} |\alpha - \beta|^{2-m} d\mu(\alpha) d\mu(\beta)$$

when $m > 2$. This is a straightforward application of the change-of-variables formula. Finally, if we define $R_\varepsilon(x_1, \dots, x_n) = \frac{1}{n} \sum_{i=1}^n \mu_{x_i}^\varepsilon$, the analogue of Proposition 4.2 would be

$$W_1(R_\varepsilon(\vec{x}), i_n(\vec{x})) \leq \varepsilon \int_{\mathbb{R}^m} |y| d\mu(y).$$

Having acquired all the tools to apply Theorem 2.2 to the case of a Coulomb gas on a compact Riemannian manifold, the next section will be devoted to prove the main theorem and its almost immediate extension.

5 Proof of the concentration inequality for Coulomb gases

Proof of Theorem 1.1. First, we notice that $e_n = \int_M H_n d\mu_{\text{eq}} = \frac{n-1}{n} e = 0$. To use Theorem 2.2 we define

$$f(r) = \frac{r^2}{2} \text{ and } R = R_t \text{ for } t = n^{-\frac{2}{m}}.$$

In this case, Proposition 4.2 tells us that $W_1(R(\vec{x}), i_n(\vec{x})) \leq C/n^{1/m}$ for some $C > 0$ independent of \vec{x} and n . This may be considered as the natural choice since $1/n^{1/m}$ is the ‘closest’ a fixed probability measure absolutely continuous with respect to π can get to an arbitrary empirical measure of n points.

If $m = 2$, by Theorem 4.4 and Proposition 4.2, we have that there exists a constant $\tilde{C} > 0$ such that

$$H_n(\vec{x}) \geq H(R(\vec{x})) - \frac{1}{8\pi n} \log(n) - \frac{\tilde{C}}{n}$$

$$W_1(R(\vec{x}), i_n(\vec{x})) \leq \frac{\tilde{C}}{\sqrt{n}}$$

for every $\vec{x} \in M^n$ and $n \geq 2$ so we can apply Theorem 2.2 to obtain the desired result with $C = \frac{\tilde{C}^2}{2} + \tilde{C}$. Similarly, if $m > 2$, by Theorem 4.4 and Proposition 4.2, we have that there exists a constant $\tilde{C} > 0$ such that

$$H_n(\vec{x}) \geq H(R(\vec{x})) - \frac{\tilde{C}}{n^{\frac{2}{m}}}$$

$$W_1(R(\vec{x}), i_n(\vec{x})) \leq \frac{\tilde{C}}{n^{\frac{1}{m}}}$$

for every $\vec{x} \in M^n$ and $n \geq 2$ so we can apply Theorem 2.2 to obtain the desired result with $C = \frac{\tilde{C}^2}{2} + \tilde{C}$. \square

Finally we present the proof of Theorem 1.2.

Proof of Theorem 1.2. To apply Theorem 2.2 we notice that Assumption A is satisfied by $f(r) = \frac{r^2}{2}$. Indeed, Theorem 3.1 is still true for this new H except for the characterization of the minimizer. In particular, H has a unique minimizer. By a calculation we can see that $e - e_n = \frac{1}{2n} \int_{M \times M} G(x, y) d\mu_{eq}(x) d\mu_{eq}(y)$ which is of order $\frac{1}{n}$ and will be absorbed by the constant C . To meet the hypotheses of Theorem 2.2, we need to compare

$$\frac{1}{n} \sum_{i=1}^n V(x_i) \quad \text{and} \quad \frac{1}{n} \sum_{i=1}^n \int_M V d\mu_{x_i}^t.$$

By using the relation

$$\mathbb{E}[V(X_t)] = V(x) + \int_0^t \mathbb{E}[\Delta f(X_s)] ds$$

where X_t is the Markov process with generator Δ starting at x we obtain

$$|\mathbb{E}[V(X_t)] - V(x)| \leq \hat{C}t$$

where \hat{C} is some upper bound to ΔV and thus

$$\left| \frac{1}{n} \sum_{i=1}^n \int_M V d\mu_{x_i}^t - \frac{1}{n} \sum_{i=1}^n V(x_i) \right| \leq \hat{C}t.$$

In conclusion, if we choose $R = R_{n - \frac{2}{m}}$, there still exists a constant $C > 0$ such that

$$H_n(\vec{x}) \geq H(R(\vec{x})) - \frac{1}{8\pi n} \log(n) - \frac{C}{n}$$

in dimension two and

$$H_n(\vec{x}) \geq H(R(\vec{x})) - \frac{C}{n^{\frac{2}{m}}}$$

in dimension $m > 2$ so that we can apply Theorem 2.2. \square

References

- [1] Thierry Aubin: *Some nonlinear problems in Riemannian geometry*. Springer Monographs in Mathematics. Springer-Verlag, Berlin, 1998. xviii+395 pp. MR-1636569
- [2] Daniel Azagra, Juan Ferrera Cuesta, Fernando López-Mesas and Yenny Rangel: Smooth approximation of Lipschitz functions on Riemannian manifolds. *Journal of Mathematical Analysis and Applications* **326**, (2007), no. 2, 1370–1378. MR-2280987
- [3] Carlos Beltrán, Nuria Corral and Juan G. Criado del Rey: Discrete and continuous Green energy on compact manifolds. *Journal of Approximation Theory* **237**, (2019), 160–185. MR-3868631
- [4] Djalil Chafaï, Adrien Hardy, and Mylène Maïda: Concentration for Coulomb gases and Coulomb transport inequalities. *Journal of Functional Analysis* **275**, (2018), no. 6, 1447–1483. MR-3820329
- [5] Isaac Chavel: *Eigenvalues in Riemannian geometry*. Academic Press, Inc., Orlando, FL, 1984. xiv+362 pp. MR-0768584

- [6] David García-Zelada: A large deviation principle for empirical measures on Polish spaces: Application to singular Gibbs measures on manifolds. *To appear in Annales de l'Institut Henri Poincaré*, arXiv:1703.02680
- [7] Alexander Grigor'yan: *Heat kernels on weighted manifolds and applications. The ubiquitous heat kernel*, 93–191, Contemporary Mathematics **398**, American Mathematical Society, Providence, RI, 2006. MR-2218016
- [8] Elton P. Hsu: *Stochastic analysis on manifolds*. Graduate Studies in Mathematics **38**. American Mathematical Society, Providence, RI, 2002. xiv+281 pp. MR-1882015
- [9] Manjunath Ramanatha Krishnapur: *Zeros of random analytic functions*. Ph. D. Thesis, U. C. Berkley, 2006. 76 pp. MR-2709142
- [10] Serge Lang: *Introduction to Arakelov theory*. Springer-Verlag, New York, 1988. x+187 pp. MR-0969124
- [11] Michel Ledoux: On optimal matching of Gaussian samples. *Zapiski Nauchnykh Seminarov POMI* **457**, (2017), 226–264. MR-3723584
- [12] Elliott Lieb and Michael Loss: *Analysis. Second edition*. Graduate Studies in Mathematics **14**. American Mathematical Society, Providence, RI, 2001. xxii+346 pp. MR-1817225
- [13] Mylène Maïda and Édouard Maurel-Segala: Free transport-entropy inequalities for non-convex potentials and application to concentration for random matrices. *Probability Theory and Related Fields* **159**, (2014), no. 1-2, 329–356. MR-3201924
- [14] Nicolas Rougerie and Sylvia Serfaty: Higher-dimensional Coulomb gases and renormalized energy functionals. *Communications on Pure and Applied Mathematics* **69**, (2016), no. 3, 519–605. MR-3455593
- [15] Sylvia Serfaty: Systems of Points with Coulomb Interactions. *Proceedings of the International Congress of Mathematicians 2018* **1**, 935–978, Rio de Janeiro, 2018. arXiv:1712.04095 MR-3837028
- [16] Cédric Villani: *Topics in optimal transportation*. Graduate Studies in Mathematics **58**. American Mathematical Society, Providence, RI, 2003. xvi+370 pp. MR-1964483