REGULARITY PROPERTIES OF THE DONSKER-VARADHAN RATE FUNCTIONAL FOR NON-REVERSIBLE DIFFUSIONS AND RANDOM EVOLUTIONS

ROSS G. PINSKY

ABSTRACT. The Donsker-Varadhan rate functional $I(\mu)$ for Markov processes has a simple form in the case that the generator of the process is self-adjoint, or equivalently, that the Markov process is reversible. In particular, in the case of a reversible diffusion generated by a second-order elliptic operator L on a compact manifold, this allows one to give a simple necessary and sufficient criterion on the measure μ in order that $I(\mu) < \infty$, and it also shows that $I(\mu)$ is continuous as a function of the coefficients of L in the sup-norm topology. In this paper, we first show that the same criterion for finiteness holds for non-reversible diffusions, and then show that $I(\mu)$ is locally Lipschitz continuous as a function of the coefficients of the generator L in the sup-norm topology. We then prove similar results in the setting of random evolutions.

1. INTRODUCTION

The Donsker-Varadhan rate functional $I(\mu)$ plays a decisive role in describing the large deviations of the occupation measure of a sufficiently well-behaved Markov process. In particular, the probability that the n-step occuapation measure for the process is in a small neighborhood of a probability measure μ is of the order $\exp(-nI(\mu))$ (in the logarithmic sense). It is thus important to know when $I(\mu) < \infty$ and when $I(\mu) = \infty$, since in the latter case the above probability is negligibly small. It follows from the classical Donsker-Varadhan papers that for reversible diffusions, $I(\mu) < \infty$ if and only if μ possesses a density in an appropriate $W^{1,2}$ space. In this paper, we prove that this same condition holds for nonreversible diffusions and also in the more general framework of random evolution equations (randomly switching diffusions). We also address the question of the stability of the rate functional with respect to the coefficients of the generator of the diffusion or random evolution. It is easy to see from the definitions that the rate functional is lower semi-continuous with respect to the coefficients. In fact, we show that the rate functional depends Lipschitz continuously on the coefficients of the generator.

We point out that these finiteness and continuity results are needed in [4] for the study of the long time behavior of the slow motion in fully coupled averaging with random evolutions as the fast motion.

2. Diffusions

Let M be a d-dimensional, compact Reimannian manifold and let L be a second-order elliptic operator written in local coordinates as

$$L = \frac{1}{2}\nabla \cdot a\nabla + b\nabla,$$

where $a = a(x)$ is a positive definite $d \times d$ matrix with entries in $C^1(M)$ and $b = b(x)$ is a continuous d-vector. Let $X = X(t)$ denote the diffusion process on M which solves the martingale problem for L, and denote the corresponding probabilities for the process starting at $x \in M$ by P_x. Let $l^t = l^t_X$ denote the occupation measure up to time t for the path X; that is, $l^t_X(B) = \int_0^t \chi_B(X(s))ds$, for $B \subset M$.

Let $\mathcal{P}(M)$ denote the space of probability measures on M with the topology of weak convergence. Define the I-function by

$$I(\mu) = -\inf_{u \in C^2(M), u>0} \int_M \frac{Lu}{u} d\mu. \tag{2.1}$$

As is well-known, the Donsker-Varahdan large deviations theory [1] states that

$$\begin{aligned} &\liminf_{t\to\infty} \frac{1}{t} \log P_x(\frac{1}{t} l^t \in G) \ge -\inf_{\mu \in G} I(\mu), \text{ for open } G \subset \mathcal{P}(M);\\ &\limsup_{t\to\infty} \frac{1}{t} \log P_x(\frac{1}{t} l^t \in C) \le -\inf_{\mu \in C} I(\mu), \text{ for closed } C \subset \mathcal{P}(M). \end{aligned} \tag{2.2}$$

(In fact, in [1], the I-function is defined with the restriction $u \in C^2(M)$ appearing in the infimum above replaced by the restriction that u be in the domain of the generator of the Markov process $X(t)$. The exact domain is generally not known. The passage to $u \in C^2(M)$, which allows for the possibility of explicit calculations, was carried out in [5].)

If the drift vector b is of the form $b = a\nabla Q$, for some C^1-function Q, then (and only then) L can be realized as a self adjoint operator—on the space $L^2(M, \mu_{\text{rev}})$, where $d\mu_{\text{rev}} = \exp(2Q)dx$; equivalently, the diffusion process $X(t)$ is reversible with respect to the measure μ_{rev}. In the self-adjoint case, the I-function was shown in [1] to have a succinct form:

(2.3)

$$I(\mu) = \begin{cases} ||(-L)^{\frac{1}{2}}(\frac{d\mu}{d\mu_{\text{rev}}})^{\frac{1}{2}}||^2_{2,\mu_{\text{rev}}}, \text{ if } \frac{d\mu}{d\mu_{\text{rev}}} \text{ exists and } (\frac{d\mu}{d\mu_{\text{rev}}})^{\frac{1}{2}} \text{ belongs to} \\ \qquad\qquad \text{the domain of } (-L)^{\frac{1}{2}}; \\ \infty, \text{ otherwise,} \end{cases}$$

where $|| \cdot ||_{2,\mu_{\text{rev}}}$ denotes the norm on $L^2(M, \mu_{\text{rev}})$. When written out explicitly, this becomes

(2.4)

$$I(\mu) = \begin{cases} \frac{1}{8}\int_M \frac{\nabla\phi a \nabla\phi}{\phi}\exp(2Q)dx = \frac{1}{2}\int_M (\nabla g a \nabla g)\exp(2Q)dx, \text{ if} \\ \qquad\qquad \phi \equiv \frac{d\mu}{d\mu_{\text{rev}}} \text{ exists and } g = \phi^{\frac{1}{2}}; \\ \infty, \text{ otherwise.} \end{cases}$$

Note in particular that (2.4) shows that in the self-adjoint case, $I(\mu)$ is finite if and only if μ possesses a density $\frac{d\mu}{dx}$ and $(\frac{d\mu}{dx})^{\frac{1}{2}} \in W^{1,2}(M)$. (We write $W^{1,2}(M)$ for $W^{1,2}(M, dx)$.) We will prove that this is also true in the non-self-adjoint case.

In order to emphasize the dependence of the I-function on the operator L, or equivalently, on a and b, from now on we will use the notation $I_{a,b}(\mu)$ in place of $I(\mu)$.

Theorem 1. *$I_{a,b}(\mu)<\infty$ if and only if μ possesses a density $\frac{d\mu}{dx}$ and $(\frac{d\mu}{dx})^{\frac{1}{2}}\in W^{1,2}(M)$.*

Remark 1. As a complement to Theorem 1, we note that in [6] it was proved that if μ possesses a density $\frac{d\mu}{dx}$ and $g\equiv(\frac{d\mu}{dx})^{\frac{1}{2}}\in W^{1,2}(M)$, then there exist positive constants $c_i,=1,2,3,4$, depending on a and b, but not on μ, such that

$$c_1\int_M|\nabla g|^2dx-c_2\le I_{a,b}(\mu)\le c_3\int_M|\nabla g|^2dx+c_4.$$

Remark 2. It follows from (2.1) that $I_{a,b}$ is lower semi-continuous on $\mathcal{P}(M)$. It is clearly not continuous since any measure μ for which $(\frac{d\mu}{dx})^{\frac{1}{2}}\in W^{1,2}(M)$ can be approximated arbitrarily closely on $\mathcal{P}(M)$ by a discrete measure, and by Theorem 1, the I-function evaluated at a discrete measure is infinite.

It also follows from (2.1) that $I_{a,b}(\mu)$ is lower semi-continuous as a function of a and b in the sup-norm topology on M. However, equation (2.4) shows that in the self-adjoint case $I_{a,b}(\mu)$ is in fact continuous as a function of a and b in the sup-norm topology. We will show that this holds in the non-self-adjoint case as well, and we will give a bound on the modulus of continuity.

Assuming a priori that μ possesses a density and that $g\equiv(\frac{d\mu}{dx})^{\frac{1}{2}}\in W^{1,2}(M)$, $I(\mu)$ has the following relatively explicit form in the general case:

(2.5)
$$\begin{aligned}I_{a,b}(\mu)&=\frac{1}{2}\int_M(\frac{\nabla g}{g}-a^{-1}b)a(\frac{\nabla g}{g}-a^{-1}b)g^2dx\\&-\inf_{h\in C^2(M)}\frac{1}{2}\int_M(\nabla h-a^{-1}b)a(\nabla h-a^{-1}b)g^2dx,\ \text{if}\quad g\equiv(\frac{d\mu}{dx})^{\frac{1}{2}}\in W^{1,2}(M).\end{aligned}$$

(See [6] and [2]. In [6], $I_{a,b}(\mu)$ is evaluated for diffusions in bounded domains with oblique reflection at the boundary, with the same smoothness requirements on the coefficients a and b as stated here, and under the assumption that μ possesses a density. Ignoring the boundary terms there gives (2.5). In [2], $I_{a,b}(\mu)$ is evaluated for diffusions on R^d under the assumption that the coefficients a and b are C^∞ and under the assumption that μ is compactly supported with a C^∞-density.)

Let $||a||_{d\times d} \equiv \sup_{x\in M, 0\neq v\in R^d} \frac{|a(x)v|}{|v|}$.

Theorem 2. *Let $\mu \in \mathcal{P}(M)$ possess a density $\frac{d\mu}{dx}$ with $g \equiv (\frac{d\mu}{dx})^{\frac{1}{2}} \in W^{1,2}(M)$. Let*

$$D(a,b,\hat{a},\hat{b}) \equiv ||a^{-1}(\hat{a}-a)||_{d\times d} \int_M (ba^{-1}b) g^2 dx + \left(\int_M (\hat{b}-b)a^{-1}(\hat{b}-b) g^2 dx\right)^{\frac{1}{2}} \left(\int_M (ba^{-1}b) g^2 dx\right)^{\frac{1}{2}}.$$

Then

$$(2.6)\qquad |I_{a,b}(\mu) - I_{\hat{a},\hat{b}}(\mu)| \le \frac{1}{2} ||a^{-1}(a-\hat{a})||_{d\times d} \int_M (\nabla g a \nabla g) dx + \left(\int_M (\hat{b}-b)a^{-1}(\hat{b}-b) g^2 dx\right)^{\frac{1}{2}} \left(\int_M (\nabla g a \nabla g) dx\right)^{\frac{1}{2}} + 2\max\Big(D(a,b,\hat{a},\hat{b}), D(\hat{a},\hat{b},a,b)\Big).$$

Remark. Theorem 2 shows that for each $\mu \in \mathcal{P}(M)$ with $(\frac{d\mu}{dx})^{\frac{1}{2}} \in W^{1,2}(M)$, $I_{a,b}(\mu)$ is locally Lipschitz continuous as a function of a and b in the sup-norm topology on M. In fact, the theorem shows that for any positive number N, the family of functions $\{I_{a,b}(\mu) : g \equiv (\frac{d\mu}{dx})^{\frac{1}{2}} \in W^{1,2}(M),\ \int_M |\nabla g|^2 dx \le N\}$ is locally Lipschitz equicontinuous as a function of a and b in the sup-norm topology on M.

Theorem 1 is proved in section 4 and Theorem 2 is proved in section 5.

3. Random Evolutions

Consider the random evolution $X = X(t) = (Y(t), \sigma(t))$, where $Y(t) \in M$ and $\sigma(t) \in \{1, \cdots, m\}$, corresponding to the generator $\mathcal{L}$, which acts on smooth vector-valued functions $u = u(y) = (u_1(y), ..., u_m(y))$, $y \in M$, by the formula

$$(\mathcal{L}u)_k(y) = L_k u_k(y) + \lambda_k(y) \sum_{l=1}^m q_{kl}(y)(u_l(y) - u_k(y)),$$

where for each k, L_k satisfies the conditions satisfied by L in section 2, where for each $y \in M$, $q(y) = \{q_{kl}(y)\}_{k,l=1}^m$ is a stochastic matrix for a Markov chain on $\{1, \cdots, m\}$, and where $\lambda_k(y) > 0$. We assume that $q_{kl}(y)$ and $\lambda(y)$ are continuous on M. In words, the diffusion component $Y(t)$ runs like a L_k-diffusion whenever the discrete component $\sigma(t) = k$, and the discrete component $\sigma(t)$ jumps with the following intensities: $P(\sigma(t+h) = l|\sigma(t) = k, Y(t) = y) = \lambda_k(y)q_{kl}(y) + o(h)$, as $h \to 0$. (See, for example, [3, chapter 5.4].)

We need to guarantee that the Markov process X is irreducible. This is equivalent to guaranteeing that the discrete part σ of the Markov process X can get from any state k to any other state l. Of course, a sufficient condition for this is that for each $y \in M$, the stochastic matrix $q(y)$ generates an irreducible Markov chain; that is, for each $y \in M$ and each $k, l \in \{1, \cdots, m\}$, there exists a positive integer n such that $(q^n)_{kl}(y) > 0$. However, it is not hard to see that in fact one can make do with considerably less. It is enough to assume that for each $k, l \in \{1, \cdots, m\}$, either there exists a $y \in M$ such that $q_{kl}(y) > 0$, or else there exists a positive integer j, integers $i_1, \cdots, i_j \in \{1, \cdots, m\}$ and points $y_0, \cdots, y_j \in M$ such that $q_{k,i_1}(y_0) \cdots q_{i_j,l}(y_j) > 0$.

The above-defined random evolution is a Feller process with nice regularity properties; consequently, the Donsker-Varadhan theory applies [1]. Let $\mathcal{P}(M \times \{1, \cdots, m\})$ denote the space of probability measures on $M \times \{1, \cdots, m\}$ with the topology of weak convergence. A measure $\mu \in \mathcal{P}(M \times \{1, \cdots, m\})$ can be written as $\mu = (\mu_1, \cdots, \mu_m)$, where the μ_k are sub-probability measures on M and $\sum_{k=1}^m \mu_k(M) = 1$. The occupation measure $l^t = l_X^t$ is given by $l_X^t(B) = \int_0^t I_B(Y(s), \sigma(s))ds$, for $B \subset M \times \{1, \cdots, m\}$. The I-function is given by

(3.1)
$$I(\mu) = -\inf \sum_{k=1}^m \int_M \frac{(\mathcal{L}u)_k}{u_k} d\mu_k, \text{ for } \mu = (\mu_1, \cdots, \mu_m) \in \mathcal{P}(M \times \{1, \cdots, m\}),$$

where the infimum is over vector-valued functions $u = (u_1, \cdots, u_m)$ satisfying $u_k > 0$ and $u_k \in C^2(M)$, for $k \in \{1, \cdots, m\}$.

It is not hard to show that the random evolution is a reversible Markov process if and only if the drifts b_k are of the form $b_k = a_k \nabla Q$ (note that Q is independent of k) and also the stochastic matrices $q(y), y \in M$, correspond to reversible Markov chains, with a common reversible measure, independent of $y \in M$. (See [7], which treats the case that λ and q do not depend on y.) We will prove the analog of Theorem 1 for general, not-necessarily reversible random evolutions.

In order to emphasize the dependence on the coefficients, from now on we will use the notation $I_{a,b,\lambda,q}$, where $a = (a_1, \cdots, a_m)$, $b = (b_1, \cdots, b_m)$ $\lambda = (\lambda_1, \cdots, \lambda_m)$ and $q = \{q_{kl}\}_{k,l=1}^m$.

Theorem 3. *$I_{a,b,\lambda,q}(\mu) < \infty$ if and only if $\mu = (\mu_1, \cdots, \mu_m) \in \mathcal{P}(M \times \{1, \cdots, m\})$ possesses a density $(\frac{d\mu_1}{dy}, \cdots, \frac{d\mu_m}{dy})$ with $g_k \equiv (\frac{d\mu_k}{dy})^{\frac{1}{2}} \in W^{1,2}(M)$, for $k = 1, \cdots, m$. Furthermore, if μ possesses a density $\frac{d\mu}{dy} = (\frac{d\mu_1}{dy}, \cdots, \frac{d\mu_m}{dy})$ and $(\frac{d\mu_k}{dy})^{\frac{1}{2}} \in W^{1,2}(M)$, $k = 1, \cdots, m$, then there exist positive constants $c_i, = 1, 2, 3, 4$, depending on a, b, λ and q, but not on μ, such that*

$$c_1 \sum_{k=1}^m \int_M |\nabla g_k|^2 dy - c_2 \le I_{a,b,\lambda,q}(\mu) \le c_3 \sum_{k=1}^m \int_M |\nabla g_k|^2 dy + c_4. \tag{3.2}$$

The same type of analysis in [6] leading to formula (2.5) for diffusions gives the following relatively explicit formula for the I-function for random evolutions:

(3.3)

$$I_{a,b,\lambda,q}(\mu) = \frac{1}{2} \sum_{k=1}^m \int_M (\frac{\nabla g_k}{g_k} - a_k^{-1} b_k) a_k (\frac{\nabla g_k}{g_k} - a_k^{-1} b_k) g_k^2 dy + \sum_{k=1}^m \int_M \lambda_k g_k^2 dy$$

$$- \inf_h \Big(\frac{1}{2} \sum_{k=1}^m \int_M (\nabla h_k - a_k^{-1} b_k) a_k (\nabla h_k - a_k^{-1} b_k) g_k^2 dy$$

$$+ \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l) dy \Big), \text{ if } g_k \equiv (\frac{d\mu_k}{dx})^{\frac{1}{2}} \in W^{1,2}(M),\ k = 1, \cdots, m,$$

where the infimum is over those functions $h = (h_1, h_2. \cdots, h_m)$ satisfying $h_k \in C^2(M)$, for $k \in \{1, 2, \cdots, m\}$.

Before continuing, we note briefly how the terms $\sum_{k=1}^m \int_M \lambda_k g_k^2 dy$ and $\sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l) dy$ enter into (3.3). In [6], one writes the test function u in (2.1) in the form $u = (g_\epsilon^2 + \delta)\exp(-h)$ for $\delta > 0$, where g_ϵ is a mollified version of g. One works with a mollified version because g might not be smooth, and one adds δ because g might not be bounded from 0. Eventually one lets $\epsilon \to 0$ and then $\delta \to 0$. In the random evolutions case, one proceeds analogously, defining $u_k = (g_{k;\epsilon}^2 + \delta)\exp(-h_k)$, where $g_{k;\epsilon}$ is a mollified version of g_k. Substituting this in (3.1), one sees how the above terms arise.

The next theorem is the analog of Theorem 2 for random evolutions.

Theorem 4. *Let $\mu = (\mu_1, \cdots, \mu_m) \in \mathcal{P}(M)$ possess a density $(\frac{d\mu_1}{dy}, \cdots, \frac{d\mu_m}{dy})$ with $g_k \equiv (\frac{d\mu_k}{dy})^{\frac{1}{2}} \in W^{1,2}(M)$, for $k = 1, \cdots, m$. Let*

$$D(a, b, \lambda, q, \hat{a}, \hat{b}, \hat{\lambda}, \hat{q}) =$$

$$\Big[\Big(\sup_{y \in M; k,l} \frac{\hat{\lambda}_k}{\lambda_k}(y)|\frac{\hat{q}_{kl}}{q_{kl}}(y) - 1| + \sup_{y \in M; k} |\frac{\hat{\lambda}_k}{\lambda_k}(y) - 1|\Big) \wedge \Big(\max_k ||a_k^{-1}(a_k - \hat{a}_k)||_{d\times d}\Big)\Big] \times$$

$$\Big(2\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy + \frac{m+1}{2}(\sup_{y \in M; k} \lambda_k(y))\Big)$$

$$+ \Big(4\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy + (m+1)(\sup_{y \in M; k} \lambda_k(y))\Big)^{\frac{1}{2}} \times \sum_{k=1}^m \Big(\int_M (b_k - \hat{b}_k) a_k^{-1} (b_k - \hat{b}_k) g_k^2 dy\Big)^{\frac{1}{2}},$$

with the convention that $\frac{\hat{q}_{kl}}{q_{kl}}(y) \equiv 1$ if $\hat{q}_{kl}(y) = q_{kl}(y) = 0$.

Then

$$
\begin{aligned}
&|I_{a,b,\lambda,q}(\mu)-I_{\hat a,\hat b,\hat\lambda,\hat q}(\mu)|\leq\\
&\frac{1}{2}\sum_{k=1}^m ||a_k^{-1}(a_k-\hat a_k)||_{d\times d}\int_M(\nabla g_k a_k\nabla g_k)dy\\
&+\sum_{k=1}^m\Big(\int_M(b_k-\hat b_k)a_k^{-1}(b_k-\hat b_k)g_k^2dy\Big)^{\frac{1}{2}}\Big(\int_M(\nabla g_k a_k\nabla g_k)dy\Big)^{\frac{1}{2}}\\
&+\sum_{k=1}^m\int_M|\lambda_k-\hat\lambda_k|g_k^2dy\\
&+\max\big(D(a,b,\lambda,q,\hat a,\hat b,\hat\lambda,\hat q),D(\hat a,\hat b,\hat\lambda,\hat q,a,b,\lambda,q)\big).
\end{aligned}
\tag{3.4}
$$

If $q(y)$ and $\hat q(y)$ are doubly stochastic for each $y\in M$, then the term $m+1$ appearing twice in D may be replaced by 2.

Remark. Theorem 4 shows that for each $\mu=(\mu_1,\cdots,\mu_m)\in\mathcal{P}(M)\times\{1,\cdots,m\}$ with $(\frac{d\mu_k}{dy})^{\frac{1}{2}}\in W^{1,2}(M)$, for $k=1,\cdots m$, $I_{a,b,\lambda,q}(\mu)$ is locally Lipschitz continuous as a function of a, b and λ in the sup-norm topology on M. It also shows that $I_{a,b,\lambda,q}(\mu)$ is locally Lipschitz continuous in q over the set $\{q=\{q_{kl}\}:\inf_{y\in M;k,l}q_{kl}(y)>0\}$. In fact, the theorem shows that for any positive number N, the family of functions $\{I_{a,b,\lambda,q}(\mu):g_k\equiv(\frac{d\mu_k}{dy})^{\frac{1}{2}}\in W^{1,2}(M),\ \int_M|\nabla g_k|^2dy\leq N,k=1,\cdots,m\}$ is locally Lipschitz equicontinuous as a function of a, b and λ in the sup-norm topology on M, and as a function of q over the set $\{q=\{q_{kl}\}:\inf_{y\in M;k,l}q_{kl}(y)>0\}$.

Theorem 3 is proved in section 6 and Theorem 4 is proved in section 7.

4. Proof of Theorem 1

Let

$$L_0=\frac{1}{2}\nabla\cdot a\nabla,$$

and let P_x^0 denote probabilities for the diffusion corresponding to L_0 and starting from $x\in M$. The operator L_0 is self-adjoint and the corresponding diffusion is reversible. Thus, in order to prove the theorem, it follows from (2.4) that it suffices to show that $I_{a,b}(\mu)=\infty$ whenever $I_{a,0}(\mu)=\infty$.

Note that from (2.2) and the lower semi-continuity of $I_{a,b}$, it follows that the condition $I_{a,b}(\mu)=\infty$ is equivalent to the condition

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log P_x(\frac{1}{t}l^t\in\bar B_{\frac{1}{n}}(\mu))=-\infty, \tag{4.1}$$

where $B_{\frac{1}{n}}(\mu)\subset\mathcal{P}(M)$ is the ball of radius ϵ centered at μ. Since, by assumption, $I_{a,0}(\mu)=\infty$, we have

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log P^0_x(\frac{1}{t}l^t\in\bar B_{\frac{1}{n}}(\mu))=-\infty. \tag{4.2}$$

Using the Girsanov transformation, we have

(4.3)

$$P_x(\frac{1}{t}l^t\in\bar B_{\frac{1}{n}}(\mu))=$$
$$E^0_x\left(\exp\left(\int_0^t a^{-1}b(X(s))d\bar X(s)-\frac{1}{2}\int_0^t(ba^{-1}b)(X(s))ds\right);\frac{1}{t}l^t_X\in\bar B_{\frac{1}{n}}(\mu)\right),$$

where $\bar X(s)=X(s)-\frac{1}{2}\int_0^s(\nabla\cdot a)(X(r))dr$. Applying Cauchy-Schwarz to (4.3) gives

(4.4)

$$P_x(\frac{1}{t}l^t\in\bar B_{\frac{1}{n}}(\mu))\le$$
$$\left(E^0_x(\exp\left(2\int_0^t a^{-1}b(X(s))d\bar X(s)-\int_0^t(ba^{-1}b)(X(s))ds\right)\right)^{\frac{1}{2}}\left(P^0_x(\frac{1}{t}l^t\in\bar B_{\frac{1}{n}}(\mu))\right)^{\frac{1}{2}}.$$

Since for $q>0$, $\exp\left(q\int_0^t a^{-1}b(X(s))d\bar X(s)-\frac{q^2}{2}\int_0^t(ba^{-1}b)(X(s))ds\right)$ is a P^0_x-martingale with mean 1, we have

$$\begin{aligned}&E^0_x(\exp\left(2\int_0^t a^{-1}b(X(s))d\bar X(s)-\int_0^t(ba^{-1}b)(X(s))ds\right)\le\\&E^0\exp\left(\int_0^t(ba^{-1}b)(X(s))ds\right)\le\exp(||ba^{-1}b||_\infty t).\end{aligned} \tag{4.5}$$

Now (4.1) follows from (4.2), (4.4) and (4.5) □

5. Proof of Theorem 2

Let

$$J_{a,b}(h)=\frac{1}{2}\int_M(\nabla h-a^{-1}b)a(\nabla h-a^{-1}b)g^2dx.$$

We need the following lemma.

Lemma 1. *If* $J_{a,b}(h) \le \frac{1}{2}\int_M (ba^{-1}b)g^2dx$*, then*

$$\frac{1}{2}\int_M (\nabla h a \nabla h) g^2 dx \le 2\int_M (ba^{-1}b)g^2dx. \tag{5.1}$$

Proof. By assumption, we have

$$\frac{1}{2}\int_M (\nabla h a \nabla h) g^2 dx - \int_M (b\nabla h) g^2 dx \le 0. \tag{5.2}$$

Using the inequality $A \cdot B \le \frac{1}{2}|A|^2 + \frac{1}{2}|B|^2$, for d-vectors A and B, we have

$$(b\nabla h)g^2 = (\sqrt{2}gba^{-\frac{1}{2}}) \cdot (\frac{g}{\sqrt{2}}a^{\frac{1}{2}}\nabla h) \le \frac{1}{4}(\nabla h a \nabla h a)g^2 + (ba^{-1}b)g^2. \tag{5.3}$$

The lemma now follows from (5.2) and (5.3). □

Consider now two sets of coefficients a, b and $\hat{a}, \hat{b}$. We need to estimate $|I_{a,b}(\mu) - I_{\hat{a},\hat{b}}(\mu)|$. Recall the formula for the I-function in (2.5), and note that the formula can be rewritten in the following equivalent form:

$$\begin{aligned} I_{a,b}(\mu) &= \frac{1}{2}\int_M (\nabla g a \nabla g) dx - \int_M (b\nabla g) g dx \\ &- \inf_{h\in C^2(M)} \left(\frac{1}{2}\int_M (\nabla h a \nabla h) g^2 dx - \int_M (b\nabla h) g^2 dx \right). \end{aligned} \tag{5.4}$$

We have

$$|\int_M \nabla g(a-\hat{a})\nabla g dx| \le ||a^{-1}(a-\hat{a})||_{d\times d}\int_M |\nabla g a \nabla g| dx, \tag{5.5}$$

and by Cauchy-Schwarz

$$|\int_M ((b-\hat{b})\nabla g) g dx| \le \left(\int_M (\hat{b}-b)a^{-1}(\hat{b}-b)g^2 dx\right)^{\frac{1}{2}} \left(\int_M \nabla g a \nabla g dx\right)^{\frac{1}{2}}. \tag{5.6}$$

Let $l_{a,b} = \inf_{h\in C^2(M)}(\frac{1}{2}\int_M (\nabla h a \nabla h)g^2dx - \int_M (b\nabla h)g^2dx)$. Let $\{h_n\}$ be a minimizing sequence satisfying $\frac{1}{2}\int_M (\nabla h_n a \nabla h_n)g^2dx - \int_M (b\nabla h_n)g^2dx \le l_{a,b} + \frac{1}{n}$. Note that $l_{a,b} = \inf_{h\in C^2(M)} J_{a,b}(h) - \frac{1}{2}\int_M (ba^{-1}b)g^2dx$. Since $J_{a,b}(0) = \frac{1}{2}\int_M (ba^{-1}b)g^2dx$, we may assume that $J_{a,b}(h_n) \le \frac{1}{2}\int_M (ba^{-1}b)g^2dx$.

We have

(5.7)
$$l_{\hat a,\hat b}\le\frac12\int_M(\nabla h_n\hat a\nabla h_n)g^2dx-\int_M(\hat b\nabla h_n)g^2dx=\frac12\int_M(\nabla h_na\nabla h_n)g^2dx$$
$$-\int_M(b\nabla h_n)g^2dx+\frac12\int_M(\nabla h_n(\hat a-a)\nabla h_n)g^2dx-\int_M((\hat b-b)\nabla h_n)g^2dx$$
$$\le l_{a,b}+\frac1n+\frac12\int_M(\nabla h_n(\hat a-a)\nabla h_n)g^2dx-\int_M((\hat b-b)\nabla h_n)g^2dx.$$

Using Lemma 1 and the fact that $J_{a,b}(h_n)\le\frac12\int_M(ba^{-1}b)g^2dx$, we have

(5.8)
$$\frac12|\int_M(\nabla h_n(\hat a-a)\nabla h_n)g^2dx|\le\frac12||a^{-1}(\hat a-a)||_{d\times d}\int_M(\nabla h_na\nabla h_n)g^2dx\le$$
$$2||a^{-1}(\hat a-a)||\int_M(ba^{-1}b)g^2dx.$$

Using Lemma 1 and the fact that $J_{a,b}(h_n)\le\frac12\int_M(ba^{-1}b)g^2dx$, along with Cauchy-Schwarz, we have

(5.9)
$$|\int_M((\hat b-b)\nabla h_n)g^2dx|\le\left(\int_M(\hat b-b)a^{-1}(\hat b-b)g^2dx\right)^{\frac12}\left(\int_M(\nabla h_na\nabla h_n)g^2dx\right)^{\frac12}$$
$$\le2\left(\int_M(\hat b-b)a^{-1}(\hat b-b)g^2dx\right)^{\frac12}\left(\int_M(ba^{-1}b)g^2dx\right)^{\frac12}.$$

Using (5.7)-(5.9) and letting $n\to\infty$ gives

$$(5.10)\qquad \begin{aligned}&l_{\hat a,\hat b}-l_{a,b}\le2||a^{-1}(\hat a-a)||\int_M(ba^{-1}b)g^2dx\\&+2\left(\int_M(\hat b-b)a^{-1}(\hat b-b)g^2dx\right)^{\frac12}\left(\int_M(ba^{-1}b)g^2dx\right)^{\frac12}.\end{aligned}$$

Reversing the roles of $l_{a,b}$ and $l_{\hat a,\hat b}$ in the above analysis, we obtain (5.10) with the roles of a,b and $\hat a,\hat b$ reversed. The theorem now follows from this last comment along with (5.4), (5.5), (5.6), and (5.10). □

6. Proof of Theorem 3

The bound (3.2) follows from the formula for the I-function in (3.3) and the analysis in the proof of Theorem 4 below.

We now turn to the finiteness condition. We will prove that if $\mu = (\mu_1, \cdots, \mu_m) \in \mathcal{P}(M \times (1, \cdots, m))$ has a component μ_k which does not possess a density in $W^{1,2}(M)$, then $I_{a,b,\lambda,q}(\mu) = \infty$. Fix such a μ and assume without loss of generality that μ_1 does not possess a density in $W^{1,2}(M)$.

Consider the process $\tilde{Y}_1$ constructed by following the process $Y = Y(t)$ only when $\sigma(t) = 1$, and stopping the clock for $\tilde{Y}_1$ when $\sigma(t) \neq 1$. That is, if $\{[s_j, t_j)\}_{j=1}^{\infty}$, with $0 \le s_1 < t_1 < s_2 < t_2 < \cdots$, is such that $\sigma(t) = 1$ when $t \in [s_j, t_j)$, for some $j = 1, 2, \cdots$, and $\sigma(t) \neq 1$ otherwise, then $\tilde{Y}_1(t) = Y(s_1 + t)$, for $t \in [0, t_1 - s_1)$, $\tilde{Y}_1(t_1 - s_1 + t) = Y(s_2 + t)$, for $t \in [0, t_2 - s_2)$, etc. Note that $\tilde{Y}_1$ has been constructed to be right-continuous. Denote probabilities for this process starting from $x \in M$ by $\tilde{P}_x$. This process is a diffusion with jumps. The diffusion is governed by L_1, and the jumps occur according to the spatially dependent exponential clock with density $\lambda_1(y)$. If a jump occurs at $y \in M$, then the after-jump distribution is given by some density $\eta_y(z)$ which is bounded and bounded from 0 in y and z. The generator $\mathcal{L}$ for $\tilde{Y}_1$ acts on smooth functions f by

$$\mathcal{L}f(y) = L_1 f(y) + \lambda_1(y) \int_M (f(z) - f(y))\eta_y(z)dz.$$

Let $J_n = J_n(\tilde{Y}_1)$ denote the time of the n-th jump of $\tilde{Y}_1$ and let $N_J(t) = N_J(t; \tilde{Y}_1)$ denote the number of jumps of $\tilde{Y}_1$ up to time t.

Recall how one defines a metric on $\mathcal{P}(M)$: for a dense sequence $\{f_n\}_{n=1}^{\infty}$ in $C(M)$, one defines $d(\theta, \nu) = \sum_{n=1}^{\infty} \frac{|\int_M f_n d\theta - \int_M f_n d\nu|}{1 + |\int_M f_n d\theta - \int_M f_n d\nu|}$. This metric extends to a metric on the space of sub-probability measures on M. We define a metric $d_m(\cdot, \cdot)$ on $\mathcal{P}(M \times (1, \cdots, m))$ in the natural way from the metric $d(\cdot, \cdot)$ on $\mathcal{P}$: $d_m(\theta, \nu) = \sum_{k=1}^{m} d(\theta_k, \nu_k)$.

For any non-zero sub-probability measure $\nu \in \mathcal{P}(M)$, we denote its normalized version by $\tilde{\nu} = \frac{\nu}{\nu(M)}$. Consider the balls $B_{\frac{1}{n}}(\mu) \subset \mathcal{P}(M \times (1, \cdots, m))$, $n = 1, 2, \cdots$. From the above description of the metrics, it is easy to see that there exists a $c > 1$ such that

$$\{\tilde{\nu}_1 : \nu = (\nu_1, \cdots, \nu_m) \in B_{\frac{1}{n}}(\mu)\} \subset B_{\frac{c}{n}}(\tilde{\mu}_1), \tag{6.1}$$

where $B_r(\tilde\mu_1)$ denotes the ball in $\mathcal{P}(M)$ of radius r, centered at $\tilde\mu_1$. By (2.2) and the lower semi-continuity of $I_{a,b,\lambda,q}$, it follows that the condition $I_{a,b,\lambda,q}(\mu)=\infty$ is equivalent to the condition

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log P_x(\frac{1}{t}l^t\in\bar B_{\frac{1}{n}}(\mu))=-\infty. \tag{6.2}$$

Let $\tilde l^t=\tilde l^t_{\tilde Y_1}$ denote the occupation measure for the process $\tilde Y_1$. By construction, the normalized version of the first component of l^t_Y is $\tilde l^t_{\tilde Y_1}$—in the notation above, $(\tilde{l^t_Y})_1=\tilde l^t_{\tilde Y_1}$. Using this with (6.1), it follows that (6.2) will hold if

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log \tilde P_x(\frac{1}{t}\tilde l^t\in\bar B_{\frac{1}{n}}(\tilde\mu_1))=-\infty. \tag{6.3}$$

Thus, to prove the theorem it suffices to show that (6.3) holds.

Let $\tilde P^U_x$ denote the probabilities for the $\tilde Y_1$ process in the particular case that $\lambda_1(y)\equiv 1$ and $\eta_y(z)=\frac{1}{|M|}$; that is, in the case that the generator operates on smooth functions f by $\mathcal{L}f(y)=L_1f(y)+\frac{1}{|M|}\int_M(f(z)-f(y))dy$. Recall the definitions of J_n and $N_J(t)$ above, and define $J_0=0$. We have

$$\begin{aligned}&d\tilde P_x(J_{n+1}>J_n+t,\tilde Y_1(J_n+s)=y(s),0\le s\le t|J_n)\\&=\exp(-\int_{J_n}^t\lambda_1(y(s))ds)\ d\tilde P_x(\tilde Y_1(J_n+s)=y(s),0\le s\le t|J_n),\end{aligned}$$

and similarly for $\tilde P^U_x$, with λ_1 replaced by 1. We also have

$$\tilde P_x(\tilde Y_1(J_n)=dz|\tilde Y_1(J_n^-))=\eta_{\tilde Y_1(J_n^-)}(z)dz,$$

and similarly for $\tilde P^U_x$, with the right hand side above replaced by $\frac{1}{|M|}dz$. In light of the above formulas, it follows that the Radon-Nikodym derivative

$\frac{d\tilde{P}_x}{d\tilde{P}_x^U}|_{\mathcal{F}_t}$ is given by

(6.4)

$$\frac{d\tilde{P}_x}{d\tilde{P}_x^U}|_{\mathcal{F}_t}(\tilde{Y}_1) =$$

$$\exp(t)\Big(\prod_{k=1}^{N_J(t)}\lambda_1(\tilde{Y}_1(J_k))\exp(-\int_{J_{k-1}}^{J_k}\lambda_1(\tilde{Y}_1(s))ds)\Big)\Big(\exp(-\int_{J_{N_J(t)}}^{t}\lambda_1(\tilde{Y}_1(s))ds\Big)\times$$

$$|M|^{N_J(t)}\prod_{k=1}^{N_J(t)}\eta_{\tilde{Y}_1(J_k^-)}(\tilde{Y}_1(J_k)),$$

where $\mathcal{F}_t$ denotes the standard filtration on right-continuous paths.

Let $\gamma > 0$. We write

(6.5)

$$\tilde{P}_x(\frac{1}{t}\tilde{l}^t \in \bar{B}_{\frac{1}{n}}(\tilde{\mu}_1),\ N_J(t) \le \gamma t) = \tilde{E}_x^U\Big(\frac{d\tilde{P}_x}{d\tilde{P}_x^U}|_{\mathcal{F}_t};\ \frac{1}{t}\tilde{l}^t \in \bar{B}_{\frac{1}{n}}(\tilde{\mu}_1),\ N_J(t) \le \gamma t\Big).$$

By (6.4) and (6.5), it follows that there exist constants $c_0, c_1 > 0$, independent of γ, such that

(6.6)

$$\exp(-c_0\gamma t - c_1 t)\tilde{P}_x(\frac{1}{t}\tilde{l}^t \in \bar{B}_{\frac{1}{n}}(\tilde{\mu}_1),\ N_J(t) \le \gamma t) \le \tilde{P}_x^U(\frac{1}{t}\tilde{l}^t \in \bar{B}_{\frac{1}{n}}(\tilde{\mu}_1),\ N_J(t) \le \gamma t)$$
$$\le \exp(c_0\gamma t + c_1 t)\tilde{P}_x(\frac{1}{t}\tilde{l}^t \in \bar{B}_{\frac{1}{n}}(\tilde{\mu}_1),\ N_J(t) \le \gamma t).$$

By standard large deviations estimates, it follows that

(6.7)

$$\lim_{\gamma\to\infty}\limsup_{t\to\infty}\tilde{P}_x(N_J(t) > \gamma t) = -\infty \text{ and } \lim_{\gamma\to\infty}\limsup_{t\to\infty}\tilde{P}_x^U(N_J(t) > \gamma t) = -\infty.$$

From (6.6) and (6.7) it follows that (6.3) holds if and only if

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log\tilde{P}_x^U(\frac{1}{t}\tilde{l}^t \in \bar{B}_{\frac{1}{n}}(\tilde{\mu}_1)) = -\infty. \tag{6.8}$$

Thus, to prove the theorem it suffices to prove (6.8).

Let $L_{1,0} = \frac{1}{2}\nabla\cdot a_1\nabla$ and let $\tilde{P}_x^{U,0}$ denote probabilities for the $\tilde{Y}_1$ process in the particular case that $L_1 = L_{1,0}$, $\lambda_1(y) \equiv 1$ and $\eta_y(z) = \frac{1}{|M|}$. Note that under $\tilde{P}_x^{U,0}$, the random variables $\{J_n\}_{n=1}^\infty$ and $\{\tilde{Y}_1(J_n)\}_{n=1}^\infty$ are independent of the paths $\{\tilde{Y}_1(t) - \tilde{Y}_1(J_n), J_n \le t < J_{n+1}\}_{n=0}^\infty$, because in this case $\lambda_1(y)$

is independent of y and $\eta_y(z)$ is independent of y and z. Because of this independence, the Girsanov transformation can be applied to give

$$\frac{d\tilde{P}_x^U}{d\tilde{P}_x^{U,0}}|_{\mathcal{F}_t} =$$
$$\exp\left(\sum_{k=0}^{\infty}\int_{J_k\wedge t}^{(J_{k+1}\wedge t)^-} a_1^{-1}b_1(\tilde{Y}_1(s))d\bar{Y}_1(s) - \frac{1}{2}\int_0^t (b_1a_1^{-1}b_1)(\tilde{Y}_1(s))ds\right),$$

where $\bar{Y}_1(s) = \tilde{Y}_1(s) - \frac{1}{2}\int_0^s (\nabla\cdot a_1)(\tilde{Y}_1(r))dr$. (We need to write the stochastic integrals in the form $\sum_{k=0}^{\infty}\int_{J_k\wedge t}^{(J_{k+1}\wedge t)^-}$ instead of simply as $\int_0^t$ in order to skirt the problem of the jumps in $\tilde{Y}_1$.) Now the same argument used in the proof of Theorem 1 shows that (6.8) will hold if

$$\lim_{n\to\infty}\limsup_{t\to\infty}\frac{1}{t}\log\tilde{P}_x^{U,0}(\frac{1}{t}\tilde{l}^t\in\bar{B}_{\frac{1}{n}}(\tilde{\mu}_1)) = -\infty. \tag{6.9}$$

We will now complete the proof of the theorem by showing that (6.9) holds.

Let $I^{U,0}$ denote the I-function for the process $\tilde{Y}_1$ under $\tilde{P}_x^{U,0}$. By (2.2) and the lower semi-continuity of the I-function, (6.9) is equivalent to $I^{U,0}(\tilde{\mu}_1) = \infty$. The generator $\mathcal{L}$ for $\tilde{Y}_1$ under $\tilde{P}_x^{U,0}$ operates on smooth functions f by $\mathcal{L}f(y) = \frac{1}{2}\nabla\cdot a_1\nabla f(y) + \frac{1}{|M|}\int_M (f(z) - f(y))dz$. It is easy to check that $\int_M g\mathcal{L}f dy = \int_M f\mathcal{L}g dy$; thus the process is reversible with respect to Lebesgue measure, and the generator $\mathcal{L}$ can be realized as a self-adjoint operator. Since $(-\mathcal{L}f, f) = \frac{1}{2}\int_M(\nabla f a_1\nabla f)dx + \int_M\int_M(f(z) - f(y))f(y)dzdy$, for smooth f, it follows that

(6.10)
$$||(-\mathcal{L})^{\frac{1}{2}}f||_{2,dx}^2 = \begin{cases} \frac{1}{2}\int_M(\nabla f a_1\nabla f)dx + \int_M\int_M(f(z) - f(y)f(y))dzdy, \\ \quad \text{if } f\in W^{1,2}(M); \\ \infty, \text{ otherwise.}\end{cases}$$

Recall that by assumption, μ_1 does not possess a density in $W^{1,2}(M)$; thus, $\tilde{\mu}_1$ also does not possess a density in $W^{1,2}(M)$. Therefore, in light of (6.10) and (2.3), which gives the formula for the I-function in the self-adjoint case, we conclude that $I^{U,0}(\tilde{\mu}_1) = \infty$. □

7. Proof of Theorem 4

Consider two sets of coefficients a, b, λ, q and $\hat{a}, \hat{b}, \hat{\lambda}, \hat{q}$. We need to estimate $|I_{a,b,\lambda,q}(\mu) - I_{\hat{a},\hat{b},\hat{\lambda},\hat{q}}(\mu)|$. The formula for the I-function in (3.3) can be rewritten in the following equivalent form:

(7.1)

$$I_{a,b,\lambda,q}(\mu) = \frac{1}{2}\sum_{k=1}^{m}\int_M (\nabla g_k a_k \nabla g_k)dy - \sum_{k=1}^{m}\int_M (b_k\nabla g_k)g_k dy + \sum_{k=1}^{m}\int_M \lambda_k g_k^2 dy$$
$$- \inf_h \Big(\frac{1}{2}\sum_{k=1}^{m}\int_M (\nabla h_k a_k \nabla h_k)g_k^2 dy - \sum_{k=1}^{m}\int_M (b_k \nabla h_k)g_k^2 dy$$
$$+ \sum_{k,l=1}^{m}\int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l)dy\Big).$$

Now the first two lines on the right hand side of (3.4) follow from the estimates in the proof of Theorem 2. The third line on the right hand side of (3.4) follows directly from (3.3). Thus, letting

$$\begin{aligned} l_{a,b,\lambda,q} = \inf_h \Big(&\frac{1}{2}\sum_{k=1}^{m}\int_M (\nabla h_k a_k \nabla h_k)g_k^2 dy - \sum_{k=1}^{m}\int_M (b_k \nabla h_k)g_k^2 dy \\ &+ \sum_{k,l=1}^{m}\int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l)dy\Big), \end{aligned} \tag{7.2}$$

it remains to show that

$$|l_{a,b,\lambda,q} - l_{\hat{a},\hat{b},\hat{\lambda},\hat{q}}| \le \max\Big(D(a,b,\lambda,q,\hat{a},\hat{b},\hat{\lambda},\hat{q}), D(\hat{a},\hat{b},\hat{\lambda},\hat{q},a,b,\lambda,q)\Big), \tag{7.3}$$

where D is as in the statement of the theorem.

Let

$$J_{a,b,\lambda,q}(h) = \frac{1}{2}\sum_{k=1}^{m}\int_M (\nabla h_k - a_k^{-1}b_k)a_k(\nabla h_k - a_k^{-1}b_k)g_k^2 dy$$
$$+ \sum_{k,l=1}^{m}\int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l)dy,$$

We need the following lemma.

Lemma 2. *If* $J_{a,b,\lambda,q}(h) \leq \frac{1}{2}\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy + \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l dy$, *then*

$$\frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k a_k \nabla h_k) g_k^2 dy + \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l) dy$$
$$\leq 2\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy + \frac{m+1}{2}(\sup_{y\in M;k} \lambda_k(y)).$$

If $q(y)$ *is doubly stochastic for each* $y \in M$, *then the term* $m+1$ *above may be replaced by 2.*

Proof. By assumption, we have

$$\begin{aligned}&\frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k a \nabla h_k) g_k^2 dy + \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l) dy \\ &\leq \sum_{k=1}^m \int_M b_k \nabla h_k g_k^2 dy + \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l dy.\end{aligned} \tag{7.4}$$

From (7.4) and (5.3) along with the fact that $\sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l) dy \geq 0$, we have

$$\begin{aligned}&\frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k a_k \nabla h_k) g_k^2 dy + \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k - h_l) dy \\ &\leq 2\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy + 2\sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l dy.\end{aligned} \tag{7.5}$$

Using the inequality $AB \leq \frac{1}{2}A^2 + \frac{1}{2}B^2$ and the fact that $0 \leq q_{kl} \leq 1$, we have

$$\begin{aligned}&\sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l dy \leq \frac{1}{2}(\sup_{y\in M;k} \lambda_k(y)) \sum_{k,l=1}^m \int_M q_{kl}(g_k^2 + g_l^2) dy \\ &\leq \frac{m+1}{2}(\sup_{y\in M;k} \lambda_k(y)).\end{aligned} \tag{7.6}$$

If $q(y)$ is doubly stochastic for each $y \in M$, then it is easy to see that $m+1$ may be replaced by 2 in the last inequality above. The lemma now follows from (7.5) and (7.6). □

Let $\{h^n\} = \{h_1^n, \cdots, h_m^n\}$ be a minimizing sequence in the definition of $l_{a,b,\lambda,q}$ in (7.2), satisfying

$$\frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k^n a_k \nabla h_k^n) g_k^2 dy - \sum_{k=1}^m \int_M (b_k \nabla h_k^n) g_k^2 dy$$
$$+ \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k^n - h_l^n) dy \le l_{a,b,\lambda,q} + \frac{1}{n}.$$

Since $l_{a,b,\lambda,q} = \inf_h J(h) - \frac{1}{2}\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy$ and since $J(0) = \frac{1}{2}\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy + \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l dy$, we may assume that

$$J_{a,b,\lambda,q}(h^n) \le \frac{1}{2}\sum_{k=1}^m \int_M (b_k a_k^{-1} b_k) g_k^2 dy + \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l dy. \tag{7.7}$$

We have

(7.8)

$$l_{\hat{a},\hat{b},\hat{\lambda},\hat{q}} \le \frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k^n \hat{a}_k \nabla h_k^n) g_k^2 dy - \sum_{k=1}^m \int_M (\hat{b}_k \nabla h_k^n) g_k^2 dy$$
$$+ \sum_{k,l=1}^m \int_M \hat{\lambda}_k \hat{q}_{kl} g_k g_l \exp(h_k^n - h_l^n) dy$$
$$= \frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k^n a_k \nabla h_k^n) g_k^2 dy - \sum_{k=1}^m \int_M (b_k \nabla h_k^n) g_k^2 dy$$
$$+ \sum_{k,l=1}^m \int_M \lambda_k q_{kl} g_k g_l \exp(h_k^n - h_l^n) dy$$
$$+ \frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k^n (\hat{a}_k - a_k) \nabla h_k^n) g_k^2 dy - \sum_{k=1}^m \int_M ((\hat{b}_k - b_k) \nabla h_k^n) g_k^2 dy$$
$$+ \sum_{k,l=1}^m \int_M \hat{\lambda}_k (\hat{q}_{kl} - q_{kl}) g_k g_l \exp(h_k^n - h_l^n) dy$$
$$+ \sum_{k,l=1}^m \int_M (\hat{\lambda}_k - \lambda_k) q_{kl} g_k g_l \exp(h_k^n - h_l^n) dy$$
$$\le l_{a,b,\lambda,q} + \frac{1}{n} + \frac{1}{2}\sum_{k=1}^m \int_M (\nabla h_k^n (\hat{a}_k - a_k) \nabla h_k^n) g_k^2 dy - \sum_{k=1}^m \int_M ((\hat{b}_k - b_k) \nabla h_k^n) g_k^2 dy$$
$$+ \sum_{k,l=1}^m \int_M \hat{\lambda}_k (\hat{q}_{kl} - \hat{q}_{kl}) g_k g_l \exp(h_k^n - h_l^n) dy + \sum_{k,l=1}^m \int_M (\hat{\lambda}_k - \lambda_k) q_{kl} g_k g_l \exp(h_k^n - h_l^n) dy.$$

We estimate from above the last two terms on the right hand side of (7.8) as follows:

$$
\begin{aligned}
&|\sum_{k,l=1}^{m}\int_M \hat\lambda_k(\hat q_{kl}-\hat q_{kl})g_kg_l\exp(h_k^n-h_l^n)dy| \\
&\le (\sup_{y\in M;k,l}\frac{\hat\lambda_k}{\lambda_k}(y)|\frac{\hat q_{kl}}{q_{kl}}(y)-1|)\sum_{k,l=1}^{m}\int_M \lambda_kq_{kl}g_kg_l\exp(h_k^n-h_l^n)dy
\end{aligned}
\tag{7.9}
$$

and

$$
\begin{aligned}
&|\sum_{k,l=1}^{m}\int_M (\hat\lambda_k-\lambda_k)q_{kl}g_kg_l\exp(h_k^n-h_l^n)dy| \\
&\le (\sup_{y\in M;k}|\frac{\hat\lambda_k}{\lambda_k}(y)-1|)\sum_{k,l=1}^{m}\int_M \lambda_kq_{kl}g_kg_l\exp(h_k^n-h_l^n)dy.
\end{aligned}
\tag{7.10}
$$

And we estimate from above the first integral term on the right hand side of (7.8) as follows:

$$
\frac{1}{2}|\sum_{k=1}^{m}\int_M(\nabla h_k^n(\hat a_k-a_k)\nabla h_k^n)g_k^2dy|\le\frac{1}{2}\sum_{k=1}^{m}||a_k^{-1}(\hat a_k-a_k)||_{d\times d}\int_M(\nabla h_k^na_k\nabla h_k^n)g_k^2dy.
\tag{7.11}
$$

Using (7.9)-(7.11) along with (7.7) and Lemma 2, we have

$$
\begin{aligned}
&\frac{1}{2}\sum_{k=1}^{m}|\int_M(\nabla h_k^n(\hat a_k-a_k)\nabla h_k^n)g_k^2dy \\
&+|\sum_{k,l=1}^{m}\int_M \hat\lambda_k(\hat q_{kl}-q_{kl})g_kg_l\exp(h_k^n-h_l^n)dy| \\
&+|\sum_{k,l=1}^{m}\int_M (\hat\lambda_k-\lambda_k)q_{kl}g_kg_l\exp(h_k^n-h_l^n)dy| \\
&\le\Big(\sup_{y\in M;k,l}\frac{\hat\lambda_k}{\lambda_k}(y)|\frac{\hat q_{kl}}{q_{kl}}(y)-1|+\sup_{y\in M;k}|\frac{\hat\lambda_k}{\lambda_k}(y)-1|\Big)\wedge\Big(\max_k||a_k^{-1}(\hat a_k-a_k)||_{d\times d}\Big)\times \\
&\Big(2\sum_{k=1}^{m}\int_M(b_ka_k^{-1}b_k)g_k^2dy+\frac{m+1}{2}(\sup_{y\in M;k}\lambda_k(y))\Big).
\end{aligned}
\tag{7.12}
$$

It follows trivially from (7.7) and Lemma 2 that

$$
\int_M(\nabla h_k^na_k\nabla h_k^n)g_k^2dy\le4\sum_{k=1}^{m}\int_M(b_ka_k^{-1}b_k)g_k^2dy+(m+1)(\sup_{y\in M;k}\lambda_k(y)),\ k=1,\cdots,m.
$$

Using this along with Cauchy-Schwarz, we estimate the second integral term on the right hand side of (7.8) as follows:

(7.13)

$$
\begin{aligned}
&|\sum_{k=1}^{m}\int_M((\hat{b}_k-b_k)\nabla h_k^n)g_k^2dy| \\
&\leq \sum_{k=1}^{m}\Big(\int_M(\nabla h_k^n a_k\nabla h_k^n)g_k^2dy\Big)^{\frac{1}{2}}\Big(\int_M(\hat{b}_k-b_k)a_k^{-1}(\hat{b}_k-b_k)g_k^2dy\Big)^{\frac{1}{2}} \\
&\leq \sum_{k=1}^{m}\Big(4\int_M(b_ka_k^{-1}b_k)g_k^2dy+(m+1)(\sup_{y\in M;k}\lambda_k(y))\Big)^{\frac{1}{2}}\Big(\int_M(\hat{b}_k-b_k)a_k^{-1}(\hat{b}_k-b_k)g_k^2dy\Big)^{\frac{1}{2}}.
\end{aligned}
$$

Using (7.8), (7.12) and (7.13) and letting $n\to\infty$, we obtain

(7.14)

$$
\begin{aligned}
&l_{\hat{a},\hat{b},\hat{\lambda},\hat{q}}-l_{a,b,\lambda,q}\leq \\
&\Big(\sup_{y\in M;k,l}\frac{\hat{\lambda}_k}{\lambda_k}(y)|\frac{\hat{q}_{kl}}{q_{kl}}(y)-1|+\sup_{y\in M;k}|\frac{\hat{\lambda}_k}{\lambda_k}(y)-1|\Big)\wedge\Big(\max_k||a_k^{-1}(\hat{a}_k-a_k)||_{d\times d}\Big)\times \\
&\Big(2\sum_{k=1}^{m}\int_M(b_ka_k^{-1}b_k)g_k^2dy+\frac{m+1}{2}(\sup_{y\in M;k}\lambda_k(y))\Big) \\
&+\sum_{k=1}^{m}\Big(4\int_M(b_ka_k^{-1}b_k)g_k^2dy+(m+1)(\sup_{y\in M;k}\lambda_k(y))\Big)^{\frac{1}{2}}\Big(\int_M(\hat{b}_k-b_k)a_k^{-1}(\hat{b}_k-b_k)g_k^2dy\Big)^{\frac{1}{2}}.
\end{aligned}
$$

Reversing the roles of $l_{a,b,\lambda,q}$ and $l_{\hat{a},\hat{b},\hat{\lambda},\hat{q}}$ in the analysis above, we obtain (7.14) with the roles of a,b,λ,q and $\hat{a},\hat{b},\hat{\lambda},\hat{q}$ reversed. This proves (7.3). □

Acknowledgement. The author thanks Yuri Kifer for posing to him the questions considered in this paper.

References

[1] Donsker, M. D. and Varadhan, S.R.S. Asymptotic evaluation of certain Markov process expectations for large time. I. *Comm. Pure Appl. Math.* **28** (1975), 1–47.

[2] Donsker, M. D. and Varadhan, S. R. S. On the principal eigenvalue of second-order elliptic differential operators. *Comm. Pure Appl. Math.* **29** (1976), 595–621.

[3] Freidlin, M. Functional Integration and Partial Differential Equations. *Annals of Mathematics Studies 109, Princeton University Press* Princeton, NJ, 1985.

[4] Kifer, Y., Large deviations and adiabatic transitions for dynamical systems and Markov processes in fully coupled averaging, *preprint.*

[5] Pinsky, R. G. On evaluating the Donsker-Varadhan I-function. *Ann. Probab.* **13** (1985), 342–362.

[6] Pinsky, R. G. The I-function for diffusion processes with boundaries. *Ann. Probab.* **13** (1985), 676–692.

[7] Pinsky, R. G. and Scheutzow, M. Some remarks and examples concerning the transience and recurrence of random diffusions. *Ann. Inst. H. Poincaré Probab. Statist* **28** (1992), 519–536.

Department of Mathematics, Technion—Israel Institute of Technology, Haifa, 32000, Israel

E-mail address: pinsky@math.technion.ac.il