Stochastic Systems
2015, Vol. 5, No. 1, 87–119
DOI: 10.1214/14-SSY138

MODERATE DEVIATIONS FOR RECURSIVE STOCHASTIC ALGORITHMS

By Paul Dupuis* and Dane Johnson†

Brown University

We prove a moderate deviation principle for the continuous time interpolation of discrete time recursive stochastic processes. The methods of proof are somewhat different from the corresponding large deviation result, and in particular the proof of the upper bound is more complicated.

1. Introduction. In this paper we consider $\mathbb{R}^d$-valued discrete time processes of the form

$$X_{i+1}^n = X_i^n + \frac{1}{n}b(X_i^n) + \frac{1}{n}\upsilon_i(X_i^n),\ X_0^n \doteq x_0,$$

where $\{\upsilon_i(\cdot)\}_{i\in\mathbb{N}_0}$ are zero mean random independent and identically distributed (iid) vector fields, and focus on their continuous time piecewise linear interpolations $\{X^n(t)\}_{0\le t\le T}$ with $X^n(i/n) = X_i^n$ (see (2.5) for the precise definition). Under certain conditions there is a law of large number limit $X^0 \in C([0,T] : \mathbb{R}^d)$, and the large deviations of X^n from this limit have been studied extensively (see, e.g., [1, 10, 12, 15]). Here we introduce a scaling $a(n)$ satisfying $a(n) \to 0$ and $a(n)\sqrt{n} \to \infty$, and study the amplified difference between X^n and its noiseless version $X^{n,0}$ (see Section 2 for the definition of $X^{n,0}$):

$$Y^n = a(n)\sqrt{n}(X^n - X^{n,0}).$$

Under Condition 2.1 stated below $\sup_{t\in[0,T]}\left\|X^0(t) - X^{n,0}(t)\right\| \sim O(1/n)$, and hence Y^n will behave the same asymptotically as $a(n)\sqrt{n}(X^n - X^0)$. We demonstrate, under weaker conditions on the noise $\upsilon_i(\cdot)$ than are necessary when considering X^n, that Y^n satisfies the large deviation principle

Received January 2014.
*Research supported in part by the Department of Energy (DE-SCOO02413, DE-SC0010539), the National Science Foundation (DMS-1317199), and the Army Research Office (W911NF-12-1-0222).
†Research supported in part by the Department of Energy (DE-SCOO02413) and the Air Force Office of Scientific Research (FA9550-12-1-0399).
MSC 2010 subject classifications: Primary 60F10, 60J05; secondary 60F17.
Keywords and phrases: Large deviations, moderate deviations, weak convergence.

on $C([0,T]:\mathbb{R}^d)$ with a "Gaussian" type rate function. As is customary for this type of scaling, we refer to this as moderate deviations.

To demonstrate this result we prove the equivalent Laplace principle, which involves evaluating limits of quantities of the form

$$a(n)^2 \log E\left[\exp\left\{-\frac{1}{a(n)^2}F(Y^n)\right\}\right]$$

when F is bounded and continuous. This is done by representing each of these quantities in terms of a stochastic control problem, and then using weak convergence methods as in [12]. Key results needed in this approach are establishing tightness of controls and controlled processes, and identifying their limits.

While one might expect the proof of this moderate deviations result to be similar to the corresponding large deviations result, there are important differences. For example, the tightness proof is significantly more complicated in the case of moderate deviations than it is in the case of large deviations. For large deviations one is able to establish an a priori bound on certain relative entropy costs associated with any sequence of nearly minimizing controls, and under this boundedness of the relative entropy costs, the empirical measures of the controlled driving noises as well as the controlled processes are tight. However, owing to the scaling in moderate deviations, even with the information that the analogous relative entropy costs decay like $O(1/a(n)^2 n)$, tightness of the empirical measures of the noises does not hold. Instead, one must consider empirical measures of the conditional means of the noises, and additional effort is required for a law of large numbers type result that shows that the conditional means are adequate to determine the limit. This extra difficulty arises for moderate deviations (even with the vanishing relative entropy costs), because with this scaling the noise is amplified by a term of the form $a(n)\sqrt{n}$.

A second way in which the proofs for large and moderate deviations differ is in their treatment of degenerate noise, i.e., problems where the support of $\upsilon_i(\cdot)$ is not all of $\mathbb{R}^d$. This leads to significant difficulties in the proof of the large deviation lower bound, and requires a delicate and involved mollification argument. In contrast, the proof in the setting of moderate deviations, though more involved than the nondegenerate case, is much more straightforward.

As a potential application of these results we mention their usefulness in the design and analysis of Monte Carlo schemes. It is well known that accelerated Monte Carlo schemes (e.g., importance sampling and splitting) benefit by using information contained in the large deviation rate function as part of the algorithm design (e.g., [3, 8, 13, 14]). In a situation where one considers events of small but not too small probability one may find the moderate deviation approximation both adequate and relatively easy to apply, since moderate deviations lead to situations where the objects needed to design an efficient scheme can be explicitly constructed in terms of solutions to the linear-quadratic regulator. These issues will be explored elsewhere.

The existing literature on moderate deviations considers various settings. Baldi [2] considers the same scaling used here but with no state dependence. For the empirical measure of a Markov chain, de Acosta [7] and de Acosta and Chen [6] prove lower and upper bounds, respectively. Guillin [17] considers inhomogeneous functionals of a "fast" continuous time ergodic Markov chain, and in [18] this is extended to a small noise diffusion whose coefficients depend on the "fast" Markov chain. There are also results for martingale differences such as Dembo [9], Gao [16], and Djellout [11]. For various reasons, the issues previously mentioned regarding the difficulties in the proof of the upper bound and the simplification in the lower bound for degenerate noise do not play a role in these papers. For instance, proving tightness in a moderate deviations setting for continuous time processes such as diffusions is typically much easier. This is because measures on path space that have bounded relative entropy with respect to Wiener measure have significantly less variability than those with bounded relative entropy with respect to a discrete time process. In particular, bounded relative entropy automatically restricts to what one could consider to be "exponential tilts" of the original distribution in continuous time, which does not happen in discrete time, and is the reason more effort must be put into the proof of tightness. This is illustrated by the convenient alternative formulations of the relative entropy representation for some continuous time processes (see [4] for Brownian motion and [5] for Poisson random measures).

The paper is organized as follows. Section 2 gives the statement of the problem and notation. Section 3 contains the proof of tightness and the characterization of limits, which account for most of the mathematical difficulties, and are also the main results needed to prove the Laplace principle. Sections 4 and 5 give the proofs of the upper and lower Laplace bounds. Although all proofs are given for the time interval $[0,1]$, they extend with only notational differences to $[0,T]$ for any $T \in (0,\infty)$.

Aknowledgement. The authors thank the referees for several suggestions that improved the paper.

2. Background and Notation. Let

$$X_{i+1}^n = X_i^n + \frac{1}{n} b(X_i^n) + \frac{1}{n} \upsilon_i(X_i^n),\ X_0^n \doteq x_0$$

where the $\{\upsilon_i(\cdot)\}_{i \in \mathbb{N}_0}$ are zero mean iid vector fields with distribution given by the stochastic kernel μ_x. Thus if $\mathcal{B}(\mathbb{R}^d)$ is the Borel σ-algebra on $\mathbb{R}^d$, then $x \to \mu_x(B)$ is measurable for all $B \in \mathcal{B}(\mathbb{R}^d)$, $\mu_x(\cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R}^d)$ for all $x \in \mathbb{R}^d$, and $P(\upsilon_i(x) \in B) = \mu_x(B)$ for all $x \in \mathbb{R}^d$, $B \in \mathcal{B}(\mathbb{R}^d)$ and $i \in \mathbb{N}_0$. Define

$$H_c(x, \alpha) \doteq \log \left(\int_{\mathbb{R}^d} e^{\langle y, \alpha \rangle} \mu_x(dy) \right)$$

for $\alpha \in \mathbb{R}^d$. The subscript c reflects the fact that this log moment generating function uses the centered distribution μ_x, rather than the usual $H(x, \alpha) = H_c(x, \alpha) + \langle \alpha, b(x) \rangle$. We will use the following.

Condition 2.1.

- *There exists $\lambda > 0$ and $K_{mgf} < \infty$ such that*

$$\sup_{x \in \mathbb{R}^d} \sup_{\|\alpha\| \leq \lambda} H_c(x, \alpha) \leq K_{mgf}. \tag{2.1}$$

- *$x \to \mu_x(dy)$ is continuous with respect to the topology of weak convergence.*
- *$b(x)$ is continuously differentiable, and the norm of both $b(x)$ and its derivative are uniformly bounded by some constant $K_b < \infty$.*

Throughout this paper we let $\|\alpha\|_A^2 = \langle \alpha, A\alpha \rangle$ for any $\alpha \in \mathbb{R}^d$ and symmetric, nonnegative definite matrix A. Define

$$A_{ij}(x) \doteq \int_{\mathbb{R}^d} y_i y_j \mu_x(dy),$$

and note that the weak continuity of μ_x with respect to x and (2.1) ensure that $A(x)$ is continuous in x and its norm is uniformly bounded by some constant K_A. Note that

$$\frac{\partial H_c(x, 0)}{\partial \alpha_i} = \int_{\mathbb{R}^d} y_i \mu_x(dy) = 0$$

and

$$\frac{\partial^2 H_c(x, 0)}{\partial \alpha_i \partial \alpha_j} = \int_{\mathbb{R}^d} y_i y_j \mu_x(dy) = A_{ij}(x)$$

for all $i, j \in \{1, \ldots, d\}$ and $x \in \mathbb{R}^d$, and that $A(x)$ is nonnegative-definite and symmetric. For $x \in \mathbb{R}^d$ we can therefore write

$$A(x) = Q(x)\Lambda(x)Q^T(x),$$

where $Q(x)$ is an orthogonal matrix whose columns are the eigenvectors of $A(x)$ and $\Lambda(x)$ is the diagonal matrix consisting of the eigenvalues of $A(x)$ in descending order. In what follows we define $\Lambda^{-1}(x)$ to be the diagonal matrix with diagonal entries equal to the inverse of the corresponding eigenvalue for the positive eigenvalues, and equal to ∞ for the zero eigenvalues. Then when we write

$$\|\alpha\|^2_{A^{-1}(x)} = \|\alpha\|^2_{Q(x)\Lambda^{-1}(x)Q^T(x)}, \tag{2.2}$$

we mean a value of ∞ for $\alpha \in \mathbb{R}^d$ not in the linear span of the eigenvectors corresponding to the positive eigenvalues, and the standard value for vectors $\alpha \in \mathbb{R}^d$ in that linear span. (Note that even if the definition of $A^{-1}(x)$ is ambiguous, in that for α in the range of $A(x)$ there may be more than one v such that $A(x)v = \alpha$, the value of $\|\alpha\|^2_{A^{-1}(x)}$ is not ambiguous. Indeed, since the eigenvectors can be assumed orthogonal, for all such v $\langle v, \alpha\rangle$ coincides with $\langle \bar{v}, \alpha\rangle$, where $\bar{v}$ is the solution in the span of eigenvectors corresponding to positive eigenvalues.) Assumption (2.1) implies there exists some $K_{DA} < \infty$ and $\lambda_{DA} \in (0, \lambda]$ (independent of x) such that

$$\sup_{x\in\mathbb{R}^d} \sup_{\|\alpha\|\le\lambda_{DA}} \max_{i,j,k} \left| \frac{\partial^3 H_c(x,\alpha)}{\partial\alpha_i \partial\alpha_j \partial\alpha_k} \right| \le \frac{K_{DA}}{d^3}, \tag{2.3}$$

and consequently for all $\|\alpha\| \le \lambda_{DA}$ and all $x \in \mathbb{R}^d$

$$\frac{1}{2}\|\alpha\|^2_{A(x)} - \|\alpha\|^3 K_{DA} \le H_c(x,\alpha) \le \frac{1}{2}\|\alpha\|^2_{A(x)} + \|\alpha\|^3 K_{DA}. \tag{2.4}$$

Define the continuous time linear interpolation of X_i^n by $X^n(i/n) = X_i^n$ for $i = 0, \ldots, n$ and

$$X^n(t) = (i + 1 - nt)X_i^n + (nt - i)X_{i+1}^n \tag{2.5}$$

for $t \in (i/n, i/n + 1/n)$. In addition, define

$$X_{i+1}^{n,0} = X_i^{n,0} + \frac{1}{n} b\left(X_i^{n,0}\right), \quad X_0^{n,0} = x_0$$

and let $X^{n,0}(t)$ be the analogous continuous time linear interpolation given by $X^{n,0}(i/n) = X_i^{n,0}$ for $i = 0, \ldots, n$ and

$$X^{n,0}(t) = (i + 1 - nt)X_i^{n,0} + (nt - i)X_{i+1}^{n,0}$$

for $t \in (i/n, i/n + 1/n)$. Clearly $X^{n,0}(t) \to X^0(t)$ in $C([0,1] : \mathbb{R}^d)$, where

$$X^0(t) = \int_0^t b(X^0(s))ds + x_0.$$

Since $E\upsilon_i(x) = 0$ for all $x \in \mathbb{R}^d$, we know that $X^n(t) \to X^0(t)$ in $C([0,1] : \mathbb{R}^d)$ in probability. One can estimate probabilities for events involving paths outside the law of large numbers limit X^0 by proving a large deviation principle and finding the corresponding rate function.

Definition 2.2. Let $\{Z^n,\ n \in \mathbb{N}\}$ be a sequence of random variables defined on a probability space $(\Omega, \mathcal{F}, P)$ and taking values in a Polish space $\mathcal{Z}$. A function $I : \mathcal{Z} \to [0,\infty]$ is called a rate function if for any $M < \infty$ the set $\{x : I(x) \leq M\}$ is compact in $\mathcal{Z}$. The sequence $\{Z^n\}$ satisfies the large deviation principle on $\mathcal{Z}$ with rate function I and sequence $r(n)$ if the following two conditions hold.

- Large Deviation Upper Bound: for each closed subset F of $\mathcal{Z}$

$$\limsup_{n\to\infty} r(n) \log P(Z^n \in F) \leq -\inf_{z\in F} I(z).$$

- Large Deviation Lower Bound: for each open subset G of $\mathcal{Z}$

$$\liminf_{n\to\infty} r(n) \log P(Z^n \in G) \geq -\inf_{z\in G} I(z).$$

Under significantly stronger assumptions, including the assumption that

$$\sup_{x\in\mathbb{R}^d} H_c(x,\alpha) < \infty$$

for all $\alpha \in \mathbb{R}^d$, it has been shown that $X^n(t)$ satisfies the large deviation principle on $C([0,1] : \mathbb{R}^d)$ with sequence $r(n) = 1/n$ and rate function

$$I_L(\phi) = \inf\left\{\int_0^1 L_c(\phi(s), u(s))ds : \phi(t) = x_0 + \int_0^t b(\phi(s))ds + \int_0^t u(s)ds, t \in [0,1]\right\},$$

where

$$L_c(x,\beta) = \sup_{\alpha\in\mathbb{R}^d}\{\langle\alpha,\beta\rangle - H_c(x,\alpha)\}$$

is the Legendre transform of $H_c(x,\alpha)$ [12, 20, 21, 22, 23].

Assume $a(n)$ satisfies

$$a(n) \to 0 \text{ and } a(n)\sqrt{n} \to \infty. \tag{2.6}$$

We define the rescaled difference

$$Y^n(t) = a(n)\sqrt{n}(X^n(t) - X^{n,0}(t)).$$

As noted in the introduction, the result stated below also holds with the interval $[0,1]$ replaced by $[0,T]$, $T \in (0,\infty)$. Let D denote the gradient operator.

THEOREM 2.3. *Assume Condition 2.1. Then $\{Y^n\}_{n\in\mathbb{N}}$ satisfies the large deviation principle on $C([0,1]:\mathbb{R}^d)$ with sequence $a(n)^2$ and rate function*

$$I_M(\phi) = \inf\left\{\frac{1}{2}\int_0^1 \|u(t)\|^2\,dt : \phi(t) = \int_0^t Db(X^0(s))\phi(s)ds \right.$$
$$\left. + \int_0^t A^{1/2}(X^0(s))u(s)ds, t \in [0,1]\right\}.$$

I_M is essentially the same as what one would obtain by using a linear approximation around the law of large numbers limit X^0 of the dynamics and a quadratic approximation of the costs in I_L. To prove the LDP, it suffices to show the Laplace principle [12, Theorem 1.2.3]

$$\begin{aligned}
&\lim_{n\to\infty} -a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] \\
&\quad = \inf_{u\in L^2([0,1]:\mathbb{R}^d)}\left\{\frac{1}{2}\int_0^1 \|u(s)\|^2\,ds + F\left(\phi^{A^{1/2}(X^0)u}\right)\right\},
\end{aligned} \tag{2.7}$$

where

$$\phi^u\,(t) = \int_0^t Db(X^0(s))\phi^u(s)ds + \int_0^t u(s)ds. \tag{2.8}$$

Note that

$$Y^n_{i+1} = Y^n_i + \frac{a(n)}{\sqrt{n}}\left(b(X^n_i) - b(X^{n,0}_i)\right) + \frac{a(n)}{\sqrt{n}}\upsilon_i(X^n_i), \quad Y^n_0 = 0.$$

For $\eta,\mu \in \mathcal{P}(\mathbb{R}^d)$ [the set of probability measures on $\mathcal{B}(\mathbb{R}^d)$] , the relative entropy of η with respect to μ is defined by

$$R(\eta\|\,\mu) \doteq \int_{\mathbb{R}^d} \log\left(\frac{d\eta}{d\mu}(x)\right)\eta(dx) \in [0,\infty]$$

if η is absolutely continuous with respect to μ, and $R(\eta\|\,\mu) \doteq \infty$ otherwise. For general properties of relative entropy we refer to [12, Section 1.4]. The variational formula [12, Proposition 1.4.2(a)] and chain rule [12, Theorem C.3.1] imply that

(2.9)
$$-a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] = \inf_{\eta} E\left[\sum_{i=0}^{n-1} a(n)^2 R(\eta_i\|\,\mu_{\bar{X}_i^n}) + F(\bar{Y}^n)\right]$$

for any bounded, continuous $F : C([0,1] : \mathbb{R}^d) \to \mathbb{R}$. Here $\eta \in \mathcal{P}((\mathbb{R}^d)^n)$ is the joint distribution of $(\bar{v}_0, \ldots, \bar{v}_{n-1})$, $\eta_i(\cdot)$ is the conditional distribution on $\bar{v}_i$ given $(\bar{v}_0, \ldots, \bar{v}_{i-1})$,

$$\bar{X}_{i+1}^n = \bar{X}_i^n + \frac{1}{n} b(\bar{X}_i^n) + \frac{1}{n}\bar{v}_i, \quad \bar{X}_0^n = x_0, \tag{2.10}$$

$$\bar{Y}_{i+1}^n = \bar{Y}_i^n + \frac{a(n)}{\sqrt{n}}\left(b(\bar{X}_i^n) - b(X_i^{n,0})\right) + \frac{a(n)}{\sqrt{n}}\bar{v}_i, \quad \bar{Y}_0^n = 0 \tag{2.11}$$

and, similar to (2.5), $\bar{X}^n(t)$ and $\bar{Y}^n(t)$ are the continuous time linear interpolations of $\{\bar{X}_i^n\}_{i=0,\ldots,n}$ and $\{\bar{Y}_i^n\}_{i=0,\ldots,n}$. Note that η_i depends on past values of the noise, but we suppress this dependence in the notation. We will prove (2.7) by proving the lower bound

$$\begin{aligned} &\liminf_{n\to\infty} -a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] \\ &\qquad \geq \inf_{u \in L^2([0,1]:\mathbb{R}^d)} \left\{\frac{1}{2}\int_0^1 \|u(s)\|^2\, ds + F\left(\phi^{A^{1/2}(X^0)u}\right)\right\} \end{aligned} \tag{2.12}$$

and the upper bound

$$\begin{aligned} &\limsup_{n\to\infty} -a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] \\ &\qquad \leq \inf_{u \in L^2([0,1]:\mathbb{R}^d)} \left\{\frac{1}{2}\int_0^1 \|u(s)\|^2\, ds + F\left(\phi^{A^{1/2}(X^0)u}\right)\right\}. \end{aligned} \tag{2.13}$$

We will use a tightness and weak convergence result in the proofs of both of these bounds, but first establish notation used in the rest of the paper.

CONSTRUCTION 2.4. Given a sequence of measures $\{\,\eta^n\}_{n\in\mathbb{N}}$ with each $\eta^n \in \mathcal{P}((\mathbb{R}^d)^n)$, define the following. Let $(\bar{v}_0^n, \ldots, \bar{v}_{n-1}^n)$ be random variables with distribution η^n, and define $\{\bar{X}_i^n\}_{i=0,\dots,n}$ and $\{\bar{Y}_i^n\}_{i=0,\dots,n}$ by (2.10) and (2.11). Let

$$\bar{X}^n(t) \doteq (i+1-nt)\bar{X}_i^n + (nt-i)\bar{X}_{i+1}^n$$

and

$$\bar{Y}^n(t) \doteq (i+1-nt)\bar{Y}_i^n + (nt-i)\bar{Y}_{i+1}^n$$

for $t \in [i/n, i/n+1/n], i = 0, \dots n-1$ be their continuous time linear interpolations. Define the conditional means of the noises

$$w^n(t) \doteq \int_{\mathbb{R}^d} y\eta_i^n(dy) \text{ for } t \in \left[\frac{i}{n}, \frac{i+1}{n}\right),$$

the amplified conditional means

$$\hat{w}^n(t) \doteq a(n)\sqrt{n}w^n(t),$$

and random measures on $\mathbb{R}^d \times [0,1]$ by

$$\hat{\eta}^n(dy \times dt) \doteq \delta_{\hat{w}^n(t)}(dy)dt = \delta_{a(n)\sqrt{n}w^n(t)}(dy)dt.$$

We will refer to this construction when given η^n to identify the associated $\bar{X}^n, \bar{Y}^n, \hat{w}^n$ and $\hat{\eta}^n$. Given $\nu \in \mathcal{P}(E_1 \times E_2)$, with each $E_i, i = 1,2$ a Polish space, let ν_2 denote the second marginal of ν, and let $\nu_{1|2}$ denote the conditional distribution on E_1 given a point in E_2.

THEOREM 2.5. *Let $\{\eta^n\}$ be a sequence of measures, with each $\eta^n \in \mathcal{P}((\mathbb{R}^d)^n)$, and define the corresponding random variables as in Construction 2.4. Assume that for some $K_E < \infty$*

$$\sup_{n\in\mathbb{N}} \left\{ a(n)^2 nE\left[\frac{1}{n}\sum_{i=0}^{n-1} R(\eta_i^n \| \mu_{\bar{X}_i^n})\right]\right\} \le K_E. \tag{2.14}$$

Then $\{(\hat{\eta}^n, \bar{Y}^n)\}_{n\in\mathbb{N}}$ is tight in $\mathcal{P}(\mathbb{R}^d \times [0,1]) \times C([0,1] : \mathbb{R}^d)$. Consider a subsequence (keeping the index n for convenience) such that $\{(\hat{\eta}^n, \bar{Y}^n)\}$ converges weakly to $(\hat{\eta}, \hat{Y})$. Then with probability 1 $\hat{\eta}_2(dt)$ is Lebesgue measure and

$$\hat{Y}(t) = \int_0^t Db(X^0(s))\hat{Y}(s)ds + \int_0^t \hat{w}(s)ds, \tag{2.15}$$

where

$$\hat{w}(t) = \int_{\mathbb{R}^d} y\hat{\eta}_{1|2}(dy\,|t).$$

In addition,
(2.16)
$$\liminf_{n\to\infty} a(n)^2 nE\left[\frac{1}{n}\sum_{i=0}^{n-1} R(\eta_i^n \| \mu_{\bar{X}_i^n})\right] \geq E\left[\int_0^1 \frac{1}{2}\|\hat{w}(s)\|^2_{A^{-1}(X^0(s))}\, ds\right].$$

3. Proof of Theorem 2.5. Assume that the bound (2.14) holds. We will show tightness of the $\{\hat{\eta}^n\}$ measures using the following lemma.

LEMMA 3.1. *Assume Condition 2.1 and let*

$$L_c(x,\beta) = \sup_{\alpha\in\mathbb{R}^d}\{\langle\alpha,\beta\rangle - H_c(x,\alpha)\} \tag{3.1}$$

be the Legendre transform of $H_c(x,\cdot)$. Then for any $x \in \mathbb{R}^d$ and $\eta \in \mathcal{P}(\mathbb{R}^d)$

$$R(\eta\| \mu_x) \geq L_c\left(x, \int_{\mathbb{R}^d} y\eta(dy)\right).$$

PROOF. While the result is likely known we could not locate a proof (see [12, Lemma 6.2.3(f)] for a proof when $H_c(x,\alpha)$ is finite for all $\alpha \in \mathbb{R}^d$), and so for completeness the details are provided. If $R(\eta\|\mu_x) = \infty$ the lemma is automatically true, so we assume $R(\eta\|\mu_x) < \infty$. Define $\ell(b) = b\log b - b + 1$ and note that for $a, b \geq 0$

$$ab \leq e^a + \ell(b). \tag{3.2}$$

From (2.1) we have

$$\int_{\mathbb{R}^d} e^{\frac{\lambda}{2^d}\|y\|}\mu_x(dy) \leq 2^d e^{dK_{\mathrm{mgf}}} < \infty.$$

Therefore

$$\begin{aligned}
&\int_{\mathbb{R}^d} \frac{\lambda}{2^d}\|y\| \frac{d\eta}{d\mu_x}(y)\mu_x(dy) \\
&\quad \leq \int_{\mathbb{R}^d} e^{\frac{\lambda}{2^d}\|y\|}\mu_x(dy) + \int_{\mathbb{R}^d} \ell\left(\frac{d\eta}{d\mu}(y)\right)\mu_x(dy) \\
&\quad \leq 2^d e^{dK_{\mathrm{mgf}}} + R(\eta\|\mu_x),
\end{aligned}$$

and consequently for any $\alpha \in \mathbb{R}^d$

$$\int_{\mathbb{R}^d} \|\alpha\|\,\|y\| \frac{d\eta}{d\mu_x}(y)\mu_x(dy) \leq \frac{2^d\|\alpha\|}{\lambda}\left(2^d e^{dK_{\mathrm{mgf}}} + R(\eta\|\mu_x)\right) < \infty. \tag{3.3}$$

Define the bounded, continuous function

$$F_K(y,\alpha) = \begin{cases} \langle \alpha, y\rangle & \text{if } |\langle \alpha, y\rangle| \le K \\ \frac{K\langle \alpha,y\rangle}{|\langle \alpha,y\rangle|} & \text{otherwise,} \end{cases}$$

and note that (3.3) and dominated convergence give

$$\lim_{K\to\infty} \int_{\mathbb{R}^d} F_K(y,\alpha)\eta(dy) = \left\langle \alpha, \int_{\mathbb{R}^d} y\eta(dy) \right\rangle .$$

In addition, dominated convergence gives

$$\lim_{K\to\infty} \int_{\{y:\langle \alpha,y\rangle<0\}} e^{F_K(y,\alpha)}\mu_x(dy) = \int_{\{y:\langle \alpha,y\rangle<0\}} e^{\langle \alpha,y\rangle}\mu_x(dy)$$

and monotone convergence gives

$$\lim_{K\to\infty} \int_{\{y:\langle \alpha,y\rangle\ge 0\}} e^{F_K(y,\alpha)}\mu_x(dy) = \int_{\{y:\langle \alpha,y\rangle\ge 0\}} e^{\langle \alpha,y\rangle}\mu_x(dy) ,$$

so

$$\lim_{K\to\infty} \log \left(\int_{\mathbb{R}^d} e^{F_K(y,\alpha)}\mu_x(dy) \right) = H_c(x,\alpha).$$

By the Donsker-Varadhan variational formula [12, Lemma 1.4.3(a)]

$$R(\eta \| \mu_x) \ge \int_{\mathbb{R}^d} F_K(y,\alpha)\eta(dy) - \log \left(\int_{\mathbb{R}^d} e^{F_K(y,\alpha)}\mu_x(dy) \right)$$

for all $K<\infty$ and $\alpha\in\mathbb{R}^d$, and so

$$R(\eta \| \mu_x) \ge \sup_{\alpha\in\mathbb{R}^d} \left\{ \left\langle \alpha, \int_{\mathbb{R}^d} y\eta(dy) \right\rangle - H_c(x,\alpha) \right\} = L_c\left(x, \int_{\mathbb{R}^d} y\eta(dy) \right) ,$$

which completes the proof of the lemma. □

The lemma implies the following theorem, which in turn will give tightness of $\{\hat{\eta}^n\}$.

Theorem 3.2. *Assume Condition 2.1 and (2.14). For the processes $\{w^n\}$ obtained in Construction 2.4*

$$\sup_{n\in\mathbb{N}} E\left[\int_0^1 a(n)\sqrt{n}\, \|w^n(s)\|\, ds \right] < \infty.$$

In addition, $\{a(n)\sqrt{n}w^n(\cdot)\}_{n\in\mathbb{N}}$ is uniformly integrable in the sense that

$$\lim_{C\to\infty} \limsup_{n\to\infty} E\left[\int_0^1 1_{\{a(n)\sqrt{n}\|w^n(s)\|>C\}} a(n)\sqrt{n}\, \|w^n(s)\|\, ds \right] = 0.$$

PROOF. We use the following inequality. Let $G \doteq (\lambda_{DA} \min_{n\in\mathbb{N}}\{a(n)\sqrt{n}\})^2$ [recall (2.6)], so that $\lambda_{DA} \geq \sqrt{G}/a(n)\sqrt{n}$ for all n. Define L_c by (3.1). Let $\bar{K} \doteq \lambda_{DA}K_{DA} + K_A/2$. Then with e_i denoting the standard unit vectors

$$\begin{aligned}
&a(n)^2 nL_c(x,\beta) \\
&\quad = \sup_{\alpha\in\mathbb{R}^d} \left[a(n)\sqrt{n}\left\langle \alpha, a(n)\sqrt{n}\beta\right\rangle - a(n)^2 nH_c(x,\alpha)\right] \\
&\quad \geq \pm a(n)\sqrt{n}\left\langle \frac{\sqrt{G}}{a(n)\sqrt{n}}e_i, a(n)\sqrt{n}\beta\right\rangle - a(n)^2 nH_c\left(x, \pm\frac{\sqrt{G}}{a(n)\sqrt{n}}e_i\right) \\
&\quad \geq \pm\sqrt{G}a(n)\sqrt{n}\beta_i - \frac{1}{2}G\left\|A(x)\right\| - G\lambda_{DA}K_{DA} \\
&\quad \geq \pm\sqrt{G}a(n)\sqrt{n}\beta_i - G\bar{K},
\end{aligned}$$

where the first inequality follows from making a specific choice of α and the second uses (2.4). Therefore

$$da(n)^2 nL_c(x,\beta) + dG\bar{K} \geq \sqrt{G}a(n)\sqrt{n}\left\|\beta\right\|. \tag{3.4}$$

Using the bound on L_c from Lemma 3.1 together with (2.14) and the last display,

$$\begin{aligned}
&d\left(\frac{K_E}{\sqrt{G}} + \sqrt{G}\bar{K}\right) \\
&\quad \geq \frac{da(n)^2 n}{\sqrt{G}}E\left[\int_0^1 L_c\left(\bar{X}^n\left(\frac{\lfloor ns\rfloor}{n}\right), w^n(s)\right)ds\right] + d\sqrt{G}\bar{K} \\
&\quad \geq E\left[\int_0^1 a(n)\sqrt{n}\left\|w^n(s)\right\| ds\right].
\end{aligned} \tag{3.5}$$

For the uniform integrability, let $C \in (1,\infty)$ be arbitrary and consider n large enough that

$$\min\{\lambda_{DA}, 1\} \geq \frac{\sqrt{C}}{a(n)\sqrt{n}}.$$

Since $\lambda_{DA} \geq 1/a(n)\sqrt{n}$ the derivation leading (3.5) holds for $G = 1$, and therefore

$$E\left[\int_0^1 a(n)\sqrt{n}\left\|w^n(s)\right\| ds\right] \leq K^* \doteq d\left(K_E + \frac{1}{2}K_A + \lambda_{DA}K_{DA}\right),$$

which implies

$$E\left[\int_0^1 1_{\{a(n)\sqrt{n}\|w^n(s)\|>C\}}ds\right] \leq \frac{K^*}{C}.$$

Since $\lambda_{DA} \geq \sqrt{C}/a(n)\sqrt{n}$ the estimate (3.4) holds with G replaced by C, and then the last display and (3.5) give

$$\begin{aligned}
&\sqrt{C} E\left[\int_0^1 1_{\{a(n)\sqrt{n}\|w^n(s)\|>C\}} a(n)\sqrt{n}\,\|w^n(s)\|\,ds\right] \\
&\quad \leq da(n)^2 nE\left[\int_0^1 L_c\left(\bar{X}^n\left(\frac{\lfloor ns\rfloor}{n}\right), w^n(s)\right)ds\right] \\
&\qquad + Cd\bar{K}E\left[\int_0^1 1_{\{a(n)\sqrt{n}\|w^n(s)\|>C\}} ds\right] \\
&\quad \leq K^* d\left(1+\bar{K}\right).
\end{aligned}$$

We conclude that

$$\lim_{C\to\infty}\limsup_{n\to\infty} E\left[\int_0^1 1_{\{a(n)\sqrt{n}\|w^n(s)\|>C\}} a(n)\sqrt{n}\,\|w^n(s)\|\,ds\right] = 0,$$

which is the claimed uniform integrability. □

We continue with the proof of Theorem 2.5. Note that $g(y,t) = \|y\|$ is a tightness function on $\mathbb{R}^d \times [0,1]$, so by [12, Theorem A.3.17]

$$G(\eta) = \int_{\mathbb{R}^d\times[0,1]} \|y\|\,\eta(dy\times dt)$$

is a tightness function on $\mathcal{P}(\mathbb{R}^d \times [0,1])$ and

$$\bar{G}(\gamma) = \int_{\mathcal{P}\left(\mathbb{R}^d\times[0,1]\right)} \int_{\mathbb{R}^d\times[0,1]} \|y\|\,\eta(dy\times dt)\gamma(d\eta)$$

is a tightness function on $\mathcal{P}(\mathcal{P}(\mathbb{R}^d \times [0,1]))$. Since

$$\begin{aligned}
\sup_{n\in\mathbb{N}} EG(\hat{\eta}^n) &= \sup_{n\in\mathbb{N}} E\left[\int \|y\|\,\hat{\eta}^n(dy\times dt)\right] \\
&= \sup_{n\in\mathbb{N}} E\left[\int_0^1 a(n)\sqrt{n}\,\|w^n(s)\|\,ds\right] < \infty,
\end{aligned}$$

$\{\hat{\eta}^n\}$ is tight and consequently there is a subsequence of $\{\hat{\eta}^n\}$ which converges weakly. To simplify notation we retain n as the index of this convergent subsequence, and denote the weak limit of $\{\hat{\eta}^n\}$ by $\hat{\eta}$. Note that for all n the second marginal of $\hat{\eta}^n(dy\times dt)$, which we denote by $\hat{\eta}_2^n(dt)$, is Lebesgue measure, and therefore $\hat{\eta}_2(dt)$ is Lebesgue measure with probability 1.

Our aim is to show that $\bar{Y}^n(t) \to \hat{Y}(t)$ weakly in $C([0,1]:\mathbb{R}^d)$, where $\hat{Y}(t)$ is given by (2.15) in terms of the weak limit $\hat{\eta}$. To achieve this we introduce the following processes which serve as intermediate steps. Let $\check{Y}_0^n = 0$ and

$$\check{Y}_{i+1}^n = \check{Y}_i^n + \frac{a(n)}{\sqrt{n}}\left(b\left(X_i^{n,0} + \frac{1}{a(n)\sqrt{n}}\check{Y}_i^n\right) - b\left(X_i^{n,0}\right)\right) + \frac{a(n)}{\sqrt{n}}w^n\left(\frac{i}{n}\right),$$

together with its continuous time linear interpolation defined for $t \in [i/n, i/n+1/n]$ by

$$\check{Y}^n(t) = (i+1-nt)\check{Y}_i^n + (nt-i)\check{Y}_{i+1}^n.$$

Also let

$$\hat{Y}^n(t) = \int_0^t Db\left(X^0(s)\right)\hat{Y}^n(s)ds + \int_0^t \hat{w}^n(s)ds \tag{3.6}$$

where

$$\hat{w}^n(t) = \int_{\mathbb{R}^d} y\hat{\eta}_{1|2}^n(dy\,|t)$$

as in Construction 2.4. These are both random variables taking values in $C([0,1]:\mathbb{R}^d)$. Note that $\bar{Y}^n$ differs from $\check{Y}^n$ because $\bar{Y}^n$ is driven by the actual noises and $\check{Y}^n$ is driven by their conditional means. While the driving terms of $\hat{Y}^n$ and $\check{Y}^n$ are the same [recall that $a(n)\sqrt{n}w^n(t) = \hat{w}^n(t)$], they differ in that $\check{Y}^n$ is still a linear interpolation of a discrete time process whereas $\hat{Y}^n$ satisfies an ODE. The goal is to show that along the subsequence where $\hat{\eta}^n \to \hat{\eta}$ weakly

$$\bar{Y}^n - \check{Y}^n \to 0, \quad \check{Y}^n - \hat{Y}^n \to 0, \quad \text{and} \quad \hat{Y}^n \to \hat{Y}$$

in $C([0,1]:\mathbb{R}^d)$, all in distribution. To show $\hat{Y}^n \to \hat{Y}$ we show that $\{\hat{Y}^n\}$ is tight in $C([0,1]:\mathbb{R}^d)$ and use the mapping defined by (3.6) from $\int_0^\cdot \hat{w}^n$ to $\hat{Y}^n$. Recall that $\sup_{x\in\mathbb{R}^d}\|Db(x)\| \le K_b$. The following lemma is an easy consequence of Gronwall's inequality.

LEMMA 3.3. *Let $u \in L^1([0,1]:\mathbb{R}^d)$ be arbitrary and ϕ^u be defined as in (2.8). Then for $0 \le s \le t \le 1$*

$$\|\phi^u(t) - \phi^u(s)\| \le (t-s)K_b e^{K_b}\int_0^1 \|u(r)\|\,dr + \int_s^t \|u(r)\|\,dr.$$

With this lemma and the uniform integrability of $\{\hat{\eta}^n\}$ given in Theorem 3.2, tightness follows.

LEMMA 3.4. *Assume Condition 2.1 and (2.14). The sequence $\{\hat{Y}^n\}$ defined in (3.6) in terms of the measures $\{\eta^n\}$ via Construction 2.4 is tight in $C([0,1]:\mathbb{R}^d)$, as is $\{\int_0^\cdot \hat{w}^n ds\}$.*

PROOF. It suffices to show that for any $\varepsilon > 0$ there is $\delta > 0$ such that

$$\limsup_{n\to\infty} P\left(\sup_{|s-t|\le\delta}\left\|\hat{Y}^n(t)-\hat{Y}^n(s)\right\| > \varepsilon\right) < \varepsilon.$$

Since $\hat{\eta}^n$ is a measure with a point mass located at $\hat{w}^n(t)$,

$$\begin{aligned} T(C) &\doteq \limsup_{n\to\infty} E\left[\int_0^1 1_{\{\|\hat{w}^n(t)\|>C\}}\|\hat{w}^n(t)\|\,dt\right] \\ &= \limsup_{n\to\infty} E\left[\int_{\{\|y\|>C\}}\|y\|\,\hat{\eta}^n(dy\times dt)\right]. \end{aligned}$$

By Theorem 3.2 $T(C)\to 0$ as $C\to\infty$. Let $K_\eta = \sup_{n\in\mathbb{N}} E\int_0^1\|\hat{w}^n(t)\|\,dt$, which is finite by Theorem 3.2, and let $\varepsilon > 0$ be arbitrary. Then for any $s<t$ satisfying $t-s\le\delta$ Lemma 3.3 implies

$$\left\|\hat{Y}^n(t)-\hat{Y}^n(s)\right\| \le \delta K_b e^{K_b}\int_0^1\|\hat{w}^n(r)\|\,dr + \int_s^t\|\hat{w}^n(r)\|\,dr.$$

Since

$$\int_s^t\|\hat{w}^n(r)\|\,dr \le C\delta + \int_0^1 1_{\{\|\hat{w}^n(r)\|>C\}}\|\hat{w}^n(r)\|\,dr,$$

it follows that

$$\begin{aligned} \left\|\hat{Y}^n(t)-\hat{Y}^n(s)\right\| &\le \delta\left(C + K_b e^{K_b}\int_0^1\|\hat{w}^n(r)\|\,dr\right) \\ &\quad + \int_0^1 1_{\{\|\hat{w}^n(r)\|>C\}}\|\hat{w}^n(r)\|\,dr. \end{aligned}$$

Hence by Markov's inequality

$$\begin{aligned} &\limsup_{n\to\infty} P\left(\sup_{|s-t|\le\delta}\left\|\hat{Y}^n(t)-\hat{Y}^n(s)\right\| > \varepsilon\right) \\ &\quad\le \frac{\delta}{\varepsilon}\limsup_{n\to\infty} E\left[\left(C + K_b e^{K_b}\int_0^1\|\hat{w}^n(r)\|\,dr\right)\right] \\ &\qquad + \frac{1}{\varepsilon}\limsup_{n\to\infty} E\left[\int_0^1 1_{\{\|\hat{w}^n(r)\|>C\}}\|\hat{w}^n(r)\|\,dr\right] \end{aligned}$$

$$\leq \frac{\delta}{\varepsilon}(C + K_b e^{K_b} K_\eta) + \frac{1}{\varepsilon} T(C).$$

Choose $C < \infty$ such that $T(C) < \varepsilon^2/2$ and then choose $\delta > 0$ so that the $\delta(C + K_b e^{K_b} K_\eta) < \varepsilon^2/2$. This shows the tightness of $\{\hat{Y}^n\}$. The tightness of $\{\int_0^\cdot \hat{w}^n ds\}$ is simpler, and follows from the bound

$$\limsup_{n\to\infty} P\left(\sup_{|s-t|\leq\delta} \int_s^t \|\hat{w}^n(r)\|\, dr > \varepsilon\right) \leq \delta\frac{C}{\varepsilon} + \frac{1}{\varepsilon} T(C).$$

□

We still need to show that $\hat{Y}^n$ converges to $\hat{Y}$. This also relies on the uniform integrability given by Theorem 3.2.

LEMMA 3.5. *Assume Condition 2.1 and (2.14). Let the sequence $\{\hat{Y}^n(t)\}$ be defined by (3.6), consider a convergent subsequence $\{(\hat{Y}^n, \hat{\eta}^n)\}$ with limit $(\hat{Y}^*, \hat{\eta})$, and let $\hat{Y}(t)$ be defined by (2.15). Then w.p.1 $\hat{Y}^* = \hat{Y}$.*

PROOF. We can write

$$\hat{Y}^n(t) = \int_0^t Db(X^0(s))\hat{Y}^n(s)ds + \int_0^t \int_{\mathbb{R}^d} y\hat{\eta}^n(dy \times ds).$$

Using the uniform integrability proved in Theorem 3.2 and that $\hat{\eta}_2$ is Lebesgue measure w.p.1, sending $n \to \infty$ and using the definition of $\hat{w}$ gives

$$\begin{aligned}
\hat{Y}^*(t) &= \int_0^t Db(X^0(s))\hat{Y}^*(s)ds + \int_0^t \int_{\mathbb{R}^d} y\hat{\eta}(dy \times ds) \\
&= \int_0^t Db(X^0(s))\hat{Y}^*(s)ds + \int_0^t \hat{w}(s)ds.
\end{aligned}$$

By uniqueness of the solution, $\hat{Y}^* = \hat{Y}$ follows. □

It remains to show $\bar{Y}^n - \check{Y}^n \to 0$ and $\check{Y}^n - \hat{Y}^n \to 0$. We begin with $\bar{Y}^n - \check{Y}^n \to 0$. Recall that the difference between $\bar{Y}^n$ and $\check{Y}^n$ is that the first is driven by the actual noises and the second is driven by their conditional means. The following theorem is a law of large numbers type result for the difference between the noises and their conditional means, and is the most complicated part of the analysis.

THEOREM 3.6. *Assume Condition 2.1 and (2.14). Consider the sequence* $\{\bar{v}_i^n\}_{i=0,\dots,n-1}$ *of controlled noises and* $\{w^n(i/n)\}_{i=0,\dots,n-1}$ *of means of the controlled noises as in Construction 2.4. For* $i \in \{1,\dots,n\}$ *let*

$$W_i^n \doteq \frac{1}{n}\sum_{j=0}^{i-1} a(n)\sqrt{n}\left(\bar{v}_i^n - w^n(i/n)\right).$$

Then for any $\delta > 0$

$$\lim_{n\to\infty} P\left[\max_{i\in\{1,\dots,n\}} \|W_i^n\| \ge \delta\right] = 0.$$

PROOF. According to (2.14)

$$\frac{1}{n}\sum_{i=0}^{n-1} E[R(\eta_i^n \,\|\, \mu_{\bar{X}_i^n})] \le \frac{K_E}{a^2(n)n}.$$

Because of this the (random) Radon-Nikodym derivatives

$$f_i^n(y) = \frac{d\eta_i^n}{d\mu_{\bar{X}_i^n}}(y)$$

are well defined and can be selected in a measurable way. We will control the magnitude of the noise when the Radon-Nikodym derivative is large by bounding

$$\frac{1}{n}\sum_{i=0}^{n-1} E[1_{\{f_i^n(\bar{v}_i^n)\ge r\}} \|\bar{v}_i^n\|]$$

for large r.

From the bound on the moment generating function (2.1),

$$\sup_{x\in\mathbb{R}^d} \int_{\mathbb{R}^d} e^{\frac{\lambda}{2^d}\|y\|}\mu_x(dy) \le 2^d e^{dK_{\mathrm{mgf}}}. \tag{3.7}$$

Let

$$\sigma = \min\{\lambda/2^{d+1}, 1\} \tag{3.8}$$

and recall the definition $\ell(b) \doteq b\log b - b + 1$. Then

$$\frac{1}{n}\sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\ge r\}} \|\bar{v}_i^n\|\right] = \frac{1}{n}\sum_{i=0}^{n-1} E\left[\int_{\{y: f_i^n(y)\ge r\}} \|y\|\, f_i^n(y)\mu_{\bar{X}_i^n}(dy)\right]$$

and the bound $ab \leq e^a + \ell(b)$ for $a, b \geq 0$ with $a = \sigma \|y\|$ and $b = f_i^n(y)$ gives that for all i

$$E\left[\int_{\{y:f_i^n(y)\geq r\}} \|y\| f_i^n(y)\mu_{\bar{X}_i^n}(dy)\right]$$
$$\leq \frac{1}{\sigma}E\left[\int_{\{y:f_i^n(y)\geq r\}} e^{\sigma\|y\|}\mu_{\bar{X}_i^n}(dy)\right] + \frac{1}{\sigma}E\left[\int_{\{y:f_i^n(y)\geq r\}} \ell(f_i^n(y))\mu_{\bar{X}_i^n}(dy)\right].$$

Since $\ell(b) \geq 0$ for all $b \geq 0$

$$E\left[\int_{\left\{y:f_i^n(y)\geq r\right\}} \ell\left(f_i^n\left(y\right)\right)\mu_{\bar{X}_i^n}\left(dy\right)\right] \leq E\left[\int_{\mathbb{R}^d} \ell(f_i^n(y))\mu_{\bar{X}_i^n}(dy)\right]$$
$$= E[R(\eta_i^n \| \mu_{\bar{X}_i^n})],$$

and by Hölder's inequality (recall (3.7) and (3.8))

$$E\left[\int_{\{y:f_i^n(y)\geq r\}} e^{\sigma\|y\|}\mu_{\bar{X}_i^n}(dy)\right]$$
$$\leq E\left[\left(\int_{\mathbb{R}^d} 1_{\{f_i^n(y)\geq r\}}\mu_{\bar{X}_i^n}(dy)\right)^{\frac{1}{2}}\left(\int_{\mathbb{R}^d} e^{2\sigma\|y\|}\mu_{\bar{X}_i^n}(dy)\right)^{\frac{1}{2}}\right]$$
$$\leq E\left[\mu_{\bar{X}_i^n}(\{y : f_i^n(y) \geq r\})^{\frac{1}{2}}\right]\left(2^d e^{dK_{\mathrm{mgf}}}\right)^{\frac{1}{2}}.$$

In addition Markov's inequality gives for $r \geq e^{-1}$

$$\mu_{\bar{X}_i^n}(\{y : f_i^n(y) \geq r\}) \leq \frac{1}{r\log r}\int \log(f_i^n(y))f_i^n(y)\mu_{\bar{X}_i^n}(dy) = \frac{R(\eta_i^n \| \mu_{\bar{X}_i^n})}{r\log r}.$$

Therefore

$$\frac{1}{n}\sum_{i=0}^{n-1} E\left[\int_{\{f_i^n(y)\geq r\}} \|y\| f_i^n(y)\mu_{\bar{X}_i^n}(dy)\right]$$
$$\leq \frac{1}{\sigma}\left(2^d e^{dK_{\mathrm{mgf}}}\right)^{\frac{1}{2}}\frac{1}{n}\sum_{i=0}^{n-1} E\left[\left(\frac{R(\eta_i^n \| \mu_{\bar{X}_i^n})}{r\log r}\right)^{\frac{1}{2}}\right] + \frac{1}{\sigma}\frac{1}{n}\sum_{i=0}^{n-1} E[R(\eta_i^n \| \mu_{\bar{X}_i^n})].$$

Since by Jensen's inequality

$$\frac{1}{n}\sum_{i=0}^{n-1} E\left[\left(\frac{R(\eta_i^n \| \mu_{\bar{X}_i^n})}{r\log r}\right)^{\frac{1}{2}}\right] \leq \left(\frac{1}{r\log r}\right)^{\frac{1}{2}}\left(\frac{1}{n}\sum_{i=0}^{n-1} E[R(\eta_i^n \| \mu_{\bar{X}_i^n})]\right)^{\frac{1}{2}},$$

we obtain the overall bound

$$\begin{aligned}
&\frac{1}{n}\sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r\}}\|\bar{v}_i^n\|\right] \\
&\quad\leq \frac{1}{\sigma}\left(2^d e^{dK_{\mathrm{mgf}}}\right)^{\frac{1}{2}}\left(\frac{1}{r\log r}\right)^{\frac{1}{2}}\left(\frac{1}{n}\sum_{i=0}^{n-1} E[R(\eta_i^n \| \mu_{\bar{X}_i^n})]\right)^{\frac{1}{2}} \\
&\qquad + \frac{1}{\sigma}\frac{1}{n}\sum_{i=0}^{n-1} E[R(\eta_i^n \| \mu_{\bar{X}_i^n})] \\
&\quad\leq \frac{1}{\sigma}\frac{K_E^{\frac{1}{2}}}{a(n)\sqrt{n}}\left(2^d e^{dK_{\mathrm{mgf}}}\right)^{\frac{1}{2}}\left(\frac{1}{r\log r}\right)^{\frac{1}{2}} + \frac{1}{\sigma}\frac{K_E}{a(n)^2 n}. \qquad (3.9)
\end{aligned}$$

Using this result we can complete the proof. Define

$$\xi_i^{n,r} \doteq \begin{cases} \bar{v}_i^n & \text{if } f_i^n(\bar{v}_i^n) < r \\ 0 & \text{otherwise.} \end{cases}$$

For any for any $\delta > 0$

$$\begin{aligned}
&P\left\{\max_{k=0,\ldots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k} a(n)\sqrt{n}\left(\bar{v}_i^n - w^n\left(\frac{i}{n}\right)\right)\right\| \geq 3\delta\right\} \\
&\quad\leq P\left\{\max_{k=0,\ldots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k} a(n)\sqrt{n}(\bar{v}_i^n - \xi_i^{n,r})\right\| \geq \delta\right\} \\
&\qquad + P\left\{\max_{k=0,\ldots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k} a(n)\sqrt{n}\left(\xi_i^{n,r} - \int_{\{y:f_i^n(y)<r\}} y\eta_i^n(dy)\right)\right\| \geq \delta\right\} \\
&\qquad + P\left\{\max_{k=0,\ldots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k} a(n)\sqrt{n}\left(w^n\left(\frac{i}{n}\right) - \int_{\{y:f_i^n(y)<r\}} y\eta_i^n(dy)\right)\right\| \geq \delta\right\}.
\end{aligned}$$

The first term satisfies

$$\begin{aligned}
&P\left\{\max_{k=0,\ldots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k} a(n)\sqrt{n}(\bar{v}_i^n - \xi_i^{n,r})\right\| \geq \delta\right\} \\
&\quad\leq \frac{1}{\delta}a(n)\sqrt{n}\frac{1}{n}\sum_{i=0}^{n-1} E\left[\|\bar{v}_i^n - \xi_i^{n,r}\|\right] \\
&\quad= \frac{1}{\delta}a(n)\sqrt{n}\frac{1}{n}\sum_{i=0}^{n-1} E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r\}}\|\bar{v}_i^n\|\right].
\end{aligned}$$

The second term involves a submartingale, and the first inequality in the next display follows from Doob's submartingale inequality. The second inequality uses a conditioning argument and that for any integrable random variable Z, $E[Z-EZ]^2 \le EZ^2$. We have

$$\begin{aligned}
&P\left\{\max_{k=0,\dots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k}a(n)\sqrt{n}\left(\xi_i^{n,r}-\int_{\{y:f_i^n(y)<r\}}y\eta_i^n(dy)\right)\right\|\ge\delta\right\}\\
&\le\frac{1}{\delta^2}E\left[\left\|\frac{1}{n}\sum_{i=0}^{n-1}a(n)\sqrt{n}\left(\xi_i^{n,r}-\int_{\{y:f_i^n(y)<r\}}y\eta_i^n(dy)\right)\right\|^2\right]\\
&=\frac{1}{\delta^2}\frac{a(n)^2}{n}\sum_{i=0}^{n-1}E\left[\left\|\left(\xi_i^{n,k}-\int_{\{y:f_i^n(y)<r\}}y\eta_i^n(dy)\right)\right\|^2\right]\\
&\le\frac{1}{\delta^2}\frac{a(n)^2}{n}\sum_{i=0}^{n-1}E\left[\left\|\xi_i^{n,k}\right\|^2\right]\\
&=\frac{1}{\delta^2}\frac{a(n)^2}{n}\sum_{i=0}^{n-1}E\left[\int_{\{y:f_i^n(y)<r\}}\|y\|^2 f_i^n(y)\mu_{\bar{X}_i^n}(dy)\right]\\
&\le\frac{r}{\delta^2}\frac{a(n)^2}{n}\sum_{i=0}^{n-1}E\left[\int_{\mathbb{R}^d}\|y\|^2\mu_{\bar{X}_i^n}(dy)\right]\\
&\le\frac{r}{\delta^2}a(n)^2K_{\mu,2},
\end{aligned}$$

where

$$K_{\mu,2}=\sup_{x\in\mathbb{R}^d}\int_{\mathbb{R}^d}\|y\|^2\mu_x(dy)<\infty,$$

and the finiteness is due to (2.1). We can use Jensen's inequality with the third term and get the same bound that was shown for the first. We have

$$\begin{aligned}
&P\left\{\max_{k=0,\dots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k}a(n)\sqrt{n}\left(w^n\left(\frac{i}{n}\right)-\int_{\{y:f_i^n(y)<r\}}y\eta_i^n(dy)\right)\right\|\ge\delta\right\}\\
&\le\frac{1}{\delta}a(n)\sqrt{n}\frac{1}{n}\sum_{i=0}^{n-1}E\left[\left\|\left(w^n\left(\frac{i}{n}\right)-\int_{\{y:f_i^n(y)<r\}}y\eta_i^n(dy)\right)\right\|\right]\\
&=\frac{1}{\delta}a(n)\sqrt{n}\frac{1}{n}\sum_{i=0}^{n-1}E\left[\left\|\int_{\{y:f_i^n(y)\ge r\}}y\eta_i^n(dy)\right\|\right]\\
&\le\frac{1}{\delta}a(n)\sqrt{n}\frac{1}{n}\sum_{i=0}^{n-1}E\left[\int_{\{y:f_i^n(y)\ge r\}}\|y\|\,\eta_i^n(dy)\right]\\
&=\frac{1}{\delta}a(n)\sqrt{n}\frac{1}{n}\sum_{i=0}^{n-1}E\left[1_{\{f_i^n(\bar{v}_i^n)\ge r\}}\left\|\bar{v}_i^n\right\|\right].
\end{aligned}$$

Combining the bounds for these three terms with (3.9) gives

$$
\begin{aligned}
P&\left\{\max_{k=0,\dots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k}a(n)\sqrt{n}\left(\bar{v}_i^n-w^n\left(\frac{i}{n}\right)\right)\right\|\geq 3\delta\right\}\\
&\leq\frac{2}{\delta}a(n)\sqrt{n}\frac{1}{n}\sum_{i=0}^{n-1}E\left[1_{\{f_i^n(\bar{v}_i^n)\geq r]}\|\bar{v}_i^n\|\right]+\frac{r}{\delta^2}a(n)^2K_{\mu,2}\\
&\leq\frac{2}{\sigma\delta}K_E^{\frac{1}{2}}\left(2^de^{dK_{\mathrm{mgf}}}\right)^{\frac{1}{2}}\left(\frac{1}{r\log r}\right)^{\frac{1}{2}}+\frac{2}{\sigma\delta}\frac{K_E}{a(n)\sqrt{n}}+a(n)^2\frac{r}{\delta^2}K_{\mu,2}.
\end{aligned}
$$

Choosing $r=1/a(n)$ and using $a(n)\to 0, a(n)\sqrt{n}\to\infty$ gives

$$
P\left\{\max_{k=0,\dots,n-1}\left\|\frac{1}{n}\sum_{i=0}^{k}a(n)\sqrt{n}\left(\bar{v}_i^n-w^n\left(\frac{i}{n}\right)\right)\right\|\geq 3\delta\right\}\to 0
$$

as $n\to\infty$, which completes the proof. $\square$

This theorem, combined with the following discrete version of Gronwall's inequality, will allow us to prove $\bar{Y}^n-\check{Y}^n\to 0$.

LEMMA 3.7. *If $\{a_n\}$, $\{b_n\}$, and $\{c_n\}$ are nonnegative sequences defined for $n=0,1,\dots$ and satisfying*

$$
a_n\leq c_n+\sum_{k=0}^{n-1}b_ka_k,
$$

then

$$
a_n\leq c_n+\sum_{k=0}^{n-1}b_kc_k\exp\left\{\sum_{i=k+1}^{n-1}b_i\right\}.
$$

THEOREM 3.8. *Under the conditions of Theorem 3.6 $\check{Y}^n-\bar{Y}^n\to 0$ in probability.*

PROOF. Recall that

$$
\bar{Y}_k^n=\sum_{i=0}^{k-1}\frac{a(n)}{\sqrt{n}}\left(b\left(X_i^{n,0}+\frac{1}{a(n)\sqrt{n}}\bar{Y}_i^n\right)-b\left(X_i^{n,0}\right)\right)+\sum_{i=0}^{k-1}\frac{a(n)}{\sqrt{n}}\bar{v}_i^n
$$

and

$$
\check{Y}_k^n=\sum_{i=0}^{k-1}\frac{a(n)}{\sqrt{n}}\left(b\left(X_i^{n,0}+\frac{1}{a(n)\sqrt{n}}\check{Y}_i^n\right)-b\left(X_i^{n,0}\right)\right)+\sum_{i=0}^{k-1}\frac{a(n)}{\sqrt{n}}w^n\left(\frac{i}{n}\right),
$$

so with W_k^n defined as in Theorem 3.6

$$\left\|\bar{Y}_k^n - \check{Y}_k^n\right\| \leq \|W_k^n\| + \sum_{i=0}^{k-1} \frac{K_b}{n} \left\|\bar{Y}_i^n - \check{Y}_i^n\right\|.$$

Using Lemma 3.7 gives

$$\begin{aligned}\left\|\bar{Y}_k^n - \check{Y}_k^n\right\| &\leq \|W_k^n\| + \sum_{i=0}^{k-1} \|W_i^n\| \frac{K_b}{n} \exp\left\{\frac{K_b}{n}(k-i-1)\right\} \\ &\leq \left(1 + K_b e^{K_b}\right) \max_{i\in\{1,\ldots,k\}} \{\|W_i^n\|\}\end{aligned}$$

so

$$\max_{i\in\{1,\ldots,n\}} \left\{\left\|\bar{Y}_i^n - \check{Y}_i^n\right\|\right\} \leq \left(1 + K_b e^{K_b}\right) \max_{i\in\{1,\ldots,n\}} \{\|W_i^n\|\}.$$

Since $\max_{i\in\{1,\ldots,n\}}\{\|W_i^n\|\} \to 0$ in probability

$$\max_{i\in\{1,\ldots,n\}} \left\{\left\|\bar{Y}_i^n - \check{Y}_i^n\right\|\right\} \to 0 \text{ and hence } \sup_{t\in[0,1]} \left\|\bar{Y}^n(t) - \check{Y}^n(t)\right\| \to 0$$

in probability. □

To complete the proof of the convergence we need to show $\check{Y}^n - \hat{Y}^n \to 0$. Recall that these two processes have the same driving terms but different drifts, in that $\hat{Y}^n$ satisfies the ODE

$$\hat{Y}^n(t) = \int_0^t Db(X^0(s))\hat{Y}^n(s)ds + \int_0^t \hat{w}^n(s)ds$$

while $\check{Y}^n$ is the linear interpolation of the discrete time process defined by $\check{Y}_0^n = 0$ and

$$\check{Y}_{i+1}^n = \check{Y}_i^n + \frac{a(n)}{\sqrt{n}}\left(b\left(X_i^{n,0} + \frac{1}{a(n)\sqrt{n}}\check{Y}_i^n\right) - b\left(X_i^{n,0}\right)\right) + \frac{1}{n}\hat{w}^n\left(\frac{i}{n}\right).$$

However, essentially the same arguments as those used in Lemma 3.4 to show tightness of $\{\hat{Y}^n\}$ can be used to prove tightness of $\{\check{Y}^n\}$, and then it easily follows as in Lemma 3.5 that any limit will satisfy the same ODE (2.15) as the limit of $\{\hat{Y}^n\}$, and therefore $\check{Y}^n - \hat{Y}^n \to 0$ follows.

Combining $\bar{Y}^n - \check{Y}^n \to 0$, $\check{Y}^n - \hat{Y}^n \to 0$, and $\hat{Y}^n \to \hat{Y}$ demonstrates that along the subsequence where $\hat{\eta}^n \to \hat{\eta}$ weakly $\bar{Y}^n \to \hat{Y}$ in distribution, which implies that along this subsequence $(\hat{\eta}^n, \bar{Y}^n) \to (\hat{\eta}, \hat{Y})$ weakly. We have already shown that with probability 1 $\hat{\eta}_2(dt)$ is Lebesgue measure and

$$\hat{Y}(t) = \int_0^t Db(X^0(s))\hat{Y}(s)ds + \int_0^t \int_{\mathbb{R}^d} y\hat{\eta}_{1|2}(dy\,|t)ds,$$

so the proof of convergence (i.e., the first part of Theorem 2.5) is complete.

To finish Theorem 2.5 we must lastly show the bound (2.16). Note that the weak convergence of $\bar{Y}^n$ implies

$$\sup_{t\in[0,1]} \left\| \bar{X}^n(\lfloor nt \rfloor /n) - X^0(t) \right\| \to 0 \text{ in probability.} \tag{3.10}$$

Define random measures on $\mathbb{R}^d \times \mathbb{R}^d \times [0,1]$ by

$$\gamma^n (dx \times dy \times dt) = \delta_{\bar{X}^n(\lfloor nt \rfloor/n)} (dx)\, \hat{\eta}^n (dy \times dt)\,.$$

Note that the tightness of $\{\gamma^n\}$ follows easily from (3.10) and from the tightness of $\{\hat{\eta}^n\}$. Thus given any subsequence we can choose a further subsequence (again we will retain n as the index for simplicity) along which $\{\gamma^n\}$ converges weakly to some limit γ on $\mathcal{P}\left(\mathbb{R}^d \times \mathbb{R}^d \times [0,1]\right)$ with

$$\gamma_{2,3} (dy \times dt) = \hat{\eta} (dy \times dt)\,,$$

where $\gamma_{2,3}$ is the second and third marginal of γ. If we establish (2.16) for this subsequence it follows for the full sequence using a standard argument by contradiction. For $\sigma > 0$ let

$$G_\sigma^{X^0} = \left\{ (x,y,t) : \left\| x - X^0 (t) \right\| \leq \sigma \right\}$$

be closed sets centered around $X^0 (t)$ in the x variable, and note that by (3.10) and weak convergence, for all $\sigma > 0$

$$1 = \limsup_{n\to\infty} E\left[\gamma^n \left(G_\sigma^{X^0} \right) \right] \leq E\left[\gamma \left(G_\sigma^{X^0} \right) \right].$$

Thus

$$E\left[\gamma \left(\cap_{n\in\mathbb{N}} G_{1/n}^{X^0} \right) \right] = 1,$$

so with probability 1, γ puts all its mass on $\left\{ (x,y,t) : x = X^0 (t) \right\}$. Therefore with probability 1, for a.e. (y,t) under $\gamma_{2,3} (dy \times dt)$,

$$\gamma_{1|2,3} \left(dx | \, y, t \right) = \delta_{X^0(t)} (dx)\,.$$

Combined with the fact that the second marginal of $\hat{\eta} (dy \times dt)$ is Lebesgue measure, this gives

$$\gamma (dx \times dy \times dt) = \delta_{X^0(t)} (dx)\, \hat{\eta} \left(dy | \, t \right) dt. \tag{3.11}$$

We will use the following Lemma to prove a lower bound on the relative entropy cost.

LEMMA 3.9. *Define*

$$\bar{L}_K(x,\beta) \doteq \sup_{\alpha\in\mathbb{R}^d} \left\{ \langle \alpha,\beta\rangle - \frac{1}{2}\|\alpha\|^2_{A(x)} - \frac{1}{2K}\|\alpha\|^2 \right\}. \tag{3.12}$$

Then uniformly in x and compact sets of β

$$\liminf_{n\to\infty} a(n)^2 n L_c(x,\beta/a(n)\sqrt{n}) \geq \bar{L}_K(x,\beta).$$

PROOF. Given a compact set B there is $C<\infty$ such that $\|\beta\|\leq C$ for $\beta\in B$. Given $K<\infty$, from the definition (3.12) there is $M_C<\infty$ (which also depends on K) such that for all x and all $\|\beta\|\leq C$,

$$\bar{L}_K(x,\beta) = \sup_{\|\alpha\|\leq M_C} \left\{ \langle \alpha,\beta\rangle - \frac{1}{2}\|\alpha\|^2_{A(x)} - \frac{1}{2K}\|\alpha\|^2 \right\}.$$

Define

$$\hat{L}_n(x,\beta) \doteq a(n)^2 n L_c(x,\beta/a(n)\sqrt{n})$$

and

$$\hat{H}_n(x,\alpha) \doteq a(n)^2 n H_c(x,\alpha/a(n)\sqrt{n}),$$

and note that $\hat{L}_n(x,\beta)$ is dual to $\hat{H}_n(x,\alpha)$ in the sense that

$$\hat{L}_n(x,\beta) = \sup_{\alpha\in\mathbb{R}^d}\{\langle\alpha,\beta\rangle - \hat{H}_n(x,\alpha)\}.$$

Then

$$\hat{L}_n(x,\beta) \geq \sup_{\|\alpha\|\leq M_C}\{\langle\alpha,\beta\rangle - \hat{H}_n(x,\alpha)\},$$

and due to (2.4) there exists some N_{M_C} such that for all x and all $n\geq N_{M_C}$

$$\begin{aligned}
&\sup_{\|\alpha\|\leq M_C}\{\langle\alpha,\beta\rangle - \hat{H}_n(x,\alpha)\}\\
&\quad\geq \sup_{\|\alpha\|\leq M_C} \left\{ \langle \alpha,\beta\rangle - \frac{1}{2}\|\alpha\|^2_{A(x)} - \frac{1}{2K}\|\alpha\|^2 \right\}\\
&\quad= \bar{L}_K(x,\beta).
\end{aligned}$$

Consequently for all x and all $\beta\in B$

$$a(n)^2 n L_c(x,\beta/a(n)\sqrt{n}) = \hat{L}_n(x,\beta) \geq \bar{L}_K(x,\beta)$$

for all $n\geq N_{M_C}$. $\square$

Let $\bar{L}_K(x,\beta)$ be given by (3.12). Note that

$$\bar{L}_K(x,\beta) \uparrow \frac{1}{2}\|\beta\|^2_{A^{-1}(x)}$$

as $K \to \infty$ for all $(x,\beta) \in \mathbb{R}^{2d}$. Combining Lemma 3.9 with Lemma 3.1 and using Fatou's lemma for weak convergence,

$$\begin{aligned}
&\liminf_{n\to\infty} a(n)^2 nE\left[\sum_{i=0}^{n-1} \frac{1}{n} R(\eta_i^n \| \mu_{\bar{X}_i^n})\right] \\
&\quad \geq \liminf_{n\to\infty} E\left[\int_{\mathbb{R}^d\times\mathbb{R}^d\times[0,1]} a(n)^2 nL_c\left(x, \frac{1}{a(n)\sqrt{n}}y\right)\gamma^n(dx\times dy\times dt)\right] \\
&\quad \geq E\left[\int_{\mathbb{R}^d\times\mathbb{R}^d\times[0,1]} \bar{L}_K(x,y)\,\gamma(dx\times dy\times dt)\right]
\end{aligned}$$

for all K. Then using the monotone convergence theorem, the decomposition (3.11), and Jensen's inequality in that order shows that

$$\begin{aligned}
&\liminf_{n\to\infty} a(n)^2 nE\left[\sum_{i=0}^{n-1} \frac{1}{n} R(\eta_i^n \| \mu_{\bar{X}_i^n})\right] \\
&\quad \geq \lim_{K\to\infty} E\left[\int_{\mathbb{R}^d\times\mathbb{R}^d\times[0,1]} \bar{L}_K(x,y)\,\gamma(dx\times dy\times dt)\right] \\
&\quad = E\left[\int_{\mathbb{R}^d\times\mathbb{R}^d\times[0,1]} \frac{1}{2}\|y\|^2_{A^{-1}(x)}\,\gamma(dx\times dy\times dt)\right] \\
&\quad = E\left[\int_0^1 \int_{\mathbb{R}^d} \frac{1}{2}\|y\|^2_{A^{-1}(X^0(t))}\,\hat{\eta}(dy|\,t)\,dt\right] \\
&\quad \geq E\left[\frac{1}{2}\int_0^1 \|\hat{w}(t)\|^2_{A^{-1}(X^0(t))}\,dt\right],
\end{aligned}$$

which is (2.16).

4. Laplace Upper Bound. The goal of this section is to prove (2.12), which due to the minus sign corresponds to the Laplace upper bound. Suppose for each n that η^n comes within ε of achieving the infimum in (2.9), so that

$$\begin{aligned}
&\liminf_{n\to\infty} -a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] + \varepsilon \\
&\quad \geq \liminf_{n\to\infty} E\left[\sum_{i=0}^{n-1} a(n)^2 R(\eta_i^n \| \mu_{\bar{X}_i^n}) + F(\bar{Y}^n)\right]. \qquad (4.1)
\end{aligned}$$

Since $\sup_{x\in\mathbb{R}^d}|F(x)| \le K_F$ for some $K_F < \infty$, we also have

$$\sup_n a(n)^2 nE\left[\sum_{i=0}^{n-1} \frac{1}{n} R(\eta_i^n \| \mu_{\bar{X}_i^n})\right] \le 2K_F + \varepsilon.$$

Consequently we can choose a subsequence of $\{\eta^n\}$ (we retain n as the index for convenience) along which the conclusions of Theorem 2.5 hold. Combining this with (4.1) gives

$$\begin{aligned}
&\liminf_{n\to\infty} -a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] + \varepsilon \\
&\quad \ge \liminf_{n\to\infty} E\left[\sum_{i=0}^{n-1} a(n)^2 R(\eta_i^n \| \mu_{\bar{X}_i^n}) + F(\bar{Y}^n)\right] \\
&\quad \ge E\left[\int_0^1 \frac{1}{2}\|\hat{w}(s)\|^2_{A^{-1}(X^0(s))}\, ds + F(\hat{Y})\right].
\end{aligned}$$

Recalling

$$\hat{Y}(t) = \int_0^t Db(X^0(s))\hat{Y}(s)ds + \int_0^t \hat{w}(s)ds,$$

it follows that

$$\begin{aligned}
&E\left[\int_0^1 \frac{1}{2}\|\hat{w}(s)\|^2_{A^{-1}(X^0(s))}\, ds + F(\hat{Y})\right] \\
&\ge \inf_{u\in L^2([0,1]:\mathbb{R}^d)} \left\{\int_0^1 \frac{1}{2}\|u(s)\|^2_{A^{-1}(X^0(s))}\, ds + F(\phi^u)\right\} \\
&= \inf_{u\in L^2([0,1]:\mathbb{R}^d)} \left\{\int_0^1 \frac{1}{2}\|u(s)\|^2\, ds + F\left(\phi^{A^{1/2}(X^0)u}\right)\right\},
\end{aligned}$$

with ϕ^u defined as in (2.8). Since $\varepsilon > 0$ is arbitrary, we have the lower bound (2.12).

5. Laplace Lower Bound. The goal of this section is to prove (2.13). Note that for $u, v \in L^2([0,1]:\mathbb{R}^d)$ and $\phi^{A^{1/2}(X^0)u}, \phi^{A^{1/2}(X^0)v}$ given by (2.8)

$$\begin{aligned}
&\phi^{A^{1/2}(X^0)u}(t) - \phi^{A^{1/2}(X^0)v}(t) \\
&\quad = \int_0^t Db(X^0(s))\left(\phi^{A^{1/2}(X^0)u}(s) - \phi^{A^{1/2}(X^0)v}(s)\right) ds \\
&\qquad + \int_0^t A^{1/2}(X^0(s))(u(s) - v(s))ds.
\end{aligned}$$

Thus by Gronwall's inequality

$$\begin{aligned}(5.1)\qquad &\sup_{t\in[0,1]}\left\|\phi^{A^{1/2}(X^0)u}(t)-\phi^{A^{1/2}(X^0)v}(t)\right\|\\ &\le (1+K_be^{K_b})\int_0^1\left\|A^{1/2}(X^0(s))u(s)-A^{1/2}(X^0(s))v(s)\right\|ds\\ &\le (1+K_be^{K_b})K_A^{1/2}\left(\int_0^1\|u(s)-v(s)\|^2\,ds\right)^{\frac{1}{2}}.\end{aligned}$$

Since $C([0,1]:\mathbb{R}^d)$ is dense in $L^2([0,1]:\mathbb{R}^d)$, the proof of the Laplace lower bound is reduced to showing that for an arbitrary $u\in C([0,1]:\mathbb{R}^d)$
(5.2)

$$\limsup_{n\to\infty} -a(n)^2\log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right]\le \frac{1}{2}\int_0^1\|u(s)\|^2\,ds+F\left(\phi^{A^{1/2}(X^0)u}\right).$$

The main difficulty is to deal with the possible degeneracy of the noise. Recall the orthogonal decomposition of $A^{-1}(x)$ (2.2). Define

$$A_K^{-1}(x)=Q(x)\Lambda_K^{-1}(x)Q^T(x)$$

where $\Lambda_K^{-1}(x)$ is the diagonal matrix such that $\Lambda_{ii,K}^{-1}(x)=\Lambda_{ii}^{-1}(x)$ when $\Lambda_{ii}^{-1}(x)\le K^2$ and $\Lambda_{ii,K}^{-1}(x)=K^2$ when $\Lambda_{ii}^{-1}(x)>K^2$. Note that by [19, Theorem 6.2.37] $A^{1/2}(x)$, $A_K^{-1}(x)$ and $A_K^{1/2}(x)$ are continuous functions of $A(x)$, and consequently they are also continuous functions of $x\in\mathbb{R}^d$. In addition define

$$u_K(s)=\begin{cases} u(s) & \text{for } \|u(s)\|\le K\\ \frac{Ku(s)}{\|u(s)\|} & \text{for } \|u(s)\|>K\end{cases}.$$

Let $\phi^{u,K}(t)=\phi^{A(X^0)A_K^{-1/2}(X^0)u_K}(t)$, and note that $\phi^{u,K}$ solves

$$\begin{aligned}\phi^{u,K}(t)&=\int_0^t Db(X^0(s))\phi^{u,K}(s)ds\\ (5.3)\qquad &\quad+\int_0^t A(X^0(s))A_K^{-1/2}(X^0(s))u_K(s)ds.\end{aligned}$$

To simplify notation we define $s_i^n\doteq i/n$ and $s^n(t)=\lfloor nt\rfloor/n$, where $\lfloor a\rfloor$ is the integer part of a. Note that $s^n(t)-t\to 0$ uniformly for $t\in[0,1]$ as $n\to\infty$. For n sufficiently large

$$\max_{0\le i\le n-1}\left\{\frac{1}{a(n)\sqrt{n}}\left\|A_K^{-1/2}\left(X^0\left(s_i^n\right)\right)u_K\left(s_i^n\right)\right\|\right\}\le\frac{1}{a(n)\sqrt{n}}K^2\le\lambda_{DA}$$

and we can define the sequence $\{(\bar{X}^{n,u,K}, \bar{Y}^{n,u,K}, \eta^{n,u,K}, \hat{\eta}^{n,u,K})\}$ as in Construction 2.4 with

$$\begin{aligned}
&\eta_i^{n,u,K}(dy) \\
&\quad = \exp\left\{\left\langle y, \frac{1}{a(n)\sqrt{n}} A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right\rangle\right. \\
&\qquad\qquad \left. - H_c\left(\bar{X}_i^{n,u,K}, \frac{1}{a(n)\sqrt{n}} A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right)\right\} \mu_{\bar{X}_i^{n,u,K}}(dy).
\end{aligned}$$

Using (2.3) and the fact that

$$\int_{\mathbb{R}^d} y \exp\{\langle y, \alpha\rangle - H_c(x,\alpha)\}\mu_x(dy) = D_\alpha H_c(x,\alpha),$$

we have for $\|\alpha\| \leq \lambda_{DA}$

$$\left\|\int_{\mathbb{R}^d} y \exp\{\langle y, \alpha\rangle - H_c(x,\alpha)\}\mu_x(dy) - A(x)\alpha\right\| \leq K_{DA}\left\|\alpha\right\|^2. \tag{5.4}$$

The next result identifies the limit in probability of the controlled processes and an asymptotic bound for the relative entropies.

THEOREM 5.1. *Let $u \in C([0,1] : \mathbb{R}^d)$ and $K < \infty$ be given, construct $\{(\bar{X}^{n,u,K}, \bar{Y}^{n,u,K}, \eta^{n,u,K}, \hat{\eta}^{n,u,K})\}$ as in this section and define $\phi^{u,K}$ by (5.3). Then*

$$\bar{Y}^{n,u,K} \to \phi^{u,K} \tag{5.5}$$

in $C([0,1] : \mathbb{R}^d)$ in probability, and

$$\begin{aligned}
&\limsup_{n\to\infty} a^2(n) n E\left[\frac{1}{n}\sum_{i=0}^{n-1} R\left(\eta_i^{n,u,K} \middle\| \mu_{\bar{X}_i^{n,u,K}}\right)\right] \\
&\qquad \leq \frac{1}{2}\int_0^1 \left\|A_K^{-1/2}(X^0(s)) u_K(s)\right\|^2_{A(X^0(s))} ds.
\end{aligned} \tag{5.6}$$

PROOF. Using (5.4) to bound the second term and (2.4) to bound the third, for n satisfying $\frac{1}{a(n)\sqrt{n}}K^2 \leq \lambda_{DA}$

$$\begin{aligned}
&R\left(\eta_i^{n,u,K} \middle\| \mu_{\bar{X}_i^{n,u,K}}\right) \\
&\quad = \int_{\mathbb{R}^d} \left\langle y, \frac{1}{a(n)\sqrt{n}} A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right\rangle \eta_i^{n,u,K}(dy)
\end{aligned}$$

$$
\begin{aligned}
&- H_c\left(\bar{X}_i^{n,u,K}, \frac{1}{a(n)\sqrt{n}} A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right) \\
&\leq \left\langle \frac{1}{a(n)\sqrt{n}} A\left(\bar{X}_i^{n,u,K}\right) A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right),\right. \\
&\qquad\qquad \left.\frac{1}{a(n)\sqrt{n}} A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right\rangle \\
&\quad - \frac{1}{2}\left\langle \frac{1}{a(n)\sqrt{n}} A\left(\bar{X}_i^{n,u,K}\right) A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right),\right. \\
&\qquad\qquad \left.\frac{1}{a(n)\sqrt{n}} A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right\rangle + \frac{2}{a(n)^3 n^{3/2}} K_{DA} K^6 \\
&= \frac{1}{2a(n)^2 n}\left\|A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right\|_{A(\bar{X}_i^{n,u,K})}^2 + \frac{2}{a(n)^3 n^{3/2}} K_{DA} K^6.
\end{aligned}
$$

Consequently

$$
\begin{aligned}
\text{(5.7)} \qquad &\limsup_{n\to\infty} a^2(n) n E\left[\frac{1}{n}\sum_{i=0}^{n-1} R\left(\eta_i^{n,u,K}\middle\| \mu_{\bar{X}_i^{n,u,K}}\right)\right] \\
&\qquad \leq \limsup_{n\to\infty} \frac{1}{2} E\left[\frac{1}{n}\sum_{i=0}^{n-1}\left\|A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right\|_{A(\bar{X}_i^{n,u,K})}^2\right],
\end{aligned}
$$

where in fact

$$
\limsup_{n\to\infty} \frac{1}{2} E\left[\frac{1}{n}\sum_{i=0}^{n-1}\left\|A_K^{-1/2}\left(X^0\left(s_i^n\right)\right) u_K\left(s_i^n\right)\right\|_{A(\bar{X}_i^{n,u,K})}^2\right] \leq \frac{1}{2} K^4 K_A.
$$

Therefore (2.14) is satisfied by $\{\eta^{n,u,K}\}$, so we can apply Theorem 2.5 and choose a subsequence (keeping n as the index for convenience) along which $\{(\hat{\eta}^{n,u,K}, \bar{Y}^{n,u,K})\}$ converges weakly to some limit $(\hat{\eta}^{u,K}, \hat{Y}^{u,K})$, where $\hat{\eta}_2^{u,K}$ is Lebesgue measure and

$$
\hat{Y}^{u,K}(t) = \int_0^t Db(X^0(s))\hat{Y}^{u,K}(s)ds + \int_0^t \int_{\mathbb{R}^d} y \hat{\eta}_{1|2}^{u,K}(dy\,|\,s)ds.
$$

This implies

$$
\text{(5.8)} \qquad \sup_{t\in[0,1]}\left\|\bar{X}^{n,u,K}(t) - X^0(t)\right\| \to 0
$$

in probability. Because of this, the uniform bound on $A^{1/2}(x)$ and the continuity of $A^{1/2}(x)$, we have (recall that $s^n(t) \doteq \lfloor nt \rfloor / n$)

$$
\sup_{t\in[0,1]}\left\|A^{1/2}(\bar{X}^{n,u,K}(s^n(t))) - A^{1/2}(X^0(s^n(t)))\right\| \to 0
$$

in probability. However, the continuity of $A^{1/2}(X^0)A_K^{-1/2}(X^0)u_K$ gives

$$\sup_{t\in[0,1]}\Big\|A^{1/2}(X^0(s^n(t)))A_K^{-1/2}(X^0(s^n(t)))u_K(s^n(t)) \\ -A^{1/2}(X^0(t))A_K^{-1/2}(X^0(t))u_K(t)\Big\| \to 0.$$

Combining these limits, and using the fact that $A_K^{-1/2}(X^0)u_K$ is uniformly bounded, shows that

$$\sup_{t\in[0,1]}\Big\|A^{1/2}(\bar{X}^{n,u,K}(s^n(t)))A_K^{-1/2}(X^0(s^n(t)))u_K(s^n(t)) \\ -A^{1/2}(X^0(t))A_K^{-1/2}(X^0(t))u_K(t)\Big\| \to 0 \tag{5.9}$$

in probability. This combined with the uniform bound on $A_K^{-1/2}(X^0)u_K$ and dominated convergence gives

$$\limsup_{n\to\infty} E\left[\frac{1}{2}\int_0^1 \Big\|A_K^{-1/2}(X^0(s^n(t)))u_K(s^n(t))\Big\|^2_{A(\bar{X}^{n,u,K}(s^n(t)))}\,dt\right] \\ = \frac{1}{2}\int_0^1 \Big\|A_K^{-1/2}(X^0(t))u_K(t)\Big\|^2_{A(X^0(t))}\,dt.$$

Combining this with (5.7) shows (5.6).

To prove (5.5) we will show that in fact

$$\hat{\eta}^{u,K}(dy\times dt) = \delta_{A(X^0(t))A_K^{-1/2}(X^0(t))u_K(t)}(dy)dt.$$

For all $\sigma > 0$ let

$$G_\sigma = \left\{(z,t)\in\mathbb{R}^d\times[0,1] : \Big\|z - A(X^0(t))A_K^{-1/2}(X^0(t))u_K(t)\Big\| \le \sigma\right\},$$

and note that by weak convergence $\limsup_{n\to\infty} E[\hat{\eta}^{n,u,K}(G_\sigma)] \le E[\hat{\eta}^{u,K}(G_\sigma)]$. Note also that

$$E[\hat{\eta}^{n,u,K}(G_\sigma)] \\ \ge P\left[\sup_{t\in[0,1]}\Big\|a(n)\sqrt{n}\int_{\mathbb{R}^d} y\eta^{n,u,K}_{\lfloor nt\rfloor}(dy) - A(X^0(t))A_K^{-\frac{1}{2}}(X^0(t))u_K(t)\Big\| \le \sigma\right].$$

However, by (5.4) we can choose n large enough to make

$$\sup_{t\in[0,1]}\Big\|a(n)\sqrt{n}\int_{\mathbb{R}^d} y\eta^{n,u,K}_{\lfloor nt\rfloor}(dy) \\ - A\big(\bar{X}^{n,u,K}\big(s^n(t)\big)\big)A_K^{-1/2}\big(X^0\big(s^n(t)\big)\big)u_K\big(s^n(t)\big)\Big\|$$

arbitrarily small, and the proof that

$$\sup_{t\in[0,1]} \Big\| A(\bar{X}^{n,u,K}(s^n(t)))A_K^{-1/2}(X^0(s^n(t)))u_K(s^n(t)) \\ -A(X^0(t))A_K^{-1/2}(X^0(t))u_K(t)\Big\| \to 0$$

in probability is identical to the proof of (5.9). Therefore for any $\sigma > 0$ $\limsup_{n\to\infty} E[\hat{\eta}^{u,K,n}(G_\sigma)] = 1$, and so $E[\hat{\eta}^{u,K}(\cap_{n\in\mathbb{N}} G_{1/n})] = 1$. This implies that with probability 1

$$\hat{\eta}^{u,K}_{1|2}(dy|\,t) = \delta_{A(X^0(t))A_K^{-1/2}(X^0(t))u_K}(dy)$$

for a.e. t. It follows that

$$\hat{Y}^{u,K}(t) = \int_0^t Db(X^0(s))\hat{Y}^{u,K}(s)ds + \int_0^t A(X^0(s))A_K^{-1/2}(X^0(s))u_K(s)ds,$$

and therefore $\bar{Y}^{n,u,K} \to \phi^{u,K}$ weakly. This implies (5.5) and completes the proof. □

The second theorem in this section allows us to approximate $F(\phi^{A^{1/2}(X^0)u})$ by $F(\phi^{u,K})$ and $\frac{1}{2}\int_0^1 \|u(s)\|^2 ds$ by

$$\frac{1}{2}\int_0^1 \left\| A_K^{-1/2}(X^0(s))u_K(s)\right\|^2_{A(X^0(s))} ds.$$

THEOREM 5.2. *Let $u \in C([0,1]:\mathbb{R}^d)$ and define $\phi^{A_K^{1/2}(X^0)u}$ by (2.8) and $\phi^{u,K}$ by (5.3). Then as $K \to \infty$*

$$\phi^{u,K} \to \phi^{A^{1/2}(X^0)u}$$

in $C([0,1]:\mathbb{R}^d)$ and

$$\sup_{K\in(0,\infty)} \frac{1}{2}\int_0^1 \left\| A_K^{-1/2}(X^0(s))u_K(s)\right\|^2_{A(X^0(s))} ds \le \frac{1}{2}\int_0^1 \|u(s)\|^2 ds.$$

PROOF. Note that

$$\left\| A^{1/2}(X^0(s))A_K^{-1/2}(X^0(s))u_K(s)\right\| \le \|u(s)\|$$

for all $s \in [0,1]$ and $K \in (0,\infty)$ so

$$\sup_{K\in(0,\infty)} \frac{1}{2}\int_0^1 \left\| A_K^{-1/2}(X^0(s))u_K(s)\right\|^2_{A(X^0(s))} ds \le \frac{1}{2}\int_0^1 \|u(s)\|^2 ds.$$

In addition,

$$A^{1/2}(X^0(s))A^{1/2}(X^0(s))A_K^{-1/2}(X^0(s))u_K(s) \to A^{1/2}(X^0(s))u(s)$$

and

$$\left\| A^{1/2}(X^0(s))A^{1/2}(X^0(s))A_K^{-1/2}(X^0(s))u_K(s) \right\| \leq \left\| A^{1/2}(X^0(s))u(s) \right\|$$

for all $s \in [0,1]$, so dominated convergence gives

$$A^{1/2}(X^0)A^{1/2}(X^0)A_K^{-1/2}(X^0)u_K \to A^{1/2}(X^0)u$$

in $L^1([0,1] : \mathbb{R}^d)$. Combining this with the second line of (5.1) shows that

$$\phi^{u,K} \to \phi^{A^{1/2}(X^0)u}$$

in $C([0,1] : \mathbb{R}^d)$. $\square$

Using (2.9) and the fact that any given control is suboptimal,

$$\begin{aligned} &- a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] \\ &\quad \leq E\left[\sum_{i=0}^{n-1} a(n)^2 R\left(\eta_i^{n,u,K} \middle\| \mu_{\bar{X}_i^{n,u,K}}\right) + F(\bar{Y}^{n,u,K})\right]. \end{aligned}$$

Using Theorem 5.1, this implies

$$\begin{aligned} &\limsup_{n\to\infty} -a(n)^2 \log E\left[e^{-\frac{1}{a(n)^2}F(Y^n)}\right] \\ &\quad \leq \frac{1}{2}\int_0^1 \left\| A_K^{-1/2}(X^0(s))u_K(s) \right\|_{A(X^0(s))}^2 ds + F(\phi^{u,K}). \end{aligned}$$

Sending $K \to \infty$ and using Theorem 5.2 gives (5.2), and hence completes the proof of the lower bound (2.13).

REFERENCES

[1] Azencott, R. and Ruget, G. (1977). Mélanges d'équations differentielles et grand écart à la loi des grandes nombres. *Z. Wahrsch. Verw. Gebiete* **38** 1–54. MR0443097

[2] Baldi, P. (1988). Large deviations and stochastic homogenization. *Ann. Mat. Pura Appl.* **151** 161–177. MR0964508

[3] Blanchet, J.H., Glynn, P. and Leder, K. (2012). On Lyapunov inequalities and subsolutions for efficient importance sampling. *TOMACS* **22** Article No. 13. MR2980826

[4] Boué, M. and Dupuis, P. (1998). A variational representation for certain functionals of Brownian motion. *Ann. Probab.* **26** 1641–1659. MR1675051
[5] Budhiraja, A., Dupuis, P. and Maroulas, V. (2011). Variational representations for continuous time processes. *Ann. de l'Inst. H. Poincaré* **47** 725–747. MR2841073
[6] Chen, X. and de Acosta, A. (1998). Moderate deviations for empirical measures of Markov chains: Upper bounds. *J. Theor. Probab.* **11** 1075–1110. MR1660920
[7] de Acosta, A. (1997). Moderate deviations for empirical measures of Markov chains: Lower bounds. *Ann. Probab.* **25** 259–284. MR1428509
[8] Dean, T. and Dupuis, P. (2009). Splitting for rare event simulation: A large deviations approach to design and analysis. *Stoch. Proc. Appl.* **119** 562–587. MR2494004
[9] Dembo, A. (1996) Moderate deviations for martingales with bounded jumps. *Elect. Comm. Probab.* **1** 11–17. MR1386290
[10] Dembo, A. and Zeitouni, O. (1993). *Large Deviations Techniques and Applications*. Jones and Bartlett, Boston. MR1202429
[11] Djellout, H. (2002). Moderate deviations for martingale differences and applications to ϕ-mixing sequences. *Stoch. Stoch. Rep.* **73** 37–63. MR1914978
[12] Dupuis, P. and Ellis, R. S. (1997). *A Weak Convergence Approach to the Theory of Large Deviations*. John Wiley & Sons, New York. MR1431744
[13] Dupuis, P. and Wang, H. (2004). Importance sampling, large deviations, and differential games. *Stoch. Stoch. Reports.* **76** 481–508. MR2100018
[14] Dupuis, P. and Wang, H. (2007). Subsolutions of an Isaacs equation and efficient schemes for importance sampling. *Math. Oper. Res.* **32** 1–35. MR2348245
[15] Freidlin, M. I. and Wentzell, A. D. (1984). *Random Perturbations of Dynamical Systems*. Springer-Verlag, New York. MR0722136
[16] Gao, F.Q. (1996). Moderate deviations for martingales and mixing random processes. *Stoch. Proc. Appl.* **61** 263–275. MR1386176
[17] Guillin, A. (2001). Moderate deviations of inhomogeneous functionals of Markov processes and application to averaging. *Stoch. Proc. Appl.* **92** 287–313. MR1817590
[18] Guillin, A. (2003). Averaging principle of SDE with small diffusion: Moderate deviations. *Ann. Probab.* **31** 413–443. MR1959798
[19] Horn, R. A. and Johnson, C. R. (1991). *Topics in Matrix Analysis*. Cambridge University Press, New York, New York. MR1091716
[20] Wentzell, A.D. (1976). Rough limit theorems on large deviations for Markov stochastic processes, I. *Theory Probab. Appl.* **21** 227–242. MR0433565
[21] Wentzell, A.D. (1976). Rough limit theorems on large deviations for Markov stochastic processes, II. *Theory Probab. Appl.* **21** 499–512. MR0433566
[22] Wentzell, A.D. (1979). Rough limit theorems on large deviations for Markov stochastic processes, III. *Theory Probab. Appl.* **24** 675–692.
[23] Wentzell, A.D. (1982). Rough limit theorems on large deviations for Markov stochastic processes, IV. *Theory Probab. Appl.* **27** 215–234. MR0657916