Large deviation for lasso diffusion process

Azzouz Dermoune ¹, Khalifa Es-Sebaiy ², Youssef Ouknine ²

¹ Laboratoire Paul Painlevé, USTL-UMR-CNRS 8524, Lille, France
Email : azzouz.dermoune@univ-lille1.fr,
² Cadi Ayyad University Av. Abdelkrim Khattabi, 40000, Gué Iz-Marrakech, Morocco
Emails : k.essebaiy@uca.ma, ouknine@uca.ma

13 juin 2018

Abstract : The aim of the present paper is to extend the large deviation with discontinuous statistics studied in [5] to the diffusion 
\[ dx^\varepsilon = -\{A^T(Ax^\varepsilon - y) + \mu \text{sgn}(x^\varepsilon)\}dt + \varepsilon dw. \]
The discontinuity of the drift of the diffusion discussed in [5] is equal to the hyperplane \( \{x \in \mathbb{R}^d : x_1 = 0\} \), however, in our case the discontinuity is more complex and is equal to the set \( \{x \in \mathbb{R}^d : \prod_{i=1}^d x_i = 0\} \).

1 Introduction

Let \( y \in \mathbb{R}^n \) be a given vector, \( A \) be a known matrix which maps the domain \( \mathbb{R}^d \) into the domain \( \mathbb{R}^n \) and \( \mu > 0 \) is a given positive real number. The sign of the real number \( u \) equals 
\[ \text{sgn}(u) = 1 \text{ if } u > 0, \text{ sgn}(u) = -1 \text{ if } u < 0 \text{ and sgn}(0) \text{ is any element of } [-1,1]. \]
The column vector \( \text{sgn}(x) := (\text{sgn}(x_1), \ldots, \text{sgn}(x_p))^\top \).

The following diffusion
\[ dx^\varepsilon = -\{A^T(Ax^\varepsilon - y) + \mu \text{sgn}(x^\varepsilon)\}dt + \varepsilon dw, \quad x^\varepsilon(0) = x(0), \]
has a discontinuous drift. Using the fact that \( A^T(Ax - y) + \mu \text{sgn}(x) \) is the subdifferential of the convex map
\[ \frac{\|Ax - y\|^2}{2} + \mu \|x\|_1 \]
we can show that the latter stochastic differential equation (sde) has a unique strong solution for any \( \varepsilon > 0 \). See ([15, 7, 8, 17]). Here \( \| \cdot \|, \| \cdot \|_1 \) denote respectively the \( l2 \) and \( l1 \) norms.

The asymptotic property as \( t \to +\infty \) is also possible. The probability density function
\[ \frac{1}{Z} \exp\left\{-\frac{2}{\varepsilon}\frac{\|Ax - y\|^2}{2} + \mu \|x\|_1\right\} := p^\varepsilon(dx) \]
is the unique invariant probability measure of \((\mathbb{N}_t^\varepsilon)\), see e.g. [1]. The mode of \(p^\varepsilon\) was introduced in linear regression by [18] and is called lasso. Lasso is the compact and convex set solution of the system

\[
A_i^\top (Ax - y) + \mu \text{sgn}(x_i) = 0, \quad i = 1, \ldots, d. \tag{2}
\]

Here \(A_i^\top\) denotes the \(i\)-th row of the matrix \(A^\top\). A large number of theoretical results has been provided for lasso see e.g. [10, 12, 18, 19, 16] and the references herein.

If \((P_t^\varepsilon)\) is the semi-groupe defined by (1) then we have the following exponential convergence

\[
\mathbb{E}_{p^\varepsilon}[|P_t^\varepsilon f - \mathbb{E}_{p^\varepsilon}(f)|^2] \leq \exp(-t/C)\text{var}_{p^\varepsilon}(f), \tag{3}
\]

where

\[
C = 4\mathbb{E}_{p^\varepsilon}[\|x - \mathbb{E}_{p^\varepsilon}(x)\|^2].
\]

The proof is a consequence of Poincaré inequality ([14, 3]):

\[
\text{var}_{p^\varepsilon}(f) \leq 4\mathbb{E}_{p^\varepsilon}[\|x - \mathbb{E}_{p^\varepsilon}(x)\|^2]\mathbb{E}_{p^\varepsilon}(\|\nabla f\|^2)
\]

valid for all lipschitz map \(f\), because \(p^\varepsilon\) is log-concave, and the fact that Poincaré inequality is equivalent to the exponential convergence (3). As a consequence we can suppose that a.s.

\[
\sup_{t \geq 0} \|x^\varepsilon(t)\| < +\infty.
\]

2 Limit as \(\varepsilon \to 0\)

Let \(U : \mathbb{R}^d \to \mathbb{R}\) be a convex map such that

\[
|\nabla U(x)| \leq L(1 + |x|), \quad \forall x,
\]

where \(L\) is a positive constant and \(\nabla U\) denotes the sub-differential of \(U\).

It’s known (see e.g. [15, 7]) that the sde

\[
dx^\varepsilon(t) \in -\nabla U(x^\varepsilon(t))dt + \varepsilon dw_t, \quad t \in [0, T]
\]

with fixed initial value \(x(0)\), has a unique strong solution. More precisely there exists a unique solution

\[
x^\varepsilon(t) = x(0) - \int_0^t v^\varepsilon(s)ds + \varepsilon w_t, \quad \forall t \in [0, T]
\]

where the measurable map \(v^\varepsilon\) is such that

\[
v^\varepsilon(t) \in \nabla U(x^\varepsilon_t) \quad \text{dt a.e.}
\]
From the linear growth of $\nabla U$ we have
\[ \|x^\varepsilon(t)\| \leq K + L \int_0^t |x^\varepsilon(s)|ds, \]
where
\[ K = \|x(0)\| + LT + \varepsilon \sup_{t \in [0,T]} \|w_t\|. \]

Gronwall’s lemma tells us that
\[ \|x^\varepsilon(t)\| \leq K \exp(LT), \quad \forall t \in [0,T]. \]

Using Ascoli theorem, we can extract a subsequence such that $x^\varepsilon \to x$ uniformly in $[0,T]$. Now using the inequality
\[ \|v^\varepsilon(t)\| \leq L(1 + \|x^\varepsilon(t)\|) \leq C, \quad \forall t, \]
we derive that the sequence $(v^\varepsilon)$ is weakly precompact in $L^p([0,T])$ for all $1 < p < +\infty$. Using Mazur’s lemma we can construct a measurable map $v$ and a subsequence such that $v^\varepsilon(t) \to v(t)$, $dt$ a.e.

From the condition $v^\varepsilon(t) \in \partial U(x^\varepsilon(t))$, the convergence $(x^\varepsilon(t), v^\varepsilon(t)) \to (x(t), v(t))$ and the fact that $\partial U$ is monotone maximal we have $v(t) \in \partial U(x(t))$ $dt$ a.e." Finally the limit $x$ is the unique solution of the differential inclusion
\[ dx_t \in -\partial U(x_t)dt. \]

See also [6].

As an application the solution $x^\varepsilon$ of (1) converges as $\varepsilon \to 0$ to the inclusion equation
\[ dx^0(t) \in -A^\top (Ax^0(t) - y) + \mu \text{sgn}(x^0), \quad x^0(0) = x(0). \tag{4} \]
The solution $x^0(t)$ converges to lasso as $t \to +\infty$ i.e. $x^0(t)$ converges to the minimizers of
\[ \frac{\|Ax - y\|^2}{2} + \mu \|x\|_1. \]
The set of the latter minimizers is compact. Hence
\[ \sup_{t \geq 0} \|x^0(t)\| < +\infty. \tag{5} \]
Then we have for some $C > 0$ and small $\varepsilon$ that
\[ \sup_{t \geq 0} \|x^\varepsilon(t)\| < C \tag{6} \]
with a big a probability.

The aim of our work is to extend Boué, Dupuis, Ellis large deviation with discontinuous statistics [5] to the diffusion (1). The discontinuity in [5] is equal to the hyperplane $\{x \in \mathbb{R}^d : x_1 = 0\}$. The discontinuity of the drift of the diffusion (1) is more complex and is equal to the set $\{x \in \mathbb{R}^d : \prod_{i=1}^d x_i = 0\}$. Before arriving to large deviation result we need some preliminary results.
3 Preliminary results

We work in the canonical probability space \((\Omega, \mathcal{F}, \mathbb{P})\) where \(\Omega = C([0, 1], \mathbb{R}^d)\) endowed with its Borel \(\sigma\)-field \(\mathcal{F}\), and its Wiener measure \(\mathbb{P}\). The canonical process \(W_t : w \in \Omega \to w(t),\ t \in [0, 1]\) is the Wiener process under \(\mathbb{P}\). The filtration \(\mathcal{F}_t := \sigma(\{W_s : \ s \leq t\}, \mathcal{N}),\ t \in [0, 1]\), where \(\mathcal{N}\) is the collection of the \(\mathbb{P}\)-null sets. The diffusion \(x^\varepsilon(1)\) is considered in the canonical probability space \((\Omega, \mathcal{F}, \mathbb{P})\). Its discontinuous drift is

\[
b(x) := -\{Ax - y + \mu \text{sgn}(x)\}.
\]

We denote by \(\mathbb{E}_{x(0)}\) the mathematical expectation under the probability distribution of the solution \(x^\varepsilon(0) = x(0)\).

I) We define for each \(i = 1, \ldots, d\), the Borel measures

\[
\begin{align*}
\gamma_{i,1}^\varepsilon(dt) &= 1_{\{x_i^\varepsilon(t) \leq 0\}} \, dt, \\
\gamma_{i,2}^\varepsilon(dt) &= 1_{\{x_i^\varepsilon(t) > 0\}} \, dt.
\end{align*}
\]

By extracting a subsequence we have \(x^\varepsilon \to x^0\) where \(x^0\) is the solution of the inclusion differential equation (4), and

\[
(\gamma_{i}^\eta(dt), i = 1, \ldots, d, \eta = 1, 2) \to (\gamma_{i}^\eta(dt), i = 1, \ldots, d, \eta = 1, 2)
\]

where the Borel measures \((\gamma_{i}^\eta(dt), i = 1, \ldots, d, \eta = 1, 2)\) satisfy

\[
\begin{align*}
\gamma_{i}^\eta(dt) &= \tilde{\gamma}_{i}^\eta(dt), \forall i = 1, \ldots, d, \eta = 1, 2, \\
\tilde{\gamma}_{i}^{1}(t) + \tilde{\gamma}_{i}^{2}(t) &= 1, \forall i = 1, \ldots, d, \\
\tilde{\gamma}_{i}^{1}(t) &= 1, \quad \text{if } x_i^{0}(t) < 0, \\
\tilde{\gamma}_{i}^{2}(t) &= 1, \quad \text{if } x_i^{0}(t) > 0, \\
\tilde{\gamma}_{i}^{2}(t) - \tilde{\gamma}_{i}^{1}(t) := sgn(x_i^{0}(t)) &\quad \text{if } x_i^{0}(t) = 0, \\
-\{A_{i}^\top (Ax^0(t) - y) - \mu\} &\geq 0, \quad \text{and } -\{A_{i}^\top (Ax^0(t) - y) + \mu\} \leq 0 \quad \text{if } x_i^{0}(t) = 0.  
\end{align*}
\]

The property (8) tells us that \(x_i^{0}(t)\) stays at zero when the strength

\[
|A_{i}^\top (Ax^0(t) - y)| \leq \mu.
\]

This phenomenon is known by physicist [2], and we can show it mathematically using a similar proof as in [5].

II) Now we fix \(f\) deterministic such that \(\int_0^1 \|f(t)\| \, dt < +\infty\). We consider the sde

\[
dx^\varepsilon(t) = \{f(t) - \mu \text{sgn}(x^\varepsilon(t))\} \, dt + \varepsilon dw(t), \quad x(0) \quad \text{is given},
\]

4
and its limit $x^0$ as $\varepsilon \to 0$ is the solution of the differential inclusion
\[dx^0(t) \in \{f(t) - \mu \text{sgn}(x^0(t))\}dt, \quad x(0) \text{ is given}.\] (10)

We have $dt$ a.e.
\[
\frac{dx^0(t)}{dt} = f(t) - \mu \{\hat{\gamma}^2(t) - \hat{\gamma}^1(t)\}.
\]

1) If $x^0_i(t) < 0$, then $\hat{\gamma}^2_i(t) = 0$, $\hat{\gamma}^1_i(t) = 1$ and
\[
\frac{dx^0_i(t)}{dt} = f_i(t) + \mu.
\]

2) If $x^0_i(t) > 0$, then $\hat{\gamma}^1_i(t) = 0$, $\hat{\gamma}^2_i(t) = 1$ and
\[
\frac{dx^0_i(t)}{dt} = f_i(t) - \mu.
\]

3) We have $dt$ a.e on the set $\{t : x^0_i(t) = 0\}$ that
\[-\mu \leq f_i(t) \leq \mu
\]

and
\[
\frac{dx^0_i(t)}{dt} = f_i(t) - \mu \{\hat{\gamma}^2_i(t) - \hat{\gamma}^1_i(t)\}
= \hat{\gamma}^2_i(t)\{f_i(t) - \mu\} + \hat{\gamma}^1_i(t)\{f_i(t) + \mu\}
= 0.
\]

It follows that
\[
f_i(t) = \mu \{\hat{\gamma}^2_i(t) - \hat{\gamma}^1_i(t)\},
1 = \hat{\gamma}^2_i(t) + \hat{\gamma}^1_i(t).
\]

Hence
\[
\hat{\gamma}^2_i(t) = \frac{f_i(t) + \mu}{2\mu},
\hat{\gamma}^1_i(t) = \frac{\mu - f_i(t)}{2\mu}.
\]

Finally, if $x^0_i(t) = 0$, then $dt$ a.e. $\frac{dx^0_i(t)}{dt} = 0$ and
\[
\hat{\gamma}^2_i(t) = \frac{f_i(t) + \mu}{2\mu},
\hat{\gamma}^1_i(t) = \frac{\mu - f_i(t)}{2\mu}.
\]
Observe that \( \beta^1(t) := f_1(t) + \mu \geq 0 \), \( \beta^2(t) := f_1(t) - \mu \leq 0 \), and \( \frac{dx^0_i(t)}{dt} = \beta^2_i(t) + \beta^1_i(t) \beta^1(t) \).

Observe also that \( x^0_i(t) \neq 0 \) if and only if \( |f_1(t)| > \mu \).

By choosing \( f_i \) piecewise constant we obtain the trajectory \( x \) having the following properties:

There exist \( 0 = \tau_1 < \ldots < \tau_r = 1 \) and the constants \( \beta_i(k), i = 1, \ldots, d, k = 1, \ldots, r \) such that

1) \( \frac{dx_i(t)}{dt} = \beta_i(k), \forall t \in [\tau_k, \tau_{k+1}) \),
2) \( x_i(t) \neq 0 \), \( \forall t \in [\tau_k, \tau_{k+1}) \), or \( x_i(t) = 0 \), \( \forall t \in [\tau_k, \tau_{k+1}) \).

We denote by \( \mathcal{N}_0 \) the set of the maps \( x : [0, 1] \rightarrow \mathbb{R}^d \) which satisfy the latter properties. It's a dense subset of \( C([0, 1]) \).

III) Now we introduce the set

\[ \mathcal{A} = \{ v : \Omega \times [0, 1] \rightarrow \mathbb{R} : \text{is progressively measurable and } \mathbb{E}_{x(0)} \left[ \int_0^1 \|v(t)\|^2 dt \right] < +\infty \} \]

and for \( v \in \mathcal{A} \) we denote by \( x^{\varepsilon, v} \) the solution

\[ dx^{\varepsilon, v}(t) = \{ b(x^{\varepsilon, v}(t)) + v(t) \} dt + \varepsilon dw(t), \quad x^{\varepsilon, v}(0) = x(0), \]

where

\[ b(x) = -\{ A^\top (Ax - y) + \mu \text{sgn}(x) \}. \]

Let \( (v^\varepsilon, \varepsilon \in (0, 1]) \subset \mathcal{A} \) be a family of progressively measurable processes such that

\[ \sup_{\varepsilon \in (0, 1]} \mathbb{E}_{x(0)} \left[ \int_0^1 \|v^\varepsilon(t)\|^2 dt \right] < +\infty. \quad (11) \]

We define for each \( i = 1, \ldots, d \), the Borel measures

\[ \nu^\varepsilon(dv, t) dt = \delta_{\nu^\varepsilon(t)}(dv) dt, \]
\[ \nu^\varepsilon,1(dv, t) = 1_{[\nu^\varepsilon(t) \leq 0]} \delta_{\nu^\varepsilon(t)}(dv), \]
\[ \nu^\varepsilon,2(dv, t) = 1_{[\nu^\varepsilon(t) > 0]} \delta_{\nu^\varepsilon(t)}(dv), \]
\[ \gamma^\varepsilon,1(dt) = 1_{[\nu^\varepsilon(t) \leq 0]} dt, \]
\[ \gamma^\varepsilon,1(t) = 1_{[\nu^\varepsilon(t) \leq 0]}, \]
\[ \gamma^\varepsilon,2(dt) = 1_{[\nu^\varepsilon(t) > 0]} dt, \]
\[ \gamma^\varepsilon,2(t) = 1_{[\nu^\varepsilon(t) > 0]}. \]

By extracting a subsequence we have \( x^{\varepsilon, v^\varepsilon} \rightarrow x^0 \), and \( \nu^\varepsilon \rightarrow \nu \). Thanks to (11), we have (see [5])

\[ \mathbb{E}_{x(0)} \left[ \int_{\mathbb{R}^d} \|v\| \mu(dv, t) \right] < +\infty. \]
The limit $x^0$ is the solution of the differential inclusion
\[ dx^0(t) \in \{f(t) - \mu \text{sgn}(x^0(t))\} dt, \]
where
\[ f(t) = -A^\top(Ax^0(t) - y) + \int_{\mathbb{R}^d} v\nu(dv,t). \]
Hence $x^0$ is exactly the solution studied in (10).

4 Boué Dupuis variational representation

The variational representation of [4] tells us that for any bounded measurable map $h : (\Omega, \mathcal{F}, \mathbb{P}) \to \mathbb{R}$
\[ H^\varepsilon(x(0)) = -\varepsilon^2 \ln \left( \mathbb{E}_{x(0)}[\exp\{-h(x^\varepsilon)\}] \right) \]
\[ = \inf \{ \mathbb{E}_{x(0)}[\frac{1}{2}\int_0^1 \|v(t)\|^2 dt + h(x^{\varepsilon,v})] : v \in \mathcal{A} \}, \quad (12) \]
where $x^{\varepsilon,v}$ denotes the solution
\[ dx^{\varepsilon,v}(t) = \{b(x^{\varepsilon,v}(t)) + v(t)\} dt + \varepsilon dw(t), \quad x^{\varepsilon,v}(0) = x(0). \]

The control $v^\varepsilon \in \mathcal{A}$ such that
\[ H^\varepsilon(x(0)) \geq \mathbb{E}_{x(0)}[\frac{1}{2}\int_0^1 \|v^\varepsilon(t)\|^2 dt + h(x^{\varepsilon,v})] - \varepsilon^2 \]
and the diffusion
\[ dx^{\varepsilon,v^\varepsilon}(t) = \{b(x^{\varepsilon,v^\varepsilon}(t)) + v^\varepsilon(t)\} dt + \varepsilon dw(t), \quad x^{\varepsilon,v^\varepsilon}(0) = x(0) \]
play the central role in the large deviation result [5], and then also in our case. We set
\[ \bar{x}^\varepsilon = x^{\varepsilon,v^\varepsilon}, \quad \bar{x} = \lim_{\varepsilon \to 0} \bar{x}^\varepsilon. \]
Observe that Condition 3.2. in [5]
\[ \sup_{\varepsilon \in (0,1]} \mathbb{E}_{x(0)}[\int_0^1 \|v^\varepsilon(t)\|^2 dt] < +\infty \]
holds also in our case.

7
5 Large deviation : upper bound

We start from the variational representation (13):

\[ H^\varepsilon(x(0)) := -\varepsilon^2 \ln \{ \mathbb{E}_{x_0} [\exp(-\frac{h(x^\varepsilon)}{\varepsilon^2})] \} = \inf_{v \in A} \mathbb{E}_{x_0} [\frac{1}{2} \int_0^1 \| v(t) \|^2 dt + h(x^\varepsilon)] \]

valid for all bounded measurable map \( h \).

From the definition of \( v^\varepsilon \) we have

\[ H^\varepsilon(x(0)) \geq \mathbb{E}_{x_0} [\frac{1}{2} \int_0^1 \| v^\varepsilon(t) \|^2 dt + h(\bar{x}^\varepsilon)] - \varepsilon^2. \]

It follows that

\[ \liminf_{\varepsilon \to 0} H^\varepsilon(x(0)) \geq \liminf_{\varepsilon \to 0} \mathbb{E}_{x_0} [\frac{1}{2} \int_{[0,1] \times \mathbb{R}^d} \| v \|^2 \nu^\varepsilon(dv, t) dt + h(\bar{x})] \]

\[ = \liminf_{\varepsilon \to 0} \mathbb{E}_{x_0} [\frac{1}{2} \int_{[0,1] \times \mathbb{R}^d} \| v \|^2 \nu^\varepsilon(dv, t) dt + h(\bar{x})]. \]

From Fatou lemma we have

\[ \liminf_{\varepsilon \to 0} \mathbb{E}_{x_0} [\int_{[0,1] \times \mathbb{R}^d} \| v \|^2 \nu^\varepsilon(dv, t) dt] = \mathbb{E}_{x_0} [\int_0^1 \liminf_{\varepsilon \to 0} \int_{\mathbb{R}^d} \| v \|^2 \nu^\varepsilon(dv, t) dt]. \]

Using the inequality

\[ \liminf_{\varepsilon \to 0} \int f(v) \mu^n(dv) \geq \int f(v) \mu(dv) \]

valid for all \( f \geq 0 \) measurable and all \( \mu^n \to \mu \) weakly, we obtain

\[ \liminf_{\varepsilon \to 0} \int_{\mathbb{R}^d} \| v \|^2 \nu^\varepsilon(dv, t) \geq \int_{\mathbb{R}^d} \| v \|^2 \nu(dv, t). \]

Finally we have

\[ \liminf_{\varepsilon \to 0} H^\varepsilon(x(0)) \geq \mathbb{E}_{x(0)} [\frac{1}{2} \sum_{i=1}^d \int_{[0,1] \times \mathbb{R}^d} |v_i|^2 \nu(dv, t) dt + h(\bar{x})]. \]

If \( \bar{x}_i(t) < 0 \), then

\[ \frac{d\bar{x}_i}{dt}(t) = b_1^i(\bar{x}(t)) + \int_{\mathbb{R}^d} v_i \nu_1^i(dv, t). \]

If \( \bar{x}_i(t) > 0 \), then

\[ \frac{d\bar{x}_i}{dt}(t) = b_2^i(\bar{x}(t)) + \int_{\mathbb{R}^d} v_i \nu_2^i(dv, t). \]
We also recall that in these cases

$$\nu_i^\eta(d\nu, t)$$

is a probability measure for $$\eta = 1, 2$$. It follows from Jensen inequality that

$$\int_{\mathbb{R}^d} |v_i|^2 \nu_i^\eta(d\nu, t) \geq \left( \int_{\mathbb{R}^d} v_i \nu_i^\eta(d\nu, t) \right)^2 \geq \frac{|d\bar{x}_i(t) - b_i^\eta(\bar{x}(t))|^2 : L_i^\eta(\bar{x}(t), \frac{d\bar{x}_i(t)}{dt}).$$

If $$\bar{x}_i(t) = 0$$, then $$0 < \tilde{\gamma}_i^1(t) < 1$$, and

$$\nu(d\nu, t) = \tilde{\gamma}_i^1(t) \frac{\nu_i^1(d\nu, t)}{\tilde{\gamma}_i^1(t)} + \frac{\tilde{\gamma}_i^2(t) \nu_i^2(d\nu, t)}{\tilde{\gamma}_i^1(t)}.$$

The measure $$\frac{\nu_i^\eta(d\nu, t)}{\tilde{\gamma}_i^\eta(t)}$$ is a probability for each $$\eta = 1, 2$$. Again from Jensen inequality we have

$$\int_{\mathbb{R}^d} |v_i|^2 \frac{\nu_i^\eta(d\nu, t)}{\tilde{\gamma}_i^\eta(t)} \geq \left( \int_{\mathbb{R}^d} \frac{v_i \nu_i^\eta(d\nu, t)}{\tilde{\gamma}_i^\eta(t)} \right)^2.$$

We recall that if $$\bar{x}_i(t) = 0$$, then

$$\beta_i^1(t) = b_i^1(\bar{x}(t)) + \int_{\mathbb{R}^d} v_i \frac{\nu_i^1(d\nu, t)}{\tilde{\gamma}_i^1(t)} \geq 0,$$

$$\beta_i^2(t) = b_i^2(\bar{x}(t)) + \int_{\mathbb{R}^d} v_i \frac{\nu_i^2(d\nu, t)}{\tilde{\gamma}_i^2(t)} \leq 0,$$

and if we denote

$$\beta_i = \frac{d\bar{x}_i}{dt}(t)$$

then

$$\tilde{\gamma}_i^1(t) \beta_i^1(t) + \tilde{\gamma}_i^2(t) \beta_i^2(t) = \beta_i.$$

It follows that

$$\int_{\mathbb{R}^d} |v_i|^2 \nu(d\nu, t) \geq \tilde{\gamma}_i^1(t) |\beta_i^1(t) - b_i^1(\bar{x}(t))|^2 + \tilde{\gamma}_i^2(t) |\beta_i^2(t) - b_i^2(\bar{x}(t))|^2 \geq \sup \{ p^1 |\beta_i^1 - b_i^1(\bar{x}(t))|^2 + p^2 |\beta_i^2 - b_i^2(\bar{x}(t))|^2 \} = L_i^0(\bar{x}(t), \frac{d\bar{x}_i(t)}{dt}).$$

The infimum is taken under the constraint

$$p^1, p^2 > 0, \quad p^1 + p^2 = 1,$$

$$p^1 \beta_i^1 + p^2 \beta_i^2 = \frac{d\bar{x}_i(t)}{dt} := \beta_i.$$
We define
\[
\tilde{L}_i(\bar{x}(t), \frac{dx_i}{dt}(t)) = \frac{d\bar{x}_i}{dt}(t) - b_i^1(\bar{x}(t)), \quad \text{if } \bar{x}_i(t) < 0,
\]
\[
\tilde{L}_i(\bar{x}(t), \frac{dx_i}{dt}(t)) = \frac{d\bar{x}_i}{dt}(t) - b_i^2(\bar{x}(t)), \quad \text{if } \bar{x}_i(t) > 0,
\]
\[
\tilde{L}_i(\bar{x}(t), \frac{dx_i}{dt}(t)) = L_0^i(\bar{x}(t), \frac{dx_i}{dt}(t)), \quad \text{if } \bar{x}_i(t) = 0.
\]

It follows for each \(i\) that
\[
\left( \int_{\mathbb{R}^d} v_i \nu(d\bar{v}, t) \right)^2 \geq \tilde{L}_i(\bar{x}(t), \frac{dx_i}{dt}(t)),
\]
and
\[
\liminf_{\varepsilon \to 0} H^\varepsilon(x(0)) \geq \mathbb{E}_{x(0)} \left[ \frac{1}{2} \sum_{i=1}^d \int_{[0,1]} \tilde{L}_i(\bar{x}(t), \frac{dx_i}{dt}(t)) dt + h(\bar{x}) \right]
\]
\[
\geq \inf \left\{ \frac{1}{2} I(\varphi) + h(\varphi) : \varphi \in C_{x(0)}([0,1]) \right\},
\]
where the rate function
\[
I_{x(0)}(\varphi) := \sum_{i=1}^d \int_0^1 \tilde{L}_i(\varphi(t), \frac{d\varphi_i}{dt}(t)) dt.
\]

The infimum is equal to \(+\infty\) if the latter set is empty.

Finally we have for any sequence \(\varepsilon\) such that \(\bar{x}^\varepsilon \to \bar{x}\), and \((\nu^\varepsilon, \nu^\varepsilon_\eta, \gamma^\varepsilon_\eta, i = 1, \ldots, d, \eta) \to (\nu, \nu^\eta_\eta, \gamma^\eta_\eta, i = 1, \ldots, d, \eta)\) that
\[
\liminf_{\varepsilon \to 0} H^\varepsilon(x(0)) \geq \inf_{\varphi \in C([0,1])} \left\{ \frac{1}{2} I(\varphi) + h(\varphi) \right\}.
\]

Using the same argument as in Boué-Dupuis-Ellis [5] we can show that
\[
\liminf_{\varepsilon \to 0} H^\varepsilon(x(0)) \geq \inf_{\varphi \in C([0,1])} \{ I_{x_0}(\varphi) + h(\varphi) \}
\]
for all \(\varepsilon \in (0, 1]\).

6 Large deviation : lower bound

6.1 Properties of \(L^\eta_0, \eta = 0, 1, 2\)

We define \(L^0_i : \mathbb{R}^d \times \mathbb{R} \to [0, +\infty)\) by
\[
L^0_i(x, \beta_i) = \inf \{ p^1 |\beta_1 - b_1^1(x)|^2 + p^2 |\beta_2 - b_2^2(x)|^2 \}
\]
(15)
The infimum is taken under the constraint

\[p^1, p^2 > 0, \quad p^1 + p^2 = 1, \quad \beta^1_i \geq 0, \quad \beta^2_i \leq 0 \]
\[p^1 \beta^1_i + p^2 \beta^2_i = \beta_i.

**Proposition.** 1) If \((x, \beta_i) \in \mathbb{R}^d \times \mathbb{R}\) are such that

\[b^2_i(x) < \beta_i < b^1_i(x)\]

then

\[L^0_i(x, \beta_i) = 0.\]

2) If \(\beta_i \leq b^2_i(x)\) then

\[L^0_i(x, \beta_i) = |\beta_i - b^2_i(x)|^2.\]

3) If \(\beta_i \geq b^2_i(x)\) then

\[L^0_i(x, \beta_i) = |\beta_i - b^1_i(x)|^2.\]

**Proof.** Observe that

\[p|\beta^1_i - b^1_i(x)|^2 + (1-p)|\beta^2_i - b^2_i(x)|^2 \geq |pb^1_i(x) + (1-p)b^2_i(x) - \beta|^2\]

for any \(p \in (0,1)\) and any couple \(\beta^1_i, \beta^2_i\) such that \(p\beta^1_i + (1-p)\beta^2_i = \beta\). If the infimum (15) is such that

\[L^0_i(x, \beta) = p|\beta^1_i - b^1_i(x)|^2 + (1-p)|\beta^2_i - b^2_i(x)|^2\]

for some \(p \in (0,1), \beta^1_i \geq 0, \beta^2_i \leq 0\), then

\[|\beta^1_i - b^1_i(x)|^2 = |\beta^2_i - b^2_i(x)|^2 = |pb^1_i(x) + (1-p)b^2_i(x) - \beta|^2.\]

Hence

\[L^0_i(x, \beta) = \inf \{|pb^1_i(x) + (1-p)b^2_i(x) - \beta|^2\}\]

where the infimum is also under the same constraint as in (15). This finishes the proof.

**Corollary.** 1) For each \(i = 1, \ldots, d, \eta = 0, 1, 2\), the maps \((x, \beta) \in \mathbb{R}^d \times \mathbb{R} \rightarrow L^0_i(x, \beta)\) are continuous.

2) If \(\beta_i \leq 0\), then \(L^0_i(x, \beta_i) \leq L^2_i(x, \beta_i)\).

3) If \(\beta_i \geq 0\), then \(L^0_i(x, \beta_i) \leq L^1_i(x, \beta_i)\).

4) For each \(i = 1, \ldots, d\) and for \(x\) fixed, the map \(\beta_i \rightarrow L^0_i(x, \beta_i)\) is convex.
Now back to the large deviation lower bound. We are going to show that
\[
\lim_{\varepsilon \to 0} \sup H^\varepsilon(x(0)) \leq I_{\mathcal{X}_0}(\varphi) + h(\varphi)
\]
for all \( \varphi \in \mathcal{N}_0 \). The map \( \varphi \) is defined by \((t_k, \beta(k)) : k = 1, \ldots, r \) such that
\[
\frac{d\varphi_i}{dt}(t) = \beta_i(k), \quad t \in [t_k, t_{k+1}],
\]
\( \varphi_i(t) \neq 0, \quad t \in [t_k, t_{k+1}] \), or \( \varphi_i(t) = 0, \quad t \in [t_k, t_{k+1}] \).

If \( \varphi_i(t) = 0 \) on \([t_k, t_{k+1}]\), then \( \beta_i(k) := \frac{d\varphi_i}{dt}(t) = 0 \) on \([t_k, t_{k+1}]\).

We consider the control
\[
v^1(x, t) = \beta^1(k) - b^1(x), \quad v^2(x, t) = \beta^2(k) - b^2(x)
\]
for \( t \in [t_k, t_{k+1}] \). Here \( \beta^1(k) = -\mu, \beta^2(k) = \mu \) if \( \beta_i(k) = 0 \). Observe that 0 = \( \frac{\beta^1(k) + \beta^2(k)}{2} \). If \( \beta_i(k) \neq 0 \), then \( \beta^1_i(k) = \beta^2_i(k) = \beta_i(k) \).

Now define for \( i = 1, \ldots, d \),
\[
v_i(x, t) = v^1_i(x, t)1_{[x_i \leq 0]} + v^2_i(x, t)1_{[x_i > 0]},
\]
and \( v(x, t) \) denotes the vector column with the components \( v_i(x, t) \). The controlled process
\[
dx^{\varepsilon, \varphi}_i(t) = b_i(x^\varepsilon(t))dt + v_i(x^\varepsilon(t), t)dt + \varepsilon dw_i(t)
\]
\[
= \{\beta^2_i(k)1_{[x^\varepsilon_i(t) > 0]} + \beta^1_i(k)1_{[x^\varepsilon_i(t) \leq 0]}\}dt + \varepsilon dw_i(t)
\]
(16)

Let us define
\[
f_i(k) = \begin{cases} 
\beta_i(k) + \mu, & t \in [t_k, t_{k+1}], \\
\beta_i(k) - \mu, & t \in [t_k, t_{k+1}], \\
0, & t \in [t_k, t_{k+1}],
\end{cases}
\varphi_i(t_k) > 0,
\varphi_i(t_k) < 0,
\varphi_i(t_k) = 0.
\]

Then we can rewrite (16) as
\[
dx^{\varepsilon, \varphi}_i(t) = \{f_i(t) - \mu \text{sgn}(x^{\varepsilon, \varphi}_i(t))\}dt + \varepsilon dw_i(t)
\]
where
\[
f_i(t) = f_i(k), \quad t \in [t_k, t_{k+1}].
\]

It follows from (9) that \( x^{\varepsilon, \varphi} \) converges to \( \varphi \) as \( \varepsilon \to 0 \).

From the variational representation
\[
H^\varepsilon(x(0)) = \inf_{\varphi \in \mathcal{A}} \mathbb{E}_{x(0)}\left[ \frac{1}{2} \int_{0}^{1} \|v(t)\|^2 dt + h(x^\varepsilon, v) \right]
\]
we have
\[
\limsup_{\varepsilon \to 0} H^\varepsilon(x(0)) \leq \limsup_{\varepsilon \to 0} \mathbb{E}_{x(0)} \left[ \frac{1}{2} \int_0^1 \| \frac{dx}{dt}(t) - b(x^\varepsilon, \phi(t)) \|^2 dt + h(x^\varepsilon, \phi) \right]
\]
\[
= \left[ \frac{1}{2} \int_0^1 \| \frac{dx}{dt}(t) - b(\phi(t)) \|^2 dt + h(\phi) \right],
\]
where
\[
\int_0^1 \| \frac{dx}{dt}(t) - b(\phi(t)) \|^2 dt = \sum_{i=1}^d \int_0^1 \| \frac{dx_i}{dt}(t) - b_i(\phi(t)) \|^2 dt
\]
and
\[
b_i(\phi(t)) = b_1^i(\phi(t)), \quad \text{if } \phi_i(t) < 0,
b_i(\phi(t)) = b_2^i(\phi(t)), \quad \text{if } \phi_i(t) > 0,
\]
in these cases
\[
\left| \frac{dx_i}{dt}(t) - b_i(\phi(t)) \right|^2 = \tilde{L}_i(\phi(t), \frac{dx_i}{dt}(t)).
\]
Moreover, on each interval \([t_k, t_{k+1})\) such that \(\beta_i(k) = 0\) we can show that
\[
\left| \frac{dx_i}{dt}(t) - b_i(\phi(t)) \right|^2 := \left| b_1^i(\beta(k)) \right|^2 \quad \text{if } b_1^i(\beta(k)) \leq 0,
\]
\[
\left| \frac{dx_i}{dt}(t) - b_i(\phi(t)) \right|^2 := \left| b_2^i(\beta(k)) \right|^2 \quad \text{if } b_2^i(\beta(k)) \geq 0,
\]
\[
\left| \frac{dx_i}{dt}(t) - b_i(\phi(t)) \right|^2 := 0, \quad b_2^i(\beta(k)) < 0 < b_1^i(\beta(k)).
\]
More precisely, if \(\beta_i(k) = 0\) then
\[
\left| \frac{dx_i}{dt}(t) - b_i(\phi(t)) \right|^2 = \tilde{L}_i^0(\phi(t), \frac{dx_i}{dt}(t)).
\]
Finally we have
\[
\int_0^1 \| \frac{dx}{dt}(t) - b(\phi(t)) \|^2 dt = \int_0^1 \tilde{L}(\phi(t), \frac{dx}{dt}(t)) dt
\]
and then for all \(\phi \in \mathcal{N}_0\)
\[
\limsup_{\varepsilon \to 0} H^\varepsilon(x(0)) \leq I_{x(0)}(\phi) + h(\phi).
\]
To finish the proof of the large deviation’s lower bound we need the following lemmas. The proof is the same as in Dupuis and Ellis book [13]. For the sake of completeness we will recall the proof.
Lemma 1. Let \( \psi \in C_x([0, 1]) \) such that \( \frac{1}{0} \int \tilde{L}(\psi(t), \psi'(t))dt < +\infty \). For each \( \delta > 0 \), there exist \( \vartheta > 0 \) and \( \xi \in C_x([0, 1]) \) with the following properties:

\[
\sup_{t \in [0, 1]} |\xi(t) - \psi(t)| \leq \delta, \quad \int_0^1 \tilde{L}(\xi(t), \frac{d\xi}{dt}(t))dt \leq \int_0^1 \tilde{L}(\psi(t), \frac{d\psi}{dt}(t))dt + \delta
\]

and

\[
\sup_{t \in [0, 1]} \left\| \frac{d\xi}{dt}(t) \right\| \leq \vartheta.
\]

Now we prove the following result.

**Proof.** Let \( c, \lambda \in (0, 1) \). We define

\[
D_\lambda = \{ t \in [0, 1] : \| \frac{d\psi}{dt}(t) \| \geq \frac{1}{\lambda} \},
\]

\[
E_\lambda = \{ t \in [0, 1] : \| \frac{d\psi}{dt}(t) \| < \frac{1}{\lambda} \}.
\]

We construct the time-rescaling map \( S_\lambda : [0, 1] \to [0, +\infty) \) as follows

\[
S_\lambda(0) = 0, \quad \frac{dS_\lambda}{dt}(t) = \begin{cases} \frac{\| \frac{d\psi}{dt}(t) \|}{c(1 - \lambda)} & \text{if } t \in D_\lambda, \\ \frac{1}{(1 - \lambda)} & \text{if } t \in E_\lambda. \end{cases}
\]

Clearly the map \( S_\lambda : [0, 1] \to [0, S_\lambda(1)] \) is one to one with \( S_\lambda(1) > 1 \). Its inverse \( T_\lambda : [0, S_\lambda(1)] \to [0, 1] \). For \( s \in [0, S_\lambda(1)] \) we define

\[
\xi_\lambda(s) = \psi(T_\lambda(s)).
\]

On the one hand

\[
\left\| \frac{d\xi_\lambda}{dt}(t) \right\| \leq \max(c(1 - \lambda), \frac{1}{\lambda}).
\]

On the other hand the hypothesis

\[
\int_0^1 \tilde{L}(\psi(t), \frac{d\psi}{dt}(t))dt := \sum_{i=1}^d \int_0^1 \tilde{L}_i(\psi(t), \frac{d\psi_i}{dt}(t))dt < +\infty
\]

implies that for each \( i \)

\[
\int_0^1 \left| \frac{d\psi_i}{dt}(t) - b_i(\psi(t)) \right| dt < +\infty,
\]

for \( \eta = 1, 2 \). From the triangular inequality

\[
\int_0^1 \left| \frac{d\psi_i}{dt}(t) \right| dt \leq \int_0^1 \left| \frac{d\psi_i}{dt}(t) - b_i(\psi(t)) \right| dt + \int_0^1 \left| b_i(\psi(t)) \right| dt
\]
we derive for each $i$ that
\[
\int_0^1 \| \frac{d\psi}{dt}(t) \| dt < +\infty.
\]
Now the rest of the proof is the same as in Dupuis et al. We prove
\[
\sup_{t \in [0,1]} \| \xi(\lambda)(t) - \psi(t) \| \to 0
\]
and
\[
\int_0^1 \tilde{L}(\xi(\lambda)(t), \frac{d\xi}{dt}(t)) dt \to \int_0^1 \tilde{L}(\psi(t), \frac{d\psi}{dt}(t)) dt
\]
as $\lambda \to 0$, which achieves the proof of Lemma 1.

**Lemma 2.** Let $\xi \in C_{x(0)}([0,1])$ such that $\int_0^1 \tilde{L}(\xi(t), \frac{d\xi}{dt}(t)) dt < +\infty$ and
\[
\sup_{t \in [0,1]} \| \frac{d\xi}{dt}(t) \| \leq \vartheta.
\]
For any $\delta > 0$ there exists $\sigma > 0$ and $\varphi^\sigma \in N_0$ such that
\[
\sup_{t \in [0,1]} |\xi(t) - \varphi^\sigma(t)| \leq \delta, \quad \text{and} \quad \int_0^1 \tilde{L}(\varphi^\sigma(t), \frac{d\varphi^\sigma}{dt}(t)) dt \leq \int_0^1 \tilde{L}(\xi(t), \frac{d\xi}{dt}(t)) dt + 2\delta
\]

**Proof.**
We define for each $i$
\[
G^0_i = \{ t \in [0,1] : \xi_i(t) = 0 \},
\]
\[
G^1_i = \{ t \in [0,1] : \xi_i(t) < 0 \},
\]
\[
G^2_i = \{ t \in [0,1] : \xi_i(t) > 0 \}.
\]
Following Dupuis et al. for all $\sigma > 0$ there exists $B^0_i = \bigcup_{j=1}^{I_i} [c_j(i), d_j(i)]$ such that $\xi_i(c_j(i)) = \xi_i(d_j(i)) = 0$ $d_j(i) - c_j(i) \leq \sigma$, $d_j(i) \leq c_{j+1}(i)$. We suppose that $c_1(i) = 0$ and $d_1(i) = 1$. We choose finitely many numbers $(e_{j}^{k}(i), k = 1, \ldots, K_j(i))$ such that
\[
d_j(i) = e_{j}^{1}(i) < \ldots < e_{j}^{K_j(i)}(i) = c_{j+1}(i)
\]
and
\[
e_{j}^{k+1}(i) - e_{j}^{k}(i) < \sigma.
\]
We define for each $i$ the function $\varphi^\sigma_i$ is piecewise linear interpolation of $\xi_i$ with interpolation points
\[
\{ c_j(i), d_j(i), e_j^k(i) : \ j = 1, \ J_i, k = 1, \ldots, K_j(i) \}.
\]
For each $\delta > 0$ there exists $\sigma_1 > 0$ such that for $0 \leq \sigma \leq \sigma_1$ implies that
$$
\sup_{t \in [0,1]} |\xi_i(t) - \varphi_i^{\sigma}(t)| \leq \delta.
$$
Clearly $\varphi_i^{\sigma}(t) = 0$ for all $t \in [c_j(i), d_j(i)]$, and $\varphi_i^{\sigma}(t) \neq 0$ for all $t \in [e_j(i), e_j^{k+1}(i)]$.

We define
\begin{align*}
a_j^{\eta}(i) &= \int_{c_j(i)}^{d_j(i)} 1_{G_i^{\eta}}(t) dt, \\
\beta_j^{\eta}(i) &= \frac{1}{a_j^{\eta}(i)} \int_{c_j(i)}^{d_j(i)} \frac{d\xi_i}{ds} 1_{G_i^{\eta}}(s) ds.
\end{align*}
Clearly
$$
\sum_{\eta=0,1,2} a_j^{\eta}(i) = d_j(i) - c_j(i).
$$
Since $\xi_i(t) = 0$ implies $\frac{d\xi_i}{dt}(t) = 0$ a.s. then
$$
\beta_j^{0}(i) = 0.
$$
Since $G_i^1$, $G_i^2$ are open, then it can be written as a countable union of open intervals at each endpoint of which $\xi_i = 0$. Hence $\beta_j^{\eta}(i) = 0$ for $\eta = 1, 2$. It follows that
\begin{align*}
L_i^1(\xi(c_j(i)), 0) &\geq L_i^0(\xi(c_j(i)), 0), \\
L_i^2(\xi(c_j(i)), 0) &\geq L_i^0(\xi(c_j(i)), 0).
\end{align*}
We have for each $i, j$ that
$$
\int_{c_j(i)}^{d_j(i)} \bar{L}_i(\xi(t), \frac{d\xi_i}{dt}(t)) dt = \sum_{\eta=0,1,2} \int_{c_j(i)}^{d_j(i)} 1_{G_i^{\eta}}(t)L_i^{\eta}(\xi(t), \frac{d\xi_i}{dt}(t)) dt.
$$
From the continuity of $L_i^{\eta}$ and $\xi$ and the fact that $\sup_{t \in [0,1]} \|\frac{d\xi_i}{dt}(t)\| \leq \vartheta$ there exists $\sigma_2 \leq \sigma_1$ such that $\sigma \leq \sigma_2$ implies
$$
\sup_{t \in [0,1]} |L_i^{\eta}(\xi(t), \frac{d\xi_i}{dt}(t)) - L_i^{\eta}(\xi(c_j(i)), \frac{d\xi_i}{dt}(t))| \leq \delta.
$$
It follows that
$$
\int_{c_j(i)}^{d_j(i)} 1_{G_i^{\eta}}(t)L_i^{\eta}(\xi(t), \frac{d\xi_i}{dt}(t)) dt \geq \int_{c_j(i)}^{d_j(i)} 1_{G_i^{\eta}}(t)L_i^{\eta}(\xi(c_j(i)), \frac{d\xi_i}{dt}(t)) dt - (d_j(i) - c_j(i))\delta.
$$
From the convexity of the function $\beta_i \rightarrow L_i^{\eta}(x, \beta_i)$ for each $x$ fixed and for each $\eta = 1, 2$, we have
$$
\frac{1}{a_j^{\eta}(i)} \int_{c_j(i)}^{d_j(i)} 1_{G_i^{\eta}}(t)L_i^{\eta}(\xi(c_j(i)), \frac{d\xi_i}{dt}(t)) dt \geq L_i^{\eta}(\xi(c_j(i)), \beta_j^{\eta}(i)) = L_i^{\eta}(\xi(c_j(i)), 0) \geq L_i^{\eta}(\xi(c_j(i)), 0).
$$
For $\eta = 0$, we have
\[ \int_{c_{j}(i)}^{d_{j}(i)} 1_{C_{i}^{0}}(t)L_{i}^{0}(\xi(c_{j}(i)), 0)dt = a_{j}^{0}(i)L_{i}^{0}(\xi(c_{j}(i)), 0). \]

Finally, we have
\[ \int_{c_{j}(i)}^{d_{j}(i)} \tilde{L}_{i}(\xi(t), \frac{d\xi}{dt}(t))dt \geq (d_{j}(i) - c_{j}(i))L_{i}^{0}(\xi(c_{j}(i)), 0) - \delta(d_{j}(i) - c_{j}(i)). \tag{17} \]

Observe that
\[ \xi_{i}(c_{j}(i)) = \varphi_{i}^{0}(c_{j}(i)), \]
but there is no guaranty that
\[ \xi_{l}(c_{j}(i)) = \varphi_{l}^{0}(c_{j}(i)) \]
for $l \neq i$. However there exist $\alpha_{l} \leq c_{j}(i) \leq \beta_{l}$ such that
\[ \xi_{l}(\alpha_{l}) = \varphi_{l}^{0}(\alpha_{l}), \quad \xi_{l}(\beta_{l}) = \varphi_{l}^{0}(\beta_{l}), \]
\[ \beta_{l} - \alpha_{l} \leq \sigma. \]

More precisely
\[ \alpha_{l} = c_{j_{l}}(l), \quad \beta_{l} = d_{j_{l}}(l) \]
for some $j_{l}$ or
\[ \alpha_{l} = e_{j_{l}}^{k_{l}}(l), \quad \beta_{l} = e_{j_{l}}^{k_{l}+1}(l). \]

It follows that for small $\sigma$ we have
\[ \xi_{l}(c_{j}(i)) \approx \varphi_{l}^{0}(c_{j}(i)) \approx \varphi^{0}(t), \quad \forall t \in [c_{j}(i), d_{j}(i)]. \]

Then for small $\sigma$ we have
\[ L_{i}^{0}(\xi(c_{j}(i)), 0) \geq L_{i}^{0}(\varphi^{0}(t), 0) - \delta, \quad \forall t \in [c_{j}(i), d_{j}(i)]. \]

Now the inequality (17) becomes
\[ \int_{c_{j}(i)}^{d_{j}(i)} \tilde{L}_{i}(\varphi^{0}(t), \frac{d\varphi^{0}}{dt}(t))dt \geq \int_{c_{j}(i)}^{d_{j}(i)} L_{i}^{0}(\varphi^{0}(t), \frac{d\varphi^{0}}{dt}(t))dt - 2\delta(d_{j}(i) - c_{j}(i)) \]
\[ = \int_{c_{j}(i)}^{d_{j}(i)} \tilde{L}_{i}(\varphi^{0}(t), \frac{d\varphi^{0}}{dt}(t))dt - 2\delta(d_{j}(i) - c_{j}(i)). \]
The same proof shows that
\[ \int_{e^{k+1}(i)}^{e^{k+1}(i)} \tilde{L}_i(\xi(t), \frac{d\xi}{dt}(t))dt \geq \int_{e^{k+1}(i)}^{e^{k+1}(i)} \tilde{L}_i(\varphi^\gamma(t), \frac{d\varphi^\gamma}{dt}(t))dt - 2\delta(e^{k+1}(i) - e^k(i)). \]

Finally
\[
\int_0^1 \tilde{L}(\xi(t), \frac{d\xi}{dt}(t))dt = \sum_{i=1}^d \int_0^1 \tilde{L}_i(\xi(t), \frac{d\xi}{dt}(t))dt = \sum_{i=1}^d \sum_{j=1}^{J_i} \int_{e^{j}(i)}^{e^{j}(i)} \tilde{L}_i(\xi(t), \frac{d\xi}{dt}(t))dt + \sum_{k=1}^{K_i(i)} \int_{e^{k+1}(i)}^{e^{k+1}(i)} \tilde{L}_i(\xi(t), \frac{d\xi}{dt}(t))dt \]
\[
\geq \int_0^1 \tilde{L}(\varphi^\gamma(t), \frac{d\varphi^\gamma}{dt}(t))dt - 2\delta.
\]

Now back to the upper bound. Let \( \tau > 0 \) and \( \psi \in C_{X(0)}([0,1]) \) such that
\[ I_{X(0)}(\psi) \leq \inf \{ I_{X(0)}(\varphi) : \varphi \in C([0,1]) \} + \tau, \]
and then
\[ \limsup_{\varepsilon \to 0} H^\varepsilon(x(0)) \leq \inf \{ I_{X(0)}(\varphi) + h(\varphi) : \varphi \in C([0,1]) \} + 2\tau \]
for any \( \tau > 0 \).

Finally we obtain for any continuous and bounded map \( h \) that
\[ \lim_{\varepsilon \to 0} H^\varepsilon(x(0)) = \inf \{ I_{X(0)}(\varphi) + h(\varphi) : \varphi \in C([0,1]) \}. \tag{18} \]

7 Different large deviation formulation

In general, a family of probability measure \( (\mathbb{P}^\varepsilon, \varepsilon > 0) \) on a metric space \((X, \Delta)\) satisfies the large deviation principle (LDP) with the rate function \( I \) if the following conditions are satisfied:

a) \( I : X \to [0, +\infty) \) is lower semi-continuous,

b) For each \( r > 0 \), \( \{ x \in X : I(x) \leq r \} \) is precompact,

c) For any \( R > 0 \), there exists a compact set \( K \) such that for any \( \delta > 0 \), we have for small \( \varepsilon \),
\[ \mathbb{P}^\varepsilon(B^\varepsilon_\delta(K)) \leq \exp(-\frac{R^2}{\varepsilon^2}), \]

d) \( \lim_{\varepsilon \to 0} \liminf_{\varepsilon \to 0} \varepsilon^2 \ln \{ \mathbb{P}^\varepsilon(B_\delta(x)) \} = \lim_{\varepsilon \to 0} \limsup_{\varepsilon \to 0} \varepsilon^2 \ln \{ \mathbb{P}^\varepsilon(B_\delta(x)) \} = -I(x) \).

Here \( B_\delta(K) \) denotes the \( \delta \)-neighborhoods of any compact set \( K \), and \( B^\varepsilon_\delta(K) \) its complement.

Let \( \mathbb{P}^\varepsilon_{X(0)} \) be the probability distribution of \( x^\varepsilon \) (1) starting from \( x(0) \). It’s known that the limit (18) is equivalent to say that the family \( (\mathbb{P}^\varepsilon_{X(0)} : \varepsilon > 0) \) satisfies the LDP with the rate function \( I_{X(0)} \) (14). See [11] for a general theory of the large deviation principle.
Références

[1] S. Albeverio, V.I. Bogachev, M. Röckner. On uniqueness of invariant measures for finite and infinite dimensional diffusions Comm. Pure Appl. Math., 52 (1999), pp. 325-362.

[2] A. Baule, E.G.D. Cohen, H. Touchette, A path integral approach to random motion with nonlinear friction, J. Phys. A : Math. Theor. 43, (2010).

[3] S. Bobkov. Isoperimetric and analytic inequalities for log-concave probability measures. Ann. Probab. 27(4) (1999), 1903-1921.

[4] M. Boué, P. Dupuis. A variational representation for certain functionals of Brownian motion. Ann. Probab. 26 (1998), 1641-1659.

[5] M. Boué, P. Dupuis, R.S. Ellis. Large deviations for small noise diffusions with discontinuous statistics. Prob. Theor. and Rel. Fields 116 (2000), pp. 125-148.

[6] R. Buckdahn, Y. Ouknine, M. Quincampoix. On limiting values of stochastic differential equations with small noise intensity tending to zero. Bull. Sci. Math. 133 (2009), no. 3, 229-237.

[7] E. Cépa. Problème de Skorohod multivoque. Ann. Probab. 26, no. 2 (1998), 500-532.

[8] E. Cépa. Équations différentielles stochastiques multivoques. Séminaire de Probabilités, XXIX, 86-107, Lecture Notes in Math., 1613, Springer, Berlin (1995).

[9] X. Chen. Stochastic differential inclusions, Thesis (2006), University of Edinburgh.

[10] S. S. Chen, D. L. Donoho, M. A. Saunders. Atomic Decomposition by Basis Pursuit. SIAM J. Sci. Comput., 20(1) (1998), 33-61.

[11] A. Dembo, O. Zeitouni. Large Deviations Techniques and Applications. Jones and Bartlett, Boston (1992).

[12] A. Dermoune, D. Ounaissi and N. Rahmania. Oscillation of Metropolis-Hasting and simulated annealing algorithms around penalized least squares estimator, Math. Comput. Simulation (2015).

[13] P. Dupuis, and R. Ellis (1997). A Weak Convergence Approach to the Theory of Large Deviations. Wiley, New York.

[14] R. Kannan, L. Lovasz, M. Simonovits. Isoperimetric Problems for Convex Bodies and a Localization Lemma. Discrete Comput. Geom. 13 (1995), 541-559.

[15] P. Kree. Diffusion equation for multivalued stochastic differential equations, J. Funct Anal. 49 (1982), pp. 73-90.

[16] D. Ounaissi. Méthodes quasi-Monte Carlo et Monte Carlo : Application aux calculs des estimateurs lasso et lasso bayésien, Thesis (2016), University of Lille 1.

[17] R. Pettersson. Existence theorem and Wong-Zakai approximations for multivalued stochastic differential equations. Probability and Mathematical Statistics, 17 (1997), pp. 29-45.
[18] R. J. Tibshirani. Regression shrinkage and selection via LASSO, J.R. Stat. Soc. Ser. B Stat. Methodol. vol. 58 no. 1, 267-288. The lasso problem and uniqueness. Electron. J. Stat. 7 (1996), 1456-1490.

[19] R. J. Tibshirani. The lasso problem and uniqueness. Electron. J. Stat. 7 (2013), 1456-1490.