filt36h3.tex

On continuous time ergodic filters with wrong initial data (2nd revised version)

A. Yu. Veretennikov* & M. L. Kleptsyna†

December 26, 2006

Abstract

For a class of non-uniformly ergodic Markov diffusions, under observations subject to a Wiener noise, it is shown that a wrong initial data is forgotten with a certain rate in certain topologies.

Keywords: Markov diffusion filters, initial data robustness, misspecified problems

MSC2000 classification: 93E11, 60J60, 60G35, 93E15

1 Introduction

We consider a continuous time filter for a Markov diffusion (X_t) with values in the Euclidean space $\mathbb{R}^d$, with observations (Y_t) from $\mathbb{R}^\ell$, satisfying the following system of nonlinear Itô's equations,

$$dX_t = b(X_t)dt + dW_t, \quad t \geq 0, \tag{1}$$

$$dY_t = h(X_t)\,dt + dB_t \quad t \geq 0, \tag{2}$$

with initial data X_0 and $Y_0 = 0$, where (W_t, B_t) is a $(d+\ell)$-dimensional Wiener process. Here $b(\cdot)$ is a d-dimensional vector-function, $h(\cdot)$ an ℓ-dimensional vector-function, random variable X_0 is independent on W and B.

The exact initial distribution of X_0 is denoted by μ_0, and the main question addressed here is about an asymptotical behaviour of the filter if this initial distribution is not known. Under *uniform ergodicity* assumptions, [1] for continuous

*University of Leeds, UK, and Institute of Information Transmission Problems, Moscow, Russia, e-mail: A.Veretennikov at leeds.ac.uk

†Université du Maine, Le Mans, France, e-mail: Marina.Kleptsyna at univ-lemans.fr

time, [17], [6] for discrete time have established a limiting independence of the optimal filter algorithm on a wrong initial data, along with certain bounds. The problem has also been discussed in [15], [19], et al.

Non-compact systems attracted attention in the last decade. Small noise in observations was tackled in [4]. In [3], observation noise is assumed to be bounded, with some additional condition on the support of its density. In [17] restrictive mixing or "pseudo-mixing" conditions on the conditional kernel are assumed, which, in particular, do not allow Gaussian noise, that is, of course, the most desirable example; the pseudo-mixing condition, in particular, includes a requirement that the observation noise is small enough. In [21], a variational approach applied to the filtering systems with a gradient type drift and linear observation part under additional assumptions. Neither of these restrictions is assumed here.

We have suggested an approach that works in a similar situation for rather general discrete time filters in [9], and in this paper present a version suitable for continuous time. The use of Birkhoff projective metric is as useful here as in compact case. Although the basic "discrete" ideas remain the same, the transition to continuous time is not straightforward, because, at some point, one has to work with densities in an infinite-dimensional space. In particular, some technical steps here are quite different from their analogues in [9]. The local contraction, or local mixing condition here is guaranteed by the Harnack inequality; its analogue in the discrete case was just due to the assumptions on the density of the noise in the signal. Notice that in earlier papers on continuous time, mixing properties were ensured by other density estimates; however, the Harnack inequality seems to be a proper and most general tool here, as it proved to be in the problem of mixing for non-conditional dynamics. One more obstacle to overcome here is that in discrete time case, a "local mixing condition" on the non-conditional kernel immediately implies a similar condition for conditional kernels, *with the same constant of mixing*. Here the latter seems to be impossible, because of the integration in the space of trajectories, say, over a unit interval, which does not appear in discrete case; to tackle this difficulty we impose some smoothness assumption on h. It is, of course, very probable that this approach could be useful for filtering diffusion systems more general than (1)–(2), because the Harnack inequality itself is known for much more general systems, see [12] or [14]. However, we start with a more simple case, in order not to overload the presentation.

The paper is arranged as follows: the section 2 contains the assumptions and the main results; the section 3 is devoted to the proof of the first main result, under the assumption (A3); the section 4 is devoted to the proof of the second main result, without (A3) but under more restrictive recurrence and other conditions. This presentation is independent from [9]; however, we found it fruitful to consider continuous case in order to understand better the discrete one, and vice versa.

2 Assumptions and main result

2.1 Assumptions

The first group of assumptions deals with the case of exponential or polynomial recurrence, and the function h is allowed to grow with any linear rate, under the only assumption that ∇h is bounded. The absolute continuity condition (A3) is most important. These assumptions are used in the Theorem 1.

(A1) We assume that the function b is locally bounded, and there exist $p = 0, 1$, $M > 0$ and $r \in (0, +\infty]$ such that

$$\left\langle b(x), \frac{x}{|x|^{1-p}} \right\rangle \le -r, \quad |x| \ge M; \tag{3}$$

in the case of $p = 1$, we understand this as a limit necessarily with $r = +\infty$, that is,

$$\lim_{|x|\to\infty} \langle b(x), x\rangle = -\infty. \tag{4}$$

(A2) The function h satisfies the condition,

$$h \in C^2, \quad \& \quad \|\nabla h\|_{C^1} < \infty.$$

(A3) The measure μ_0 is absolutely continuous with respect to ν_0, and

$$\left\|\frac{d\mu_0}{d\nu_0}\right\|_{L_\infty(\nu_0)} \le C < \infty.$$

Moreover, both initial measures μ_0 and ν_0 possess some exponential moment, that is, there exists $c > 0$ such that

$$\int e^{c|x|}(\mu_0(dx) + \nu_0(dx)) < \infty. \tag{5}$$

Remark 1. The assumption on exponential moments for initial measures in (A3) can be relaxed in the case $p = 1$. In (A3) the uniform bound in $L_\infty(\nu_0)$ can be replaced by any $L_a(\nu_0)$, $a > 1$; this would just change the constants in the final bounds, but not the exponential or better than any polynomial rate of convergence. Notice, however, that under $a = 1$ we can only get convergence in probability, without any useful rate. One special case when some convergence rate is still available with $a = 1$ is provided by the Theorem 2 below. There are tools to relax the assumption $\|\nabla h\|_{C^1} < \infty$ in order to allow a certain growth of $\|\nabla^2 h\|$ at infinity, in the spirit of large deviations, however, we do not pursue this here.

The second group of assumptions is used in the case when (A3) may fail, and will be used in the Theorem 2.

(A′1) We assume that

$$\lim_{|x|\to\infty} \left\langle b(x), \frac{x}{|x|} \right\rangle = -\infty; \tag{6}$$

the vector-function b is locally bounded.

(A′2) The vector-function h is bounded, and, moreover, $h \in C_b^2$, that is, bounded along with its (continuous) partial derivatives up to the second order.

(A′3) The measure μ_0 is absolutely continuous with respect to ν_0:

$$\mu_0 << \nu_0.$$

Moreover, both initial measures μ_0 and ν_0 possess all exponential moments, that is, for any $c > 0$,

$$\int e^{c|x|}(\mu_0(dx) + \nu_0(dx)) < \infty. \tag{7}$$

2.2 Setting and main results

In order to explain the setting, we have to start with the algorithm that solves the "exact filtering problem". This algorithm depends, both explicitly and implicitly, on the initial data. "Wrong initial data" means that we are going to plug in a new initial measure instead of the exact one, in this algorithm. To explain this accurately, we need to show the details of the latter.

The filtered or conditional measure for X_t given the observations $Y = (Y_s, 0 \le s \le t)$ is constructed as follows. Consider the following family of stochastic exponentials,

$$\rho_{0,t}(X, Y) = \exp(\int_0^t h(X_s)\, dY_s - (1/2)\int_0^t h^2(X_s)\, ds),$$

where $h^2 = |h|^2$, $hY \equiv h^*Y$ (here * means transposition), and

$$\gamma_{0,t}(X, Y) = \exp(-\int_0^t h(X_s)\, dB_s - (1/2)\int_0^t h^2(X_s)\, ds).$$

Since X and B are independent, γ is a probability density.

Notice that $\rho = \gamma^{-1}$, both with respect to the original measure $\mathbb{P}$, and $\mathbb{P}^\gamma$, where $d\mathbb{P}^\gamma = \gamma\, d\mathbb{P}$, see [18]. Then, due to the Bayes – Kallianpur – Striebel formula, see [8, chapter 11],

$$\mathbb{E}_{\mu_0}(f(X) \mid Y) = \frac{\mathbb{E}^\gamma_{\mu_0}(f(X)\rho_{0,t}(X,Y) \mid Y)}{\mathbb{E}^\gamma_{\mu_0}(\rho_{0,t}(X,Y) \mid Y)}\,. \tag{8}$$

Here and in the sequel, any expression like $\mathbb{E}(\cdot \mid Y)$ means $\mathbb{E}(\cdot \mid \mathcal{F}_t^Y)$. This should not lead to any confusion because t is fixed throughout the paper. Moreover, this convention is reasonable, for when we claim that the sigma-algebra $\mathcal{F}_t^Y$ is given, this is always interpreted as though we know the trajectory $(Y_s,\, 0 \le s \le t)$, so that the idenfication of $\mid Y$ and $\mid \mathcal{F}^Y_t$ is, indeed, natural. Notice that $f(X)$ here may denote a function of X_t, or, more generally, a function of the whole trajectory X_s, $0 \le s \le t$. For the reader's convenience and for methodical purposes, we remind a short proof of the latter formula.

For any bounded measurable function $g = g(Y)$ we can check the definition of the conditional probability as follows,

$$\mathbb{E}_{\mu_0} g \frac{\mathbb{E}^\gamma_{\mu_0}\left(f(X)\gamma^{-1} \mid Y\right)}{\mathbb{E}^\gamma_{\mu_0}(\gamma^{-1} \mid Y)} = \mathbb{E}^\gamma_{\mu_0} \gamma^{-1} g \frac{\mathbb{E}^\gamma_{\mu_0}\left(f(X)\gamma^{-1} \mid Y\right)}{\mathbb{E}^\gamma_{\mu_0}(\gamma^{-1} \mid Y)}$$

$$= \mathbb{E}^\gamma_{\mu_0} \left(g \frac{\mathbb{E}^\gamma_{\mu_0}\left(f(X)\gamma^{-1} \mid Y\right)}{\mathbb{E}^\gamma_{\mu_0}(\gamma^{-1} \mid Y)} \left(\mathbb{E}^\gamma_{\mu_0}\left(\gamma^{-1}\right) \mid Y\right) \right)$$

$$= \mathbb{E}^\gamma_{\mu_0}\left(g\, \mathbb{E}^\gamma_{\mu_0}\left(f(X)\gamma^{-1} \mid Y\right)\right) = \mathbb{E}^\gamma_{\mu_0}\left(g\, f(X)\gamma^{-1}\right) = \mathbb{E}_{\mu_0}\left(g\, f(X)\right).$$

Hence, we get the formula (8), as required. As explained below, both parts of the ratio in (8) can be regarded as *continuous functions* with respect to the trajectory Y in the topology of uniform continuous function space. Our filtering algorithm will use exactly these versions.

Now, let us define the following operator S^Y_t,

$$\mu S^Y_t(A) := \int \mathbb{E}^\gamma_{x_0}\left(1(X_t \in A)\rho_{0,t}(X,Y) \mid Y\right) \mu(dx_0).$$

Then, by (8), for every $t > 0$, $\mathbb{P}_{\mu_0}$-almost surely,

$$\mathbb{P}_{\mu_0}(X_t \in \cdot \mid Y) = d^{\mu_0}_t(Y) \mu_0 S^Y_t(\cdot), \tag{9}$$

where $d^{\mu_0}_t(Y)$ is a normalization constant, that is,

$$d^{\mu_0}_t(Y)^{-1} = \mu_0 S^Y_t(\mathbb{R}^d).$$

Moreover, $\mathbb{P}$-a.s.,

$$\mathbb{E}^\gamma_{\mu_0}(f(X)\rho_{0,t}(X,Y) \mid Y) = \langle \mu_0 S^Y_t, f\rangle,$$

and

$$\mathbb{E}^\gamma_{\mu_0}(\rho_{0,t}(X,Y) \mid Y) = \langle \mu_0 S^Y_t, 1\rangle;$$

here the first equality simply means that

$$\mathbb{E}^\gamma_{\mu_0}(\rho_{0,t}(X,Y) 1(X_t \in \cdot) \mid Y) = \mu_0 S^Y_t(\cdot).$$

Notice that we can consider the function $\mathbb{P}_{\mu_0}(X_t \in \cdot \mid Y)$ as a measure due to the existence of regular conditional distributions with a "desintegration" property, – that is, Chapman – Kolmogorov's equations for any finite set of integer nonrandom moments, – cf., e.g., [20], or [13]. Hence, we can identify the right and left hand sides in (9).

Similarly, we can define the filtering measure with a wrong initial data as

$$d_t^{\nu_0}(Y)\nu_0 S_t^Y(\cdot), \quad \text{that is,} \quad d_t^{\nu_0}(\tilde Y)\nu_0 S_t^{\tilde Y}(\cdot)\mid_{\tilde Y = Y},$$

where $(\tilde X, \tilde Y)$ is a solution of the same SDE system (1)–(2) with a new initial measure ν_0 on some *independent* probability space; this independence is not essential here, but will be convenient in the sequel. The assumption (A3) serves as a sufficient condition for this operation of substitution of Y instead of $\tilde Y$ to be well-defined; namely, it guarantees that "$\mathbb{P}_{\mu_0}$-almost surely" implies "$\mathbb{P}_{\nu_0}$-almost surely". Remind that another sufficient reasoning about this substitution is just to use continuity with respect to Y mentioned above; in turn, (A3) will serve another important task later in the proof. The conditional probability $\mathbb{P}_{\mu_0}(X_t \in \cdot \mid Y)$ can also be defined via a non-linear operator $\bar S_t^{Y,\mu_0}$,

$$\mu_0 \bar S_t^{Y,\mu_0}(\cdot) := d_t^{\mu_0}\mu_0 S_t^Y(\cdot).$$

The *main question* here is about the discrepancy of measures,

$$(\mu_0 \bar S_t^{Y,\mu_0} - \nu_0 \bar S_t^{Y,\nu_0})(dx_t).$$

We will be studying the *mean* total variation distance; *non-averaged* bounds are also available, for example, for the sequence $t = 1, 2, \ldots$, via Bienaymé – Chebyshev's inequality and Borel–Cantelli lemmae. Let us reiterate that the setting is based upon the notion of exact solution algorithm, and this algorithm must be presented before any statements. We use a particular representation for this algorithm, and no other presentation is discussed in this paper. In particular, as was mentioned above, this algorithm uses conditional expectations continuous with respect to the trajectory Y. Questions about any use of other versions could be just not accepted. We notice, however, that the final results below, – i.e., the formulae (10) and (11), – relate to the *non-conditional expectations* of the difference of two functions of Y. This means that for any other version of those conditional expectations in both formulae, both results will remain valid. At the same time, the use of continuous versions of more complicated conditional expectations of functions of Y and $\bar Y$ *in the proof* is simply a useful proof trick, nothing more. So, questions about other versions of conditional expectations *in the proof* are not relevant.

Theorem 1 *Under the assumptions (A1) – (A3) above, the following bounds hold true:*

$$\mathbb{E}_{\mu_0}\|\mu_0 \bar S_t^{Y,\mu_0} - \nu_0 \bar S_t^{Y,\nu_0}\|_{TV} \le \left\{ \begin{array}{ll} C_m t^{-m}, & p = 1, \quad \forall m > 0, \\ C\exp(-ct), & p = 0. \end{array} \right. \tag{10}$$

Theorem 2 *Under the assumptions (A′1) – (A′3) above, the following bound holds true: there exist C, $c > 0$ such that*

$$\mathbb{E}_{\mu_0}\|\mu_0 \bar S_t^{Y,\mu_0} - \nu_0 \bar S_t^{Y,\nu_0}\|_{TV} \le C\exp(-ct). \tag{11}$$

Remark 2. Notice that in both Theorems the constants in the bounds can be chosen uniformly over appropriate classes of problems, namely, with uniformly bounded values of the integrals in the assumption on the initial measures, all coefficients, and all other constants in the assumptions. The theorem 2 is a non-divergent version close to the (divergent) result in [21], in particular, because of the "large deviation" type conditions, although the proof is quite different. In this respect notice that the theorem 1 provides a version of exponential convergence rate under considerably less restrictive stability conditions.

Remark 3. If there is no even absolute continuity $\mu_0 << \nu_0$, but there is an absolute continuity at some $k_0 > 0$, one can repeat all considerations below starting from this k_0, that would not change the final conclusion about the convergence rate. However, one should take care about the filter algorithm itself: if the observations which are given do not correspond to the initial measure, the algorithm, generally speaking, may not be able to run at all. One possible solution could be to find or model some other imaginary observations, that suit the wrong initial measure, and run the algorithm until it can work with given observations Y. We do not discuss further details here.

3 Proof of Theorem 1

1. Let us consider the process $Z_t = (X_t, \tilde X_t)$; remind that the components X and $\tilde X$ are independent random processes satisfying the same equation (1) with independent versions of Wiener processes, correspondingly, W and $\tilde W$, however with different initial distributions, $X_0 \sim \mu_0$ and $\tilde X_0 \sim \nu_0$.

First of all, let us notice that all conditional expectations like $\mathbb{E}_{\mu_0,\nu_0}(f(X_t, \tilde X_t) \mid Y, \tilde Y)$, with bounded Borel f, which will be frequently in use throughout the proof, are *Hölder* continuous in $C([0,t];\mathbb{R}^d) \times C([0,t];\mathbb{R}^d)$ with respect to $Y, \tilde Y$, and only these versions are considered in all cases. This continuity follows from the same calculus related to Girsanov exponentials as in the step **3** below, via the formula (8), integration by parts in Girsanov exponentials and due to exponential martingale inequalities under the measure $\mathbb{P}^{\gamma}_{\mu_0,\nu_0}$. It is needless to say that no vicious circle arises here. Notice that another explanation of continuity which appears reasonable is based on the PDE representations of both parts in the ratio in the Bayes – Kallianpur – Striebel formula. However, our point of view is that this way is more involved, see the comment after the formula (26) below.

Denote $n := [t]$, – the integer part of t. Let us introduce some indicators. First of all, in this proof, X stands for the whole continuous trajectory on $[0,t]$, and the same for the trajectories $\tilde X$ and Y. Next, we denote a (non-random) vector of dimension n with coordinates 1 or 0 at every place by δ, and the following events and indicators,

$$D_i := \left\{\max\left(|X_i|, |\tilde X_i|\right) \le R;\ \max\left(\sup_{i\le s\le i+1}|X_s|, \sup_{i\le s\le i+1}|\tilde X_s|\right) < R+1\right\},$$

and

$$1_\delta(X,\tilde X) := \prod_{i=0}^{n-1} \left(1\left(D_i\right)\right)^{\delta_i} \times \left(1 - 1\left(D_i\right)\right)^{1-\delta_i},$$

with a convention $0^0 = 1$. We have,

$$\|\mu_0 \bar S_t^{Y,\mu_0} - \nu_0 \bar S_t^{Y,\nu_0}\|_{TV}$$

$$= 2 \sup_{A\in\mathcal{B}(\mathbb{R}^d)} |\mathbb{E}_{\mu_0,\nu_0}(1(X_t\in A) \mid Y) - \mathbb{E}_{\mu_0,\nu_0}(1(\tilde X_t\in A) \mid \tilde Y)|_{\tilde Y = Y}|$$

$$= 2 \sup_{A\in\mathcal{B}(\mathbb{R}^d)} |\mathbb{E}_{\mu_0,\nu_0}(\textstyle\sum_{\delta\in\Delta} 1_\delta(X,\tilde X)1(X_t\in A) \mid Y)$$

$$-\mathbb{E}_{\mu_0,\nu_0}(\textstyle\sum_{\delta\in\Delta} 1_\delta(X,\tilde X)1(\tilde X_t\in A) \mid \tilde Y)|_{\tilde Y = Y}|$$

$$\le 2 \sup_{A\in\mathcal{B}(\mathbb{R}^d)} \textstyle\sum_{\delta\in\Delta} |\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,\tilde X)1(X_t\in A) \mid Y)$$

$$-\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,\tilde X)1(\tilde X_t\in A) \mid \tilde Y)|_{\tilde Y = Y}|, \tag{12}$$

where Δ is the whole set of possible values of δ.

Next, denote by θ_s the shift operator on the trajectories, that is, $(\theta_{t'}X)_s = X_{t'+s}$, $(\theta_{t'}\tilde X)_s = \tilde X_{t'+s}$, etc. For $0\le t'\le t$, by $1_\delta(\theta_{t'}X,\theta_{t'}\tilde X)$ we will understand the indicator $1_\delta(X',\tilde X')$ with the trajectories

$$X'_s := \begin{cases} 0, & s\le t', \\ X_s, & t'\le s\le t, \\ 0, & s\ge t, \end{cases} \qquad \text{and} \qquad \tilde X'_s := \begin{cases} 0, & s\le t', \\ \tilde X_s, & t'\le s\le t, \\ 0, & s\ge t. \end{cases}$$

Let us define new operators on the spaces of normalized and non-normalized measures on $\mathbb{R}^{2d} = \mathbb{R}^d\times\mathbb{R}^d$, as follows, for $0\le t' < t'' \le t$,

$$(\mu_{t'},\nu_{t'}) S_{t':t''}^{Y;R;\delta}(A\times B)$$

$$= \textstyle\int \mathbb{E}^\gamma_{x_0,\tilde x_0}\Big(1(X_{t''-t'}\in A, \tilde X_{t''-t'}\in B) 1_\delta(\theta_{t'}X,\theta_{t'}\tilde X)$$

$$\times \rho_{0,t''-t'}(X,\theta_{t'}Y)\rho_{0,t''-t'}(\tilde X,\theta_{t'}Y) \mid Y\Big)\,\mu_{t'}(dx_0)\nu_{t'}(d\tilde x_0),$$

and

$$(\mu,\nu)\bar S_t^{Y;R;\delta;\mu_0,\nu_0}(A\times B)$$

$$= d_t^{\mu_0} d_t^{\nu_0} \int \mathbb{E}^\gamma_{x_0,\tilde x_0}\Big(1(X_t\in A,\tilde X_t\in B)1_\delta(X,\tilde X)\rho_{0,t}(X,Y)\rho_{0,t}(\tilde X,Y) \mid Y\Big)\,\mu(dx_0)\nu(d\tilde x_0),$$

and

$$(\mu,\nu)S_t^{Y;R;\delta}(A\times B)$$

$$= \int \mathbb{E}^{\gamma}_{x_0,\tilde{x}_0}\Big(1(X_t\in A, \tilde{X}_t\in B)1_\delta(X,\tilde{X})\rho_{0,t}(X,Y)\rho_{0,t}(\tilde{X},Y) \mid Y\Big)\,\mu(dx_0)\nu(d\tilde{x}_0).$$

Moreover, due to the *desintegration* property of regular conditional distributions (see, e.g., [20], [13]),

$$(\mu,\nu)S_t^{Y;R;\delta}(A\times B) \ = (\mu,\nu)\left(\prod_{i=0}^{n-1} S_{i:i+1}^{Y;R;\delta}\right) S_{n:t}^{Y;R;\delta}(A\times B).$$

Notice that in the latter expression we keep the general notation, but, in fact, the operator, $S_{n:t}^{Y;R;\delta}$ here does not depend neither on δ, nor on R, by definition, – because we do not impose any restriction on the "remainder" interval $[n,t]$.

For every δ, let

$$e_t^{Y;\delta;\mu_0,\nu_0} := (\mu_0,\nu_0)\bar{S}_t^{Y;R;\delta}(\mathbb{R}^{2d}).$$

Then, $\mathbb{P}_{\mu_0,\nu_0}$-almost surely,

$$e_t^{Y;\delta;\mu_0,\nu_0} = \mathbb{E}_{\mu_0,\nu_0}(1_\delta(Z)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}.$$

Notice that by the symmetry in the definition of the operators $\bar{S}_t^{Y;R;\delta}$,

$$e_t^{Y;\delta;\mu_0,\nu_0} = e_t^{Y;\delta;\nu_0,\mu_0}, \quad \mathbb{P}_{\mu_0,\nu_0}\text{-almost surely.}$$

Indeed, since all restrictions on x are the same as on $\tilde{x}$, or, in other words, because $1_\delta(x,\tilde{x}) = 1_\delta(\tilde{x},x)$, we conclude,

$$e_n^{Y;\delta;\mu_0,\nu_0} = (\mu_0,\nu_0)\bar{S}_t^{Y;R;\delta}(\mathbb{R}^{2d})$$

$$= d_t^{\mu_0}d_t^{\nu_0}\int \mathbb{E}^{\gamma}_{x_0,\tilde{x}_0}\Big(1_\delta(X,\tilde{X})\rho_{0,t}(X,Y)\rho_{0,t}(\tilde{X},Y)\mid Y\Big)\,\mu(dx_0)\nu(d\tilde{x}_0)$$

$$= d_t^{\mu_0}d_t^{\nu_0}\int \mathbb{E}^{\gamma}_{x_0,\tilde{x}_0}\Big(1_\delta(\tilde{X},X)\rho_{0,t}(\tilde{X},Y)\rho_{0,t}(X,Y)\mid Y\Big)\,\mu(dx_0)\nu(d\tilde{x}_0)$$

(by change of variables, $X\longleftrightarrow\tilde{X}$ and $x_0\longleftrightarrow\tilde{x}_0$, keeping $X_0=x_0$, $\tilde{X}_0=\tilde{x}_0$)

$$= d_t^{\mu_0}d_t^{\nu_0}\int \mathbb{E}^{\gamma}_{x_0,\tilde{x}_0}\Big(1_\delta(X,\tilde{X})\rho_{0,t}(X,Y)\rho_{0,t}(\tilde{X},Y)\mid Y\Big)\,\nu(dx_0)\mu(d\tilde{x}_0)$$

$$= (\nu_0,\mu_0)\bar{S}_t^{Y;R;\delta}(\mathbb{R}^{2d}) = e_n^{Y;\delta;\nu_0,\mu_0},$$

as required. Next, denote

$$(\mu,\nu)\hat{S}_t^{Y;R;\delta;\mu_0,\nu_0}(A\times B) \ := (e_t^{Y;\delta;\mu_0,\nu_0})^{-1}(\mu,\nu)\bar{S}_t^{Y;R;\delta;\mu_0,\nu_0}(A\times B).$$

The sense of the last notation is that the result of this action is a *normalized* measure restricted to the event $(X,\tilde X)\in\delta$. Hence, we have, with $D\in\mathcal{B}(\mathbb{R}^{2d})$,

$$\|\mu_0\bar S_t^{Y;\mu_0} - \nu_0\bar S_t^{\tilde Y;\nu_0}|_{\tilde Y=Y}\|_{TV}$$

$$\le \sum_{\delta\in\Delta} \|(\mu_0,\nu_0)\bar S_t^{Y;R;\delta;\mu_0,\nu_0} - (\nu_0,\mu_0)\bar S_t^{Y;R;\delta;\mu_0,\nu_0}\|_{TV}$$

$$= 2\sum_{\delta\in\Delta} \sup\nolimits_{D\in\mathcal{B}(\mathbb{R}^{2d})}(e_t^{Y;\delta;\mu_0,\nu_0}(\mu_0,\nu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D) - e_t^{Y;\delta;\nu_0,\mu_0}(\nu_0,\mu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D))$$

$$= 2\sum_{\delta\in\Delta} e_t^{Y;\delta;\mu_0,\nu_0} \sup\nolimits_{D\in\mathcal{B}(\mathbb{R}^{2d})}((\mu_0,\nu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D) - (\nu_0,\mu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D)), \quad (13)$$

$\mathbb{P}_{\mu_0,\nu_0}$-almost surely. Here the first inequality can be explained as follows,

$$\|\mu_0\bar S_n^{Y;\mu_0} - \nu_0\bar S_n^{\tilde Y;\nu_0}\Big|_{\tilde Y=Y}\|_{TV}$$

$$= 2\sup_{A\in\mathcal{B}(\mathbb{R}^d)}\left|\mathbb{P}_{\mu_0}\left(1(X_n\in A)\mid Y\right) - \mathbb{P}_{\nu_0}\left(1(\tilde X_n\in A)\mid \tilde Y\right)\Big|_{\tilde Y=Y}\right|$$

$$= 2\sup_{A\in\mathcal{B}(\mathbb{R}^d)}\left|\left(\mathbb{P}_{\mu_0,\nu_0}\left(1(X_n\in A)\mid Y,\tilde Y\right) - \mathbb{P}_{\mu_0,\nu_0}\left(1(\tilde X_n\in A)\mid Y,\tilde Y\right)\right)\Big|_{\tilde Y=Y}\right|$$

(because (X,Y) does not depend on $\tilde Y$, nor $(\tilde X,\tilde Y)$ on Y)

$$= 2\sup_{A\in\mathcal{B}(\mathbb{R}^d)}\left|\mathbb{P}_{\mu_0,\nu_0}\left(\sum_{\delta\in\Delta} 1_\delta(X,\tilde X)1(X_n\in A)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right.$$

$$\left. - \mathbb{P}_{\mu_0,\nu_0}\left(\sum_{\delta\in\Delta} 1_\delta(X,\tilde X)1(\tilde X_n\in A)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right|$$

$$= 2\sup_{A\in\mathcal{B}(\mathbb{R}^d)}\left|\sum_{\delta\in\Delta} \mathbb{E}_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)1(X_n\in A)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right.$$

$$\left. - \sum_{\delta\in\Delta} \mathbb{E}_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)1(\tilde X_n\in A)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right|$$

(because of linearity)

$$= 2\sup_{A\in\mathcal{B}(\mathbb{R}^d)}\left|\sum_{\delta\in\Delta} \mathbb{E}_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)\left(1(X_n\in A) - 1(\tilde X_n\in A)\right)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right|$$

(again due to linearity)

$$\le 2\sup_{A\in\mathcal{B}(\mathbb{R}^d)}\sum_{\delta\in\Delta}\left|\mathbb{E}_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)\left(1(X_n\in A) - 1(\tilde X_n\in A)\right)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right|$$

$$= 2 \sup_{A\in\mathcal{B}(\mathbb{R}^d)} \sum_{\delta\in\Delta} \Big| \mathbb{E}_{\mu_0,\nu_0} \left(1_\delta(X,\tilde X) 1(X_n\in A) \mid Y, \tilde Y \right) \Big|_{\tilde Y = Y}$$

$$- \mathbb{E}_{\mu_0,\nu_0} \left(1_\delta(X,\tilde X) 1(\tilde X_n\in A) \mid Y, \tilde Y \right) \Big|_{\tilde Y = Y} \Big|$$

$$\le 2 \sup_{D\in\mathcal{B}(\mathbb{R}^{2d})} \sum_{\delta\in\Delta} \Big| \mathbb{E}_{\mu_0,\nu_0} \left(1_\delta(X,\tilde X) 1((X_n,\tilde X_n)\in D) \mid Y, \tilde Y \right) \Big|_{\tilde Y = Y}$$

$$- \mathbb{E}_{\mu_0,\nu_0} \left(1_\delta(\tilde X, X) 1((\tilde X_n, X_n)\in D) \mid Y, \tilde Y \right) \Big|_{\tilde Y = Y} \Big| .$$

In the last line we used symmetry $1_\delta(\tilde X_n, X_n) \equiv 1_\delta(X_n, \tilde X_n)$. Remind that in all terms we use versions of conditional expectations continuous with respect to trajectories Y and $\tilde Y$.

To estimate the terms $\sup_D((\mu_0,\nu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D) - (\nu_0,\mu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D))$ in (13), we will use the Birkhoff metric for positive measures, – other names for this metric are Hilbert and projective, see [10], and also [1], [17], –

$$h(\mu,\nu) = \begin{cases} \ln \dfrac{(\inf u:\ u>0,\ \mu\le u\nu)}{(\sup v:\ v>0,\ \mu\ge v\nu)}, & \text{if finite,} \\ +\infty, & \text{otherwise.} \end{cases}$$

In other words,

$$h(\mu,\nu) = \ln \|\frac{d\mu}{d\nu}\| + \ln \|\frac{d\nu}{d\mu}\|,$$

where $\|\frac{d\mu}{d\nu}\| = \operatorname{ess\,sup} |\frac{d\mu}{d\nu}|$. For normalized measures, we have the following basic inequality (see [1] or [17]),

$$2 \sup_D ((\mu_0,\nu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D) - (\nu_0,\mu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}(D))$$

$$\le \frac{2}{\ln 3}\, h((\mu_0,\nu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}, (\nu_0,\mu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}).$$

It was duly noticed by Christoph Leuridan [private communication, 2006] that the constant $\frac{2}{\ln 3}$ in this inequality may be replaced by one; for our aim, any constant would do.

Moreover, for any $0\le i\le n-1$,

$$h\left((\mu_i,\nu_i) S_{i:i+1}^{Y;R;\delta}, (\nu_i,\mu_i) S_{i:i+1}^{Y;R;\delta} \right) \le h\left((\mu_i,\nu_i), (\nu_i,\mu_i) \right), \qquad (14)$$

as well as

$$h\left((\mu_n,\nu_n) S_{n:t}^{Y;R;\delta}, (\nu_n,\mu_n) S_{n:t}^{Y;R;\delta} \right) \le h\left((\mu_n,\nu_n), (\nu_n,\mu_n) \right), \qquad (15)$$

and for any $L > 0$ there is a constant $\pi_R(L) < 1$, such that for every $i \in J := \{j : \delta_j = \delta_{j+1} = 1\}$,

$$h\left((\mu_i, \nu_i) S^{Y;R;\delta}_{i:i+1}, (\nu_i, \mu_i) S^{Y;R;\delta}_{i:i+1}\right) \le \pi_R(L)\, h\left((\mu_i, \nu_i), (\nu_i, \mu_i)\right), \tag{16}$$

if the trajectory Y on $i \le t \le i+1$ satisfies the inequality, $\Delta_i(Y) := \sup_{i \le s \le i+1} |Y_s - Y_i| \le L$. Emphasize that this boundedness condition for Y can be checked with a probability close to one, if L is large enough. The idea of this contraction which occurs or does not occur randomly, depending on Y, is that it can be satisfied *frequently* as i runs its values, $i = 0, 1, \ldots, n$, and this suffices for a final exponential or polynomial bound, depending on $p = 0$ or $p = 1$. If such a contraction does not occur frequently enough, – see below for precise expressions, – we will not pass to the Birkhoff metric, but will estimate the probability of this unlikely event. The non-strict bounds (14) and (15) are due to the fact that any kernel, conditional or not, does not increase the Birkhoff distance between two measures.

2. At this stage of the proof we will show how the Harnack inequality implies (16). Denote by Γ the parabolic boundary of the cylinder $\{(t, x, \tilde x) : 0 \le t \le 1; |x| \le R + 1, |\tilde x| \le R + 1\}$. Then the version of the parabolic Harnack inequality from [12] applied to the pair $(X, \tilde X)$, which is a diffusion process of dimension $2d$, may be formulated in the following way: for every non-negative function φ on Γ such that $\varphi(t, x, \tilde x) \equiv 0$, $\forall\, 0 \le t \le 1/4$, and $\mathbb{E}_{0,0} 1(D_0) \varphi(\tau, X_\tau, \tilde X_\tau) > 0$, we have,

$$\sup_{|x_0(1)|, |\tilde x_0(1)|, |x_0(2)|, |\tilde x_0(2)| \le R} \frac{\mathbb{E}_{x_0(1), \tilde x_0(1)}\, 1(D_0)\, \varphi(\tau, X_\tau, \tilde X_\tau)}{\mathbb{E}_{x_0(2), \tilde x_0(2)}\, 1(D_0)\, \varphi(\tau, X_\tau, \tilde X_\tau)} \le C, \tag{17}$$

where $\tau = \inf(t : 0 \le t \le 1, |X_t| \ge R + 1)$, with the convention $\inf(\varnothing) = 1$.

The inequality (17) implies (16). Indeed, as $|x_0(1)| \vee |\tilde x_0(1)| \vee |x_0(2)| \vee |\tilde x_0(2)| \le R$, the inequality (17) implies,

$$\mathbb{P}_{x_0(1), \tilde x_0(1)}(D_0 \cap \{(X_1, \tilde X_1) \in D\}) \le C\, \mathbb{P}_{x_0(2), \tilde x_0(2)}(D_0 \cap \{(X_1, \tilde X_1) \in D\}). \tag{18}$$

Now we will establish a similar inequality for the conditional linearized dynamics, under the assumption $\Delta_0(Y) = \sup_{0 \le s \le 1} |Y_s - Y_0| \le L$, and with some new constant C, in particular, depending on L and R.

In the Girsanov exponentials here we perform integration by parts (since $h \in C^2_b$), or Itô's formula, to get the following, – we show the term $\rho_{0,1}(X, Y)$ only, and drop the part $\exp(-\frac{1}{2} \int_0^1 h^2(X_s)\, ds)$, because it is bounded and bounded away from zero on D_0,

$$\exp\Big(\int_0^1 h(X_s)\, dY_s\Big) = \exp\Big(h(X_1)^*(Y_1 - Y_0) - \int_0^1 (Y_s - Y_0)\, dh(X_s)\Big)$$

$$= \exp\Big(h(X_1)^*(Y_1 - Y_0) - \int_0^1 (Y_s - Y_0)^* \nabla h(X_s)\, (dW_s + b(X_s)\, ds)\Big)$$

$$\times \exp(-\frac{1}{2}\int_0^1 (Y_s - Y_0)\Delta h(X_s)\, ds)$$

$$\equiv \exp(-\int_0^1 (Y_s - Y_0)^* \nabla h(X_s)\, dW_s - \frac{1}{2}\int_0^1 \|(Y_s - Y_0)^* \nabla h(X_s)\|^2\, ds)$$

$$\times \exp(h(X_1)(Y_1 - Y_0) - \frac{1}{2}\int_0^1 (Y_s - Y_0)^* \Delta h(X_s)\, ds)$$

$$\times \exp(\int_0^1 \left\{\frac{1}{2}\|(Y_s - Y_0)^* \nabla h(X_s)\|^2 - (Y_s - Y_0)^* \nabla h(X_s) b(X_s)\right\} ds).$$

Here everywhere Δh is understood as a vector with components Δh^i, where $h = (h^i,\, 1 \le i \le \ell)$; ∇h is a matrix $\ell \times d$; and all expressions like Yh mean scalar products, $\langle Y, h\rangle$; in some more complicated cases we also use transposition signs, like $Y^* \nabla h\, dW_s$.

The second exponential with $\tilde X$, can be similarly transformed to

$$\exp(\int_0^1 h(\tilde X_s)\, dY_s) = \exp(h(\tilde X_1)(Y_1 - Y_0) - \int_0^1 (Y_s - Y_0)^*\, dh(\tilde X_s))$$

$$= \exp(-\int_0^1 (Y_s - Y_0)^* \nabla h(\tilde X_s)\, d\tilde W_s - \int_0^1 \|(Y_s - Y_0)^* \nabla h(\tilde X_s)\|^2\, ds)$$

$$\times \exp(h(\tilde X_1)(Y_1 - Y_0) - \int_0^1 (Y_s - Y_0)\Delta h(\tilde X_s)\, ds)$$

$$\times \exp(\int_0^1 \left\{\frac{1}{2}\|(Y_s - Y_0)^* \nabla h(\tilde X_s)\|^2 - (Y_s - Y_0)^* \nabla h(\tilde X_s) b(\tilde X_s)\right\} ds).$$

We will treat here the term

$$\hat\gamma := \exp(-\int_0^1 (Y_s - Y_0)^* \nabla h(X_s)\, dW_s - \int_0^1 (Y_s - Y_0)^* \nabla h(\tilde X_s)\, d\tilde W_s)$$

$$\times \exp(-\frac{1}{2}\int_0^1 \|(Y_s - Y_0)^* \nabla h(X_s)\|^2\, ds - \frac{1}{2}\int_0^1 \|(Y_s - Y_0)^* \nabla h(\tilde X_s)\|^2\, ds)$$

as a new Girsanov exponential which changes the drift $b(X_s)$ of our diffusion X by a new one, $b(X_s) - (Y_s - Y_0)^* \nabla h(X_s)$, and, similarly, the drift $b(\tilde X_s)$ of the component $\tilde X$ is to be changed to $b(\tilde X_s) - (Y_s - Y_0)^* \nabla h(\tilde X_s)$. In the other words, by changing measure, we transform our process $(X, Y, \tilde X, \tilde Y)$ on $[0, 1]$, which satisfies the system of equations under the original measure $\mathbb{P}$,

$$\begin{cases} dX_s = b(X_s)\, ds + dW_s, \\ dY_s = h(X_s)\, ds + dB_s, \\ d\tilde X_s = b(\tilde X_s)\, ds + d\tilde W_s, \\ d\tilde Y_s = h(\tilde X_s)\, ds + d\tilde B_s, \end{cases} \tag{19}$$

with initial data $X_0, Y_0, \tilde{X}_0, \tilde{Y}_0$, to the following one with respect to $\hat{\mathbb{P}}$, on the same interval $[0,1]$,

$$
\left\{
\begin{array}{l}
dX_s = (b(X_s) - (Y_s - Y_0)^* \nabla h(X_s))\, ds + dW_s, \\
dY_s = dB_s, \\
d\tilde{X}_s = (b(\tilde{X}_s) - (Y_s - Y_0)^* \nabla h(\tilde{X}_s))\, ds + d\tilde{W}_s, \\
d\tilde{Y}_s = d\tilde{B}_s,
\end{array}
\right.
\tag{20}
$$

and with the same initial data. Of course, all Wiener processes in (20) are new. The measure $\hat{\mathbb{P}}$ reads,

$$
\frac{d\hat{\mathbb{P}}}{d\mathbb{P}^\gamma}(\omega) := \hat{\gamma}(\omega).
$$

Notice that this operation, of course, does not change trajectories of all processes $X, \tilde{X}, Y, \tilde{Y}$. Due to the Bayes – Kallianpur – Striebel formula (8) again, we have,

$$
\begin{aligned}
&\mathbb{E}^\gamma_{x,\tilde{x}} \Big(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\}) \rho_{0,1}(X,Y) \rho_{0,1}(\tilde{X},\tilde{Y}) \mid Y, \tilde{Y}\Big) \\
&= \frac{\hat{\mathbb{E}}_{x,\tilde{x}} \Big(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\}) \rho_{0,1}(X,Y) \rho_{0,1}(\tilde{X},\tilde{Y}) \hat{\gamma}^{-1}(X,\tilde{X},Y) \mid Y, \tilde{Y}\Big)}{\hat{\mathbb{E}}_{x,\tilde{x}} \Big(\hat{\gamma}^{-1}(X,\tilde{X},Y) \mid Y, \tilde{Y}\Big)}.
\end{aligned}
\tag{21}
$$

Correspondingly,

$$
\begin{aligned}
&\mathbb{E}^\gamma_{\tilde{x},x} \Big(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\}) \rho_{0,1}(X,Y) \rho_{0,1}(\tilde{X},\tilde{Y}) \mid Y, \tilde{Y}\Big) \\
&= \frac{\hat{\mathbb{E}}_{\tilde{x},x} \Big(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\}) \rho_{0,1}(X,Y) \rho_{0,1}(\tilde{X},\tilde{Y}) \hat{\gamma}^{-1}(X,\tilde{X},Y) \mid Y, \tilde{Y}\Big)}{\hat{\mathbb{E}}_{\tilde{x},x} \Big(\hat{\gamma}^{-1}(X,\tilde{X},Y) \mid Y, \tilde{Y}\Big)}.
\end{aligned}
\tag{22}
$$

Let us consider the enumerator here. The "main part", – that is, the part with stochastic integrals, – of the expression

$$
\rho_{0,1}(X,Y) \rho_{0,1}(\tilde{X},\tilde{Y}) \hat{\gamma}^{-1}(X,\tilde{X},Y)
$$

is the following exponential,

$$
\begin{aligned}
&\exp\Big(-\textstyle\int_0^1 (Y_s - Y_0)^* \nabla h(X_s)\, dW_s - \int_0^1 (Y_s - Y_0)^* \nabla h(\tilde{X}_s)\, d\tilde{W}_s\Big) \times \\
&\times \exp\Big(+\textstyle\int_0^1 (Y_s - Y_0)^* \nabla h(X_s)\, dW_s + \int_0^1 (Y_s - Y_0)^* \nabla h(\tilde{X}_s)\, d\tilde{W}_s\Big) \equiv 1.
\end{aligned}
$$

In the other words, the stochastic part here vanishes, providing a multiple which equals one. The rest is bounded on the set $\{\sup_{0 \le s \le 1} |Y_s - Y_0| \le L\}$ (remind that we are working with the system (20) under the measure $\hat{\mathbb{P}}$, so that $\tilde{Y}$ does not show up here):

$$\exp(h(X_1)(Y_1 - Y_0) - \frac{1}{2}\int_0^1 (Y_s - Y_0)\Delta h(X_s)\,ds - \frac{1}{2}\int_0^1 \|(Y_s - Y_0)^*\nabla h(X_s)\|^2\,ds)$$

$$\times \exp(\int_0^1 \left\{\frac{1}{2}\|(Y_s - Y_0)^*\nabla h(X_s)\|^2 - (Y_s - Y_0)^*\nabla h(X_s)b(X_s)\right\} ds)$$

$$\times \exp(h(\tilde{X}_1)(Y_1 - Y_0) - \frac{1}{2}\int_0^1 (Y_s - Y_0)\Delta h(\tilde{X}_s)\,ds - \frac{1}{2}\int_0^1 \|(Y_s - Y_0)^*\nabla h(\tilde{X}_s)\|^2\,ds)$$

$$\times \exp(\int_0^1 \left\{\frac{1}{2}\|(Y_s - Y_0)^*\nabla h(\tilde{X}_s)\|^2 - (Y_s - Y_0)^*\nabla h(\tilde{X}_s)b(\tilde{X}_s)\right\} ds) \le C(L,R) < \infty,$$

due to the assumptions; we also have a similar lower bound for the left hand side, $\ge C(L,R)^{-1}$, which is not used here but will be useful in the sequel. Hence, we have the following bound,

$$\hat{\mathbb{E}}_{x,\tilde{x}}\left(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\})\rho_{0,1}(X,Y)\rho_{0,1}(\tilde{X},\tilde{Y})\hat{\gamma}^{-1}(X,\tilde{X},Y) \mid Y,\tilde{Y}\right)$$

$$\le C(L,R)\hat{\mathbb{E}}_{x,\tilde{x}}\left(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\}) \mid Y,\tilde{Y}\right)$$

$$= C(L,R)\hat{\mathbb{E}}_{x,\tilde{x}}\left(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\}) \mid Y\right),$$

the last step because $\tilde{Y}$ is independent of $X, \tilde{X}, Y$ under $\hat{\mathbb{P}}$, due to the system (20).

Here is it reasonable to modify slightly the definition of D_0. Indeed, the first two requirements, $|X_0| \le R$ and $|\tilde{X}_0| \le R$ are not important, and can be dropped without any harm to the meaning, because they are restrictions on x and $\tilde{x}$. Now we will allow any values $x, \tilde{x} \in \mathbb{R}^{2d}$, but will use the conclusions related only to $x, \tilde{x}$ separated from the boundary, as this is needed for Harnack's inequality. Hence, let

$$D_0' := \left\{\max\left(\sup_{0\le s\le 1}|X_s|, \sup_{0\le s\le 1}|\tilde{X}_s|\right) < R+1\right\}.$$

Then, for $|x|$, $|\tilde{x}| \le R$, we have an identity,

$$\hat{\mathbb{E}}_{x,\tilde{x}}\left(1(D_0 \cap \{(X_1, \tilde{X}_1) \in D\}) \mid Y\right) = \hat{\mathbb{E}}_{x,\tilde{x}}\left(1(D_0' \cap \{(X_1, \tilde{X}_1) \in D\}) \mid Y\right).$$

We are going to show the inequality for regular conditional measures,

$$\begin{aligned} &\hat{\mathbb{E}}_{x,\tilde{x}}\left(1(D_0' \cap \{(X_1, \tilde{X}_1) \in D\}) \mid Y\right) \\ &\le C(L,R)\hat{\mathbb{E}}_{x',\tilde{x}'}\left(1(D_0' \cap \{(X_1, \tilde{X}_1) \in D\}) \mid Y\right), \end{aligned} \tag{23}$$

on the set $\{\sup_{0\le s\le 1}|Y_s - Y_0| \le L\}$. To this aim, it suffices to show a similar inequality with any non-negative continuous bounded function φ, or even for any non-negative $\varphi \in C_b^2$,

$$\begin{aligned} &\hat{\mathbb{E}}_{x,\tilde x}\left(1(D_0')\varphi(X_1,\tilde X_1) \mid Y\right) \\ &\le C(L,R)\hat{\mathbb{E}}_{x',\tilde x'}\left(1(D_0')\varphi\,(X_1,\tilde X_1) \mid Y\right). \end{aligned} \tag{24}$$

The expression $\hat{\mathbb{E}}_{x,\tilde x}\left(1(D_0')\varphi(X_1,\tilde X_1)) \mid Y\right)$ can be treated as a solution $u^\psi(s,x,\tilde x)$ at $s=0$ of a linear parabolic PDE with a non-homogeneous first order term determined by the random process Y,

$$\begin{aligned} &u_s + \frac{1}{2}u_{xx} + \frac{1}{2}u_{\tilde x\tilde x} + (b(x) - (\psi_s - \psi_0)^*\nabla h(x))u_x \\ &+(b(\tilde x) - (\psi_s - \psi_0)^*\nabla h(\tilde x))u_{\tilde x} = 0, \qquad \max(|x|,|\tilde x|) < R+1, \;\; 0\le s<1, \\ &u(1,x,\tilde x) = \varphi(x,\tilde x), \quad \& \quad u(s,x,\tilde x)=0, \;\; 0<s<1, \; \max(|x|,|\tilde x| = R+1), \end{aligned} \tag{25}$$

with a replacement $\psi = Y$; hence, we can use the notation u^Y. In all cases the curve $(\psi_s,\, 0\le s\le 1)$ will be considered continuous; in particular, it is bounded. For every $\varphi \in C_b^2$, there is a unique solution of the equation (25) in the Sobolev class of functions (see [16]); for $\varphi \in C_b$ only this is also correct, see [23]). Notice that discontinuous initial functions φ could be used, too, however, PDE references are simpler with φ smooth. Let us justify the equality,

$$\hat{\mathbb{E}}_{0,x,\tilde x}\left(1(D_0')\varphi(X_1,\tilde X_1) \mid Y\right) = u^Y(0,x,\tilde x). \tag{26}$$

Here the right hand side in (26) is just a solution of the equation (25) with a substitution $\psi = Y$.

The almost sure equality (26) is intuitively evident, given that Y is solved independently of X and $\tilde X$, and that Y and $(X,\tilde X)$ are independent under $\mathbb{P}^\zeta$. However, for the completeness of presentation we prefer to give an independent justification below. Notice that the most natural way, – to fix Y, use Itô – Krylov's formula and take conditional expectation, – requires some technical measurability lemmae for solutions of SDEs which depend on the trajectory Y as a parameter, as well as SDE uniqueness theorems *in the sense of ODEs* which we decided to avoid. By the way, if we used the original measure, we would not get a linear PDE as above; this shows that despite the intuitive evidence, some precautions should be made at this point.

Notice that due to the a priori bounds for solutions of PDEs [16], the function u^ψ depends on the trajectory ψ continuously, and this will be used directly. Let us take any bounded continuous function $g(\cdot)$ defined on $C[0,1]$, and show that

$$\hat{\mathbb{E}}_{0,x,\tilde{x}} g(Y) \hat{\mathbb{E}}_{0,x,\tilde{x}} \left(1(D_0') \varphi(X_1, \tilde{X}_1) \mid Y\right) = \hat{\mathbb{E}}_{0,x,\tilde{x}} g(Y) u^Y(0,x,\tilde{x}), \qquad (27)$$

which is another form of the equality (26). One more version is,

$$\hat{\mathbb{E}}_{0,x,\tilde{x}} g(Y) 1(D_0') \varphi(X_1, \tilde{X}_1) = \hat{\mathbb{E}}_{0,x,\tilde{x}} g(Y) u^Y(0,x,\tilde{x}), \qquad (28)$$

Consider the exponential,

$$\zeta = \zeta(Y) := \exp\left(\int_0^1 (Y_s - Y_0)^* \nabla h(X_s) dW_s + \int_0^1 (Y_s - Y_0)^* \nabla h(\tilde{X}_s) d\tilde{W}_s \right.$$
$$\left. - \frac{1}{2} \int_0^1 |(Y_s - Y_0)^* \nabla h(X_s)|^2 + |(Y_s - Y_0)^* \nabla h(\tilde{X}_s)|^2) ds \right),$$

and the measure

$$d\mathbb{P}^\zeta / d\hat{\mathbb{P}}(\omega) := \zeta(Y).$$

Then, (28) may be rewritten as

$$\mathbb{E}^\zeta_{0,x,\tilde{x}} \zeta^{-1}(Y) g(Y) 1(D_0') \varphi(X_1, \tilde{X}_1) = \mathbb{E}^\zeta_{0,x,\tilde{x}} \zeta^{-1}(Y) g(Y) u^Y(0,x,\tilde{x}), \qquad (29)$$

Under $\mathbb{P}^\zeta$, the system becomes

$$\begin{cases} dX_s = b(X_s)\, ds + dW_s, \quad X_0 = x, \\ dY_s = dB_s, \\ d\tilde{X}_s = b(\tilde{X}_s)\, ds + d\tilde{W}_s, \\ d\tilde{Y}_s = d\tilde{B}_s, \end{cases}$$

as usual, with new Wiener processes.

3. Next argument is to use an epsilon net in order to approximate Y, the latter being a Wiener process under both $\hat{\mathbb{P}}$ and $\mathbb{P}^\zeta$. Due to the properties of a d-dimensional Wiener process, and Ulam's Theorem on measure tightness, given any $\epsilon' > 0$, we can find a compact $K \subset C[0,1]$, such that

$$\mathbb{P}^\zeta(Y - Y_0 \in K) > 1 - \epsilon'. \qquad (30)$$

Now, with any $\epsilon > 0$, let us choose some ϵ-net, $\{\psi^i,\, 1 \le i \le N\}$ for this compact K. Denote by induction,

$$U^1 := \{\psi : \|\psi - \psi^1\|_C < \epsilon\};$$
$$U^2 := \{\psi : \|\psi - \psi^2\|_C < \epsilon\} \setminus U^1;$$
$$\dots$$
$$U^{k+1} := \{\psi : \|\psi - \psi^{k+1}\|_C < \epsilon\} \setminus U^k;$$
$$\dots$$

Then, since Y and $(X,\tilde X)$ are independent under $\mathbb{P}^\zeta$, we have,

$$
\begin{aligned}
&\mathbb{E}^\zeta_{0,x,\tilde x}\zeta^{-1}(Y)g(Y)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})\\
&=\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)\zeta^{-1}(Y)g(Y)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})\\
&+\mathbb{E}^\zeta_{0,x,\tilde x}1(Y\not\in\bigcup_i U^i)\zeta^{-1}(Y)g(Y)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})\\
&\approx\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)\zeta^{-1}(\psi^i)g(\psi^i)1(D'_0\cap\{(X_1,\tilde X_1)\in D\}),
\end{aligned}
$$

the latter approximative equality is claimed as $\epsilon\to 0$. Indeed, we estimate,

$$
\begin{aligned}
&|\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)\zeta^{-1}(Y)g(Y)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})\\
&+\mathbb{E}^\zeta_{0,x,\tilde x}1(Y\not\in\bigcup_i U^i)\zeta^{-1}(Y)g(Y)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})\\
&-\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)\zeta^{-1}(\psi^i)g(\psi^i)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})\,|\\
&\le \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\not\in\bigcup_i U^i)\zeta^{-1}(Y)g(Y)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})\\
&+\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)|\zeta^{-1}(Y)g(Y)-\zeta^{-1}(\psi^i)g(\psi^i)|1(D'_0\cap\{(X_1,\tilde X_1)\in D\}).
\end{aligned}
$$

Here the term

$$
\mathbb{E}^\zeta_{0,x,\tilde x}1(Y\not\in\bigcup_i U^i)\zeta^{-1}(Y)g(Y)1(D'_0\cap\{(X_1,\tilde X_1)\in D\})
$$

is small due to (30). Moreover, since g is bounded and continuous in $C[0,1]$, it suffices to estimate, for example, the expression,

$$
\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)|\zeta^{-1}(Y)-\zeta^{-1}(\psi^i)|1(D'_0\cap\{(X_1,\tilde X_1)\in D\}),
$$

and, more than that, the last indicator may be dropped here while estimating from above. We have,

$$
\begin{aligned}
&\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)|\zeta^{-1}(Y)-\zeta^{-1}(\psi^i)| = \sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)\zeta^{-1}(Y)|1-\zeta^{-1}(\psi^i)\,\zeta(Y)|\\
&=\sum_i \mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)\zeta^{-1}(Y)\mathbb{E}^\zeta_{0,x,\tilde x}\left(\left|1-\exp\left(\int_0^1(Y_s-Y_0)^*\nabla h(X_s)dW_s\right.\right.\right.\\
&+\int_0^1(Y_s-Y_0)^*\nabla h(\tilde X_s)d\tilde W_s-\int_0^1(\psi^i_s-\psi^i_0)^*\nabla h(X_s)dW_s-\int_0^1(\psi^i_s-\psi^i_0)^*\nabla h(\tilde X_s)d\tilde W_s\\
&-\frac12\int_0^1(|(Y_s-Y_0)^*\nabla h(X_s)|^2-|(\psi^i_s-\psi^i_0)^*\nabla h(X_s)|^2\\
&\left.\left.\left.+|(Y_s-Y_0)\nabla h(\tilde X_s)|^2-|(\psi^i_s-\psi^i_0)^*\nabla h(\tilde X_s)|^2)\,ds\right)\right|\;|\;Y\right).
\end{aligned}
$$

Here, using the fact that under $\mathbb{P}^\zeta$, the processes Y and $(X, \tilde{X}, W, \tilde{W})$ are *independent*, we can integrate with respect to W and $\tilde{W}$, treating Y as a non-random trajectory. Let us apply the Cauchy – Bouniakovsky – Schwarz (CBS) inequality for conditional expectations, and then estimate this conditional expectation *squared*:

$$\Bigg(\mathbb{E}^\zeta_{0,x,\tilde{x}}\Bigg(\Bigg|1-\exp\Bigg(\int_0^1 (Y_s-Y_0)^*\nabla h(X_s)dW_s$$

$$+\int_0^1 (Y_s-Y_0)^*\nabla h(\tilde{X}_s)d\tilde{W}_s-\int_0^1 (\psi^i_s-\psi^i_0)^*\nabla h(X_s)dW_s-\int_0^1 (\psi^i_s-\psi^i_0)^*\nabla h(\tilde{X}_s)d\tilde{W}_s$$

$$-\frac{1}{2}\int_0^1 (\ |(Y_s-Y_0)^*\nabla h(X_s)|^2-|(\psi^i_s-\psi^i_0)^*\nabla h(X_s)|^2$$

$$+|(Y_s-Y_0)\nabla h(\tilde{X}_s)|^2-|(\psi^i_s-\psi^i_0)^*\nabla h(\tilde{X}_s)|^2)\,ds\Bigg)\Bigg|\ \mid Y\Bigg)^2$$

$$\le \mathbb{E}^\zeta_{0,x,\tilde{x}}\Bigg(\Bigg|1-\exp\Bigg(\int_0^1 (Y_s-Y_0)^*\nabla h(X_s)dW_s+\int_0^1 (Y_s-Y_0)^*\nabla h(\tilde{X}_s)d\tilde{W}_s$$

$$-\int_0^1 (\psi^i_s-\psi^i_0)^*\nabla h(X_s)dW_s-\int_0^1 (\psi^i_s-\psi^i_0)^*\nabla h(\tilde{X}_s)\,d\tilde{W}_s$$

$$-\frac{1}{2}\int_0^1 (\ |(Y_s-Y_0)^*\nabla h(X_s)|^2-|(\psi^i_s-\psi^i_0)^*\nabla h(X_s)|^2$$

$$+|(Y_s-Y_0)\nabla h(\tilde{X}_s)|^2-|(\psi^i_s-\psi^i_0)^*\nabla h(\tilde{X}_s)|^2)\,ds\Bigg)\Bigg|^2\ \mid Y\Bigg).$$

$$=1+\mathbb{E}^\zeta_{0,x,\tilde{x}}\Bigg(\exp\Bigg(\int_0^1 2(Y_s-Y_0)^*\nabla h(X_s)dW_s+\int_0^1 2(Y_s-Y_0)^*\nabla h(\tilde{X}_s)d\tilde{W}_s$$

$$-\int_0^1 2(\psi^i_s-\psi^i_0)^*\nabla h(X_s)dW_s-\int_0^1 2(\psi^i_s-\psi^i_0)^*\nabla h(\tilde{X}_s)d\tilde{W}_s$$

$$-\int_0^1 (\ |(Y_s-Y_0)^*\nabla h(X_s)|^2-|(\psi^i_s-\psi^i_0)^*\nabla h(X_s)|^2$$

$$+|(Y_s - Y_0)\nabla h(\tilde X_s)|^2 - |(\psi^i_s - \psi^i_0)^*\nabla h(\tilde X_s)|^2)\, ds \Bigg) \mid Y\Bigg)$$

$$-2\mathbb{E}^{\zeta}_{0,x,\tilde x} \exp\Bigg(\int_0^1 (Y_s - Y_0)^*\nabla h(X_s)dW_s + \int_0^1 (Y_s - Y_0)^*\nabla h(\tilde X_s)d\tilde W_s$$

$$- \int_0^1 (\psi^i_s - \psi^i_0)^*\nabla h(X_s)dW_s - \int_0^1 (\psi^i_s - \psi^i_0)^*\nabla h(\tilde X_s)d\tilde W_s$$

$$-\frac{1}{2}\int_0^1 \big(\ |(Y_s - Y_0)^*\nabla h(X_s)|^2 - |(\psi^i_s - \psi^i_0)^*\nabla h(X_s)|^2$$

$$+|(Y_s - Y_0)\nabla h(\tilde X_s)|^2 - |(\psi^i_s - \psi^i_0)^*\nabla h(\tilde X_s)|^2)\, ds \Bigg) \mid Y\Bigg).$$

Clearly, on the set $Y \in U^i$, both expectations in the latter expression are uniformly close to one, so that the whole sum is of the order $o_\epsilon(1)$. Thus, the whole expression is uniformly close to zero. More precisely, by using exponential martingale properties, we easily get,

$$o_\epsilon(1) \ \text{ is of the order } \epsilon. \tag{31}$$

This shows *Hölder* continuity of the order 1/2 of each term of the sum

$$\sum_i \mathbb{E}^{\zeta}_{0,x,\tilde x} 1(Y \in U^i)\zeta^{-1}(\psi^i)g(\psi^i)1(D'_0 \cap \{(X_1, \tilde X_1) \in D\}) \mid \tag{32}$$

with respect to ψ. We notice that the same arguments provide us Hölder continuity of conditional expectations mentioned in the beginning of the proof. Let us show (31). Indeed, for generic adapted processes y_s, z_s, $0 \le s \le 1$, under a condition $\sup_s(|y_s| + |z_s|) \le \epsilon < \infty$, with $\epsilon > 0$ small, we have, with a d-dimensional Wiener process W,

$$\left(\mathbb{E}\left(\exp\left(\int_0^1 z^*_s dW_s - \int_0^1 y_s\, ds\right)\right)\right)^{1/2}$$

$$= \left(\mathbb{E}\left(\exp\left(\int_0^1 z^*_s dW_s \mp \int_0^1 |z_s|^2\, ds\right) - \int_0^1 y_s\, ds\right)\right)^{1/2}$$

$$\le \left(\mathbb{E}\exp\left(\int_0^1 2z^*_s dW_s - 2\int_0^1 |z_s|^2\, ds\right)\right)^{1/4} \left(\mathbb{E}\exp\left(\int_0^1 2(|z_s|^2 - y_s)\, ds\right)\right)^{1/4}$$

$$\le \exp(C\epsilon) \sim 1 + C\epsilon.$$

Similarly, due to the version of the same inequality that could be called *CBS inequality from below*, $\mathbb{E}\xi\eta \geq \mathbb{E}\xi^{1/2}(\mathbb{E}\eta^{-1})^{-1/2}$ for $\xi, \eta > 0$, there is a lower bound,

$$\mathbb{E}\left(\exp\left(\int_0^1 z_s^* dW_s - \int_0^1 y_s\, ds\right)\right)$$

$$= \mathbb{E}\left(\exp\left(-\int_0^1 z_s^* dW_s \pm \frac{1}{2}\int_0^1 |z_s|^2\, ds\right) - \int_0^1 y_s\, ds\right)$$

$$\geq \left(\mathbb{E}\exp\left(\int_0^1 z_s^* dW_s - \frac{1}{2}\int_0^1 |z_s|^2\, ds\right)\right)^{-1/2} \left(\mathbb{E}\exp\left(\int_0^1 (-\frac{1}{4}|z_s|^2 - \frac{1}{2}y_s)\, ds\right)\right)$$

$$\geq \exp(-C\epsilon) \sim 1 - C\epsilon.$$

Here in both cases $\sim$ means *equivalence after we subtract* 1; we have used

$$\xi = \exp\left(-\frac{1}{2}\int_0^1 |z_s|^2\, ds - \int_0^1 y_s\, ds\right), \text{ and } \eta = \exp\left(\int_0^1 z_s^* dW_s + \frac{1}{2}\int_0^1 |z_s|^2\, ds\right).$$

In other words, the difference between $\mathbb{E}\exp\left(\int_0^1 z_s^* dW_s - \int_0^1 y_s\, ds\right)$ and 1 is of the order ϵ, at most. Hence, the value

$$\mathbb{E}^{\zeta}_{0,x,\tilde{x}}\left(\left|1 - \exp\left(\int_0^1 (Y_s - Y_0)^* \nabla h(X_s) dW_s\right.\right.\right.$$

$$+ \int_0^1 (Y_s - Y_0)^* \nabla h(\tilde{X}_s) d\tilde{W}_s - \int_0^1 (\psi^i_s - \psi^i_0)^* \nabla h(X_s) dW_s - \int_0^1 (\psi^i_s - \psi^i_0)^* \nabla h(\tilde{X}_s) d\tilde{W}_s$$

$$\left.\left.\left. -\frac{1}{2}\int_0^1 ((Y_s - Y_0) - (\psi^i_s - \psi^i_0))^* \nabla h(X_s)|^2 + |((Y_s - Y_0) - (\psi^i_s - \psi^i_0))^* \nabla h(\tilde{X}_s)|^2) ds\right)\right| \mid Y\right)$$

is, at most, of the order ϵ, which shows (31). Notice that using Hölder's inequality instead of CBS, it is possible to improve it up to any order less than one; but for our goal any positive order suffices.

To establish continuity of the right hand side of (29), we will use an approach based on PDE estimates, also mentioned in the beginning of the proof.

By virtue of [16, Theorem 4.9.1], for every ψ with $\|\psi\|_C \leq K$, there exists a unique solution u^ψ of the equation (25), in the class

$$C([0,1] \times B^{\otimes 2}_{R+1}) \cap \cap_{p>1} W^{1,2}_p([0,1] \times B^{\otimes 2}_{R+1}),$$

where $B^{\otimes 2}_{R+1} = \{(x,\tilde x):\, |x|\le R+1, |\tilde x|\le R+1\}$. In particular, for every $p>0$,

$$u^\psi_x \in L_p,$$

moreover, for a given K, – an upper bound for $\|\psi\|_C$, – the norms $\|u^\psi_x\|_{L_p}$ are uniformly bounded. Hence, for any two trajectories ψ, ψ' with the same bound K, the difference $v = u^\psi - u^{\psi'}$ satisfies the equation

$$L^\psi v = -(\psi_t - \psi'_t)u^{\psi'}_x,$$

with a zero boundary and initial values. If we treat here the right hand side as a given vector-function $u^{\psi'}_x \in \cap_{p>1} L_p$ with a small multiple $-(\psi_t-\psi'_t)$, the Theorem 4.9.1 [16] provides the *a priori* bound,

$$\|u^\psi - u^{\psi'}\|_{W^{1,2}_p} \le N\|\psi-\psi'\|_C.$$

Therefore, due to Sobolev's embedding theorems [16, Lemma 2.3.3], see also, [16, Corollary to the Theorem 4.9.1], we conclude that there exist $N,\alpha>0$ (here N is, generally speaking, different, and α could be chosen arbitrarily close to 1 such that

$$\|u^\psi - u^{\psi'}\|_C \le N\|\psi-\psi'\|^\alpha_C.$$

Hence, we also have,

$$\mathbb{E}^\zeta \zeta^{-1}(Y)g(Y)u^Y(0,x,\tilde x)$$

$$= \sum_i \mathbb{E}^\zeta 1(Y\in U^i)\zeta^{-1}(Y)g(Y)u^Y(0,x,\tilde x)$$

$$+\mathbb{E}^\zeta 1(Y\not\in \bigcup_i U^i)\zeta^{-1}(Y)g(Y)u^Y(0,x,\tilde x)$$

$$\approx \sum_i \mathbb{E}^\zeta 1(Y\in U^i)\zeta^{-1}(\psi^i)g(\psi^i)u^{\psi^i}(0,x,\tilde x).$$

The multiples $g(\psi^i)$ and $\mathbb{E}^\zeta_{0,x,\tilde x}1(Y\in U^i)$ in each term of the two sums,

$$\sum_i \mathbb{E}^\zeta 1(Y\in U^i)\mathbb{E}^\zeta_{0,x,\tilde x}\zeta^{-1}(\psi^i)g(\psi^i)u^{\psi^i}(0,x,\tilde x),$$

and (32), are the same, therefore, to prove (29) and, hence, (27), it suffices to show that

$$\mathbb{E}^\zeta_{0,x,\tilde x}\zeta^{-1}(\psi^i)1(D'_0)\varphi(X_1,\tilde X_1) = u^{\psi^i}(0,x,\tilde x)\mathbb{E}^\zeta\zeta^{-1}(\psi^i). \qquad (33)$$

Since $\mathbb{E}^\zeta\zeta^{-1}(\psi^i)=1$, the right hand side is simply the solution of the equation

$$u_s + \frac12\Delta_{xx}u + \frac12\Delta_{\tilde x\tilde x}u + (b(x) - (\psi^i_s-\psi^i_0)\nabla h(x))\nabla u$$

$$+(b(\tilde x) - (\psi^i_s-\psi^i_0)\nabla h(\tilde x))\nabla u = 0, \qquad 0\le s<1,\ \max(|x|,|\tilde x| < R+1), \qquad (34)$$

$$u(1,x,\tilde x) = \varphi(x,\tilde x), \quad \& \quad u(s,x,\tilde x) = 0,\ \forall\, 0\le s<1,\ \max(|x|,|\tilde x| = R+1),$$

in the corresponding Sobolev function class. The boundary conditions are also evident. But the fact that the left hand side of (33) satisfies (34) is well known, see, e.g., [11].

Thus, (26) holds true (a.s.). Next, all coefficients of the equation (20) and, hence, (25) are bounded, and the Harnack inequality, see [12], gives us,

$$u^Y(0,x,\tilde x) \le C(L,R) u^Y(0,x',\tilde x'),$$

with a constant $C(L,R)$ which depends only on the bounds on coefficients, but not on D. This can be rewritten exactly as (24), or, equivalently, as (23).

The denominator common for (21) and (22) can be estimated as follows,

$$\hat{\mathbb{E}}_{\tilde x, x}\left(\hat\gamma^{-1}(X,\tilde X, Y) \mid Y\right)$$

$$= \hat{\mathbb{E}}_{\tilde x, x}\Big(\exp(+\int_0^1 (Y_s - Y_0)^*\nabla h(X_s)\, dW_s + \int_0^1 (Y_s - Y_0)^*\nabla h(\tilde X_s)\, d\tilde W_s)$$

$$\times \exp(+\frac12 \int_0^1 \|(Y_s - Y_0)^*\nabla h(X_s)\|^2\, ds + \frac12 \int_0^1 \|(Y_s - Y_0)^*\nabla h(\tilde X_s)\|^2\, ds) \mid Y, \tilde Y\Big)$$

$$= \hat{\mathbb{E}}_{\tilde x, x}\Big(\exp(+\int_0^1 (Y_s - Y_0)^*\nabla h(X_s)\, dW_s + \int_0^1 (Y_s - Y_0)^*\nabla h(\tilde X_s)\, d\tilde W_s)$$

$$\times \exp(-\frac12 \int_0^1 \|(Y_s - Y_0)^*\nabla h(X_s)\|^2\, ds - \frac12 \int_0^1 \|(Y_s - Y_0)^*\nabla h(\tilde X_s)\|^2\, ds)$$

$$\times \exp(+\int_0^1 \|(Y_s - Y_0)^*\nabla h(X_s)\|^2\, ds + \int_0^1 \|(Y_s - Y_0)^*\nabla h(\tilde X_s)\|^2\, ds) \mid Y, \tilde Y\Big).$$

On the set $\{\Delta_0(Y) \le L\}$, here the latter exponential possesses the bounds,

$$C^{-1} \le \exp(+\int_0^1 \|(Y_s - Y_0)^*\nabla h(X_s)\|^2\, ds + \int_0^1 \|(Y_s - Y_0)^*\nabla h(\tilde X_s)\|^2\, ds) \le C,$$

with some $C = C(L,R)$.

Combining all representations and inequalities after the bounds (21) and (22), we finally get, on the set $\Delta_0(Y) = \sup_{0\le s\le 1}|Y_s - Y_0| \le L$,

$$\sup_{x,\tilde x \in B_R} \mathbb{E}^\gamma_{x,\tilde x}\left(1(D_0 \cap \{(X_1,\tilde X_1) \in D\})\rho_{0,1}(X,Y)\rho_{0,1}(\tilde X, Y) \mid Y\right)$$

$$\le C(L,R) \inf_{x',\tilde x' \in B_R} \mathbb{E}^\gamma_{x',\tilde x'}\left(1(D_0 \cap \{(X_1,\tilde X_1) \in D\})\rho_{0,1}(X,Y)\rho_{0,1}(\tilde X, Y) \mid Y\right), \quad (35)$$

as required. This is a so-called "mixing condition" in the sense of [17] for conditional measures, or (conditional) kernels, and, due to the Proposition 3.9 [17], this bound implies (16), as required, with some $\pi_R(L) \le \dfrac{C(L,R)^2-1}{C(L,R)^2+1}$, where $C(L,R)$ is a constant from (35). For the sake of completeness, we show the calculus that leads to (35). For any $|x|, |x'|, |\tilde x|, |\tilde x'| \le R$, and always on the set $\Delta_0(Y) = \sup_{0\le s\le 1}|Y_s - Y_0| \le L$,

$$
\begin{aligned}
&\mathbb{E}^{\gamma}_{x,\tilde x}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\rho_{0,1}(X,Y)\rho_{0,1}(\tilde X,Y)\mid Y\right)\\
&= \frac{\hat{\mathbb{E}}_{x,\tilde x}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\rho_{0,1}(X,Y)\rho_{0,1}(\tilde X,Y)\hat\gamma^{-1}\mid Y\right)}{\hat{\mathbb{E}}_{x,\tilde x}\left(\hat\gamma^{-1}\mid Y\right)}\\
&\le C(L,R)\hat{\mathbb{E}}_{x,\tilde x}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\mid Y\right)\\
&\le C(L,R)\hat{\mathbb{E}}_{x',\tilde x'}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\mid Y\right),
\end{aligned}
$$

and similarly,

$$
\begin{aligned}
&\mathbb{E}^{\gamma}_{x,\tilde x}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\rho_{0,1}(X,Y)\rho_{0,1}(\tilde X,Y)\mid Y\right)\\
&= \frac{\hat{\mathbb{E}}_{x,\tilde x}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\rho_{0,1}(X,Y)\rho_{0,1}(\tilde X,Y)\hat\gamma^{-1}\mid Y\right)}{\hat{\mathbb{E}}_{x,\tilde x}\left(\hat\gamma^{-1}\mid Y\right)}\\
&\ge C(L,R)^{-1}\hat{\mathbb{E}}_{x,\tilde x}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\mid Y\right)\\
&\ge C(L,R)^{-1}\hat{\mathbb{E}}_{x',\tilde x'}\left(1(D_0\cap\{(X_1,\tilde X_1)\in D\})\mid Y\right).
\end{aligned}
$$

Thus, indeed, (35) follows.

4. Denote by $\#1(\delta)$ the (non-random) number of ones in the vector δ, and by $\#1(\delta, L)$ the (random) number of those of them, say, with an index "i", which enjoys the property $\Delta_i(Y) \le L$ on the corresponding unit intervals. Then, by induction from n to 2, we get from (16),

$$
\begin{aligned}
&h((\mu_0,\nu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}, (\nu_0,\mu_0)\hat S_t^{Y;R;\delta;\mu_0,\nu_0}) \equiv h\left((\mu_0,\nu_0)S_t^{Y;R;\delta}, (\nu_0,\mu_0)S_t^{Y;R;\delta}\right)\\
&= h\left((\mu_0,\nu_0)\prod_{i=0}^{n-1} S_{i:i+1}^{Y;R;\delta}, (\nu_0,\mu_0)\prod_{i=0}^{n-1} S_{i:i+1}^{Y;R;\delta}\right) \le C\pi_{\mathbb{R}}^{(\#1(\delta,L)-2)_+}. \qquad (36)
\end{aligned}
$$

We use the fact that even if the starting measures μ_0 and ν_0 are not comparable in the sense of the Birkhoff metric, they become comparable after the first application of the inequality (35). So, we may lose up to two units from $\#1(\delta, L)$ in the exponential bound. From (13) and (36), it follows that

$$\mathbb{E}_{\mu_0}\|\mu_0\bar{S}_n^{Y;\mu_0} - \nu_0\bar{S}_n^{Y;\nu_0}\|_{TV} \le C\,\mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\delta\in\Delta}\ \pi_R^{(\#1(\delta,L)-2)_+}e_n^{Y;\delta;\mu_0,\nu_0}$$

$$= C\,\mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\#1(\delta,L)\ge\epsilon n}\ \pi_R^{\#1(\delta,L)-2}e_n^{Y;\delta;\mu_0,\nu_0}$$

$$+C\,\mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\#1(\delta,L)<\epsilon n}\ \pi_R^{(\#1(\delta,L)-2)_+}e_n^{Y;\delta;\mu_0,\nu_0}. \qquad (37)$$

Notice that whatever $\epsilon > 0$ (assuming only $\epsilon n \ge 2$), we have,

$$\mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\delta:\,\#1(\delta,L)\ge\epsilon n}\ \pi_R^{\#1(\delta,L)-2}e_n^{Y;\delta;\mu_0,\nu_0}$$

$$\le \mathbb{E}_{\mu_0,\nu_0}\pi_R^{\epsilon n-2}\sum\nolimits_{\delta:\,\#1(\delta,L)\ge\epsilon n}\ \mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}$$

$$= \pi_R(L)^{\epsilon n-2}\mathbb{E}_{\mu_0,\nu_0}\mathbb{E}_{\mu_0,\nu_0}(\bigcup\nolimits_{\delta:\,\#1(\delta,L)\ge\epsilon n} 1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}\le \pi_R(L)^{\epsilon n-2}. \qquad (38)$$

In the second sum for the terms with $\#\mathbf{1}(\delta, L) < 2$ we will not use the Birkhoff metric at all, nor will we use contraction in that metric in such a case. In particular, the question of comparability of μ_0 and ν_0 will not be important.

5. Hence, our main task remains to estimate the second term of the sum,

$$\mathbb{E}_{\mu_0,\nu_0}\sum_{\delta:\,\#1(\delta,L)<\epsilon n}\pi_R^{(\#1(\delta,L)-2)_+}\left(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}\right)$$

$$\le \mathbb{E}_{\mu_0,\nu_0}\sum_{\delta:\,\#1(\delta,L)<\epsilon n}\left(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}\right).$$

In turn, we will split this sum into two parts, as follows. Let $\epsilon' > \epsilon$. Then,

$$\mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\delta:\,\#1(\delta,L)<\epsilon n}\left(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}\right)$$

$$= \mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\delta:\,\#1(\delta,L)<\epsilon n,\,\#1(\delta)<\epsilon' n}\left(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}\right)$$

$$+\mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\delta:\,\#1(\delta,L)<\epsilon n,\,\#1(\delta)\ge\epsilon' n}\left(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}\right)$$

$$\le \mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\delta:\,\#1(\delta)<\epsilon' n}\left(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y)\mid Y,\tilde{Y})\mid_{\tilde{Y}=Y}\right)$$

$$+\mathbb{E}_{\mu_0,\nu_0}\sum\nolimits_{\delta:\,\#1(\delta,L)<\epsilon n,\,\#1(\delta)\ge\epsilon' n}\ 1. \qquad (39)$$

Let us show how to tackle the second term in (39). We fix any ϵ satisfying $0 < \epsilon < \epsilon'$. Denote $p_L = \mathbb{P}(\sup_{0\le s\le 1}|Y_s - Y_0| \le L)$, and $q_L = 1 - p_L$; notice that $p_L \approx 1$, if L is large enough. We have,

$$\mathbb{E}_{\mu_0,\nu_0} \textstyle\sum_{\delta:\, \#1(\delta,L)<\epsilon n,\, \#1(\delta)\ge \epsilon' n} 1$$

$$= \mathbb{P}_{\mu_0,\nu_0}\left(\#1(\delta, L) < \epsilon n,\, \#1(\delta) \ge \epsilon' n\right)$$

$$= \mathbb{P}_{\mu_0,\nu_0}\left(\text{there is at least } (\epsilon' - \epsilon)n \text{ unit intervals where } \Delta_i(Y) > L\right)$$

$$\le \textstyle\sum_{k=(\epsilon'-\epsilon)n}^{n} C_n^k q_L^k p_L^{n-k} \le (2\, q_L^{\epsilon'-\epsilon})^n, \tag{40}$$

which can be made less than any exponential by choosing L large enough. Notice that up to this point, the proof of the Theorem 2 will be identical.

6. Let us estimate the first term in (39). We have,

$$\sum_{\delta:\, \#1(\delta)<\epsilon n} \mathbb{E}_{\mu_0}\left(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y, \tilde Y)\Big|_{\tilde Y = Y}\right)$$

$$= \mathbb{E}_{\mu_0}\left(\sum_{\delta:\, \#1(\delta)<\epsilon n} \mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,\tilde X) \mid Y, \tilde Y)\Big|_{\tilde Y = Y}\right)$$

$$= \mathbb{E}_{\mu_0}\left(\mathbb{E}_{\mu_0,\nu_0}(\sum_{\delta:\, \#1(\delta)<\epsilon n} 1_\delta(X,\tilde X) \mid Y, \tilde Y)\Big|_{\tilde Y = Y}\right). \tag{41}$$

Let us introduce some new notations:

$$\#1(X)_R := \sum_{k=0}^{n-1} 1(|X_k| \le R,\, \sup_{k\le s\le 1} |X_s| < R+1),$$

$$\#0(X)_R := n - \#1(X)_R.$$

By the Dirichlet principle we notice,

$$1(\#1(X)_R \ge \frac{1+\epsilon}{2} n,\, \#1(\tilde X)_R \ge \frac{1+\epsilon}{2} n) \sum_{\delta:\, \#1(\delta)<\epsilon n} 1_\delta(X,\tilde X) = 0, \tag{42}$$

Indeed, notice that $\sum_{\delta:\, \#1(\delta)<\epsilon n} 1_\delta(X,\tilde X) = 1(\sum_{i=0}^{n-1} 1(D_i) < \epsilon n)$. If $\#1(X)_R = n$, $\#1(\tilde X)_R = n$, then $\sum_{i=0}^{n-1} 1(D_i) = n$. If we decrease either of the terms $\#1(X)_R = n$ or $\#1(\tilde X)_R = n$ by one, this can decrease the value $\sum_{i=0}^{n-1} 1(D_i)$ *at most* by one. Therefore, to make this sum less than ϵn, we must have $\#0(X)_R + \#0(\tilde X)_R > n - \epsilon n$, and so, either $\#0(X)_R > (1-\epsilon)n/2$, or $\#0(\tilde X)_R >$

$(1-\epsilon)n/2$. But both inequalities contradict $\#1(X)_R \ge \frac{1+\epsilon}{2}n$, & $\#1(\tilde X)_R \ge \frac{1+\epsilon}{2}n$, which implies (42), as required.

Hence, we get,

$$
\begin{aligned}
&\mathbb{E}_{\mu_0}\left(\mathbb{E}_{\mu_0,\nu_0}(\sum_{\delta:\,\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X)\mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right) \\
&\le \mathbb{E}_{\mu_0}\left(\mathbb{E}_{\mu_0,\nu_0}\left(1(\#1(X)_R<\frac{1+\epsilon}{2}n)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right) \\
&+\mathbb{E}_{\mu_0}\left(\mathbb{E}_{\mu_0,\nu_0}\left(1(\#1(\tilde X)_R<\frac{1+\epsilon}{2}n)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right). \\
&= \mathbb{E}_{\mu_0}\left(1(\mathbb{E}_{\mu_0}\left(1(\#1(X)_R<\frac{1+\epsilon}{2}n)\mid Y\right)\right) \\
&+\mathbb{E}_{\mu_0}\left(\mathbb{E}_{\nu_0}\left(1(\#1(\tilde X)_R<\frac{1+\epsilon}{2}n)\ \mid \tilde Y\right)\Big|_{\tilde Y=Y}\right),
\end{aligned}
\tag{43}
$$

because X does not depend on $\tilde Y$, nor $\tilde X$ depends on Y. Remind that we always use versions of conditional expectations continuous with respect to $Y, \tilde Y$. We estimate,

$$
\mathbb{E}_{\mu_0}\left(\mathbb{E}_{\mu_0}\left(1(\#1(X)_R<\frac{1+\epsilon}{2}n)\mid Y\right)\right) = \mathbb{E}_{\mu_0}1(\#1(X)_R<\frac{1+\epsilon}{2}n).
$$

Next, we estimate the other term, with the help of (A3), because $\mathbb{E}_{\mu_0}F(X,Y) \le C\mathbb{E}_{\nu_0}F(X,Y)$, $\forall F \ge\ 0$, and $\mathbb{E}_{\nu_0}F(X,Y) = \mathbb{E}_{\nu_0}F(\tilde X,\tilde Y)$,

$$
\begin{aligned}
&\mathbb{E}_{\mu_0}\left(\mathbb{E}_{\nu_0}\left(1(\#1(\tilde X)_R<\frac{1+\epsilon}{2}n)\ \mid \tilde Y\right)\Big|_{\tilde Y=Y}\right) \\
&\le C\,\mathbb{E}_{\nu_0}\left(\mathbb{E}_{\nu_0}\left(1(\#1(\tilde X)_R<\frac{1+\epsilon}{2}n)\ \mid \tilde Y\right)\Big|_{\tilde Y=Y}\right) \\
&= C\,\mathbb{E}_{\nu_0}\left(1(\#1(\tilde X)_R<\frac{1+\epsilon}{2}n)\right),
\end{aligned}
$$

7. Due to the bounds which easily follow from [24] and [25], the latter expectation possesses an appropriate bound, exponential or polynomial, depending on the value p, if R is chosen large enough, namely,

$$
\mathbb{E}_{\mu_0}1(\#1(X)_R<\epsilon'' n) \le \left\{\begin{array}{ll} C_m n^{-m}, & p=1,\ \ \forall\, m>0, \\ C\exp(-cn), & p=0. \end{array}\right. \tag{44}
$$

The same inequality holds true for ν_0 and $\tilde X$, also. This follows from the hitting time estimates for $\hat\tau = \inf(t\ge 0:\ |X_t|\le R)$, from [24] and [25], – see some details about a reduction to a simple one-dimensional diffusion below in the proof of the Theorem 2, –

$$\left\{\begin{array}{lll} \mathbb{E}_x\hat\tau^k \le C_m(1+|x|^m) & (\forall m>2k) & (p=1,\ \forall k>0), \\ \mathbb{E}_x\exp(\alpha\hat\tau) \le C\exp(c|x|) & (\exists C,c,\alpha>0) & (p=0), \end{array}\right. \tag{45}$$

due to the bound $\mathbb{P}_{\mu_0}(\#1(X)_R < \epsilon'' n) \le \mathbb{P}_{\mu_0}(\hat\tau_{\epsilon'' n} > n)$, – where $\hat\tau_1 = \hat\tau$, and by induction $\hat\tau_{n+1} := \inf(t \ge \hat\tau_n : |X_t| \le R)$, $n \ge 1$, – and due to exponential Chebyshev's inequality in the case $p = 0$, and by standard inequalities for semi-martingales in the case $p = 1$. Indeed, by virtue of (45), we have by induction, in each case:

[$p=0$] We have,

$$\mathbb{P}_x(\hat\tau_{\epsilon'' n} > n) \le \exp(-\alpha n + \epsilon'' n \ln C + c|x|),$$

where the value $C := \sup_{|x|\le R}\mathbb{E}\exp(\alpha\hat\tau)$ can be done arbitrarily close to one, by choosing R large enough. This provides an exponential bound for the probability $\mathbb{P}_x(\hat\tau_{\epsilon'' n} > n)$.

[$p=1$] Let $\epsilon'' < \epsilon''' < 1$, and let R be large enough, so that $\epsilon'' \sup_{|x'|\le R}\mathbb{E}_{x'}((\hat\tau_k - \hat\tau_{k-1}) \mid X_{\hat\tau_{k-1}}) < \epsilon''' < 1$. As it was proved in [25], with any $k > 0$ and any $m > 2k > 0$,

$$\mathbb{P}_x(\hat\tau_{\epsilon'' n} > n) = \mathbb{P}_x(\sum_{i=1}^{\epsilon'' n}(\hat\tau_k - \hat\tau_{k-1}) > n) \le C_m(1+|x|^m)((1-\epsilon''')n)^{-k} n^{k/2}.$$

This is a consequence of the simple fact that the moment $E_x\left(\sum_{i=1}^{\epsilon'' n}(\hat\tau_k - \hat\tau_{k-1}) - \mathbb{E}(\hat\tau_k - \hat\tau_{k-1} \mid \mathcal{F}_{\hat\tau_{k-1}})\right)^k$ grows not faster than $n^{k/2}$. The inequality gives any polynomial bound for the probability $\mathbb{P}_x(\hat\tau_{\epsilon'' n} > n)$. Hence, in both cases we have the bound (44).

8. Due to (43) and (44) we get estimate for the expression in (41):

$$\mathbb{E}_{\mu_0}\left(\mathbb{E}_{\mu_0,\nu_0}(\sum_{\delta:\,\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X) \mid Y,\tilde Y)\Big|_{\tilde Y = Y}\right) \le \left\{\begin{array}{ll} C_m n^{-m}, & p=1, \\ C\exp(-cn), & p=0. \end{array}\right.$$

for all $m > 0$ in the case $p = 1$. Combining this with the earlier inequalities (37) – (39), and (40), we obtain the final estimate, (10), in the case $p = 0$ possibly with a new constant c in the exponential. The Theorem 1 is proved.

4 Proof of Theorem 2

1 (6). Remind that we continue the proof starting from the estimates (39)–(40), and our chosen enumeration style of its steps (**(1 (6)), (2 (7))**, etc.) refers to this point. In particular, the assumption $h \in C^2$ has been already used. Remind also that now we cannot use the assumption (A3), and it was not used earlier.

Let us apply the formula (8) to the combined process $(X, Y, \tilde X, \tilde Y)$:

$$\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X, \tilde X) \mid Y, \tilde Y)$$

$$= \frac{\mathbb{E}^\gamma_{\mu_0,\nu_0}\left(1_\delta(X, \tilde X)\gamma^{-1}(X, Y; \tilde X, \tilde Y) \mid Y, \tilde Y\right)}{\mathbb{E}^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde Y)}\,. \tag{46}$$

Next, due to the Cauchy–Bouniakovsky–Schwarz inequality, for the conditional expectation we estimate the numerator in (46),

$$\mathbb{E}^\gamma_{\mu_0,\nu_0}\left(1_\delta(X, \tilde X)\gamma^{-1} \mid Y, \tilde Y\right)$$
$$\le \left(\mathbb{E}^\gamma_{\mu_0,\nu_0}\left(1_\delta(X, \tilde X) \mid Y, \tilde Y\right)\right)^{1/2} \left(\mathbb{E}^\gamma_{\mu_0,\nu_0}\left(\gamma^{-2} \mid Y, \tilde Y\right)\right)^{1/2}.$$

The denominator will be treated separately.

The further plan uses the following idea. Firstly, under $\mathbb{P}^\gamma$, the couples of processes $(X, \tilde X)$ and $(Y, \tilde Y)$ are independent on $[0, t]$, – in fact, here even all four components are independent, – so that

$$\mathbb{E}^\gamma_{\mu_0,\nu_0}\left(1_\delta(X, \tilde X) \mid Y, \tilde Y\right) = \mathbb{E}^\gamma_{\mu_0,\nu_0} 1_\delta(X, \tilde X) = \mathbb{E}_{\mu_0,\nu_0} 1_\delta(X, \tilde X),$$

which is a non-random value. Secondly, we will show that the *expectation* of the second factor divided by the denominator, with respect to $\mathbb{P}$ (*not* $\mathbb{P}^\gamma$),

$$\mathbb{E}_{\mu_0}\frac{\left(\mathbb{E}^\gamma_{\mu_0,\nu_0}\left(\gamma^{-2} \mid Y, \tilde Y\right)|_{\tilde Y = Y}\right)^{1/2}}{\mathbb{E}^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde Y)\,|_{\tilde Y = Y}}$$
$$= \mathbb{E}^\gamma_{\mu_0}\gamma^{-1}\,\frac{\left(\mathbb{E}^\gamma_{\mu_0,\nu_0}\left(\gamma^{-2} \mid Y, \tilde Y\right)|_{\tilde Y = Y}\right)^{1/2}}{\mathbb{E}^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde Y)\,|_{\tilde Y = Y}}\,, \tag{47}$$

does not exceed some exponential $\exp(Ct)$, with a constant C that depends only on $\|h\|_B$ but not on R. Finally, it will be shown that the expression $\mathbb{E}^\gamma_{\mu_0,\nu_0} 1_\delta(X, Y)$ can be made smaller that $\exp(-Ct)$ with *any* $C > 0$, by an appropriate choice of R, if $\#1(\delta, L) < \epsilon n$. Hence, we will get an exponential bound for the first part of the sum (37). Let us start this programme.

2 (7). Denominator in (47). We are going to estimate it from below. We have,

$$\mathbb{E}^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde Y)\,|_{\tilde Y = Y}$$

$$= \mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(\int_0^t h(X_s)\,dY_s - \frac{1}{2}\int h^2(X_s)\,ds$$

$$+ \int_0^t h(\tilde{X}_s)\,d\tilde{Y}_s - \frac{1}{2}\int_0^t h^2(\tilde{X}_s)\,ds) \mid Y, \tilde{Y}) \mid_{\tilde{Y}=Y}$$

$$\geq e^{-Ct}\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(\int_0^t h(X_s)\,dY_s + \int_0^t h(\tilde{X}_s)\,d\tilde{Y}_s \mid Y, \tilde{Y}) \mid_{\tilde{Y}=Y}$$

$$= e^{-Ct}\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(\int_0^t h(X_s)\,dY_s) \mid Y)\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(\int_0^t h(\tilde{X}_s)\,d\tilde{Y}_s) \mid \tilde{Y}) \mid_{\tilde{Y}=Y}$$

$$\geq e^{-Ct}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-\int_0^t h(X_s)\,dY_s) \mid Y))^{-1}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-\int_0^t h(\tilde{X}_s)\,d\tilde{Y}_s) \mid \tilde{Y}))^{-1} \mid_{\tilde{Y}=Y}.$$

In other words,

$$\left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde{Y}) \mid_{\tilde{Y}=Y}\right)^{-1}$$

$$\leq e^{+Ct}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-\int_0^t h(X_s)\,dY_s) \mid Y))(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-\int_0^t h(\tilde{X}_s)\,d\tilde{Y}_s) \mid \tilde{Y})) \mid_{\tilde{Y}=Y}.$$

Remind that both conditional expectations here are continuous functions of Y and $\tilde{Y}$, correspondingly, and both suit well our further applications of the CBS inequality. We have, with $p > 1$, $r = 2p$,

$$\left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-\int_0^t h(\tilde{X}_s)\,dY_s) \mid Y))^p\right)^{1/p}$$

$$\leq \left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-p\int_0^t h(\tilde{X}_s)\,dY_s) \mid Y))\right)^{1/p}$$

$$= \left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}\exp(-p\int h(\tilde{X}_s)\,dY_s)\right)^{1/p}$$

$$= \left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}\exp(-p\int_0^t h(\tilde{X}_s)\,dY_s - r\int_0^t h(\tilde{X}_s)^2\,ds + r\int_0^t h(\tilde{X}_s)^2\,ds)\right)^{1/p}$$

$$\leq \left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}\exp(-2p\int_0^t h(\tilde{X}_s)\,dY_s - 2r\int_0^t h(\tilde{X}_s)^2\,ds)\right)^{1/2p}$$

$$\times \left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}\exp(2r\int_0^t h(\tilde{X}_s)^2\,ds)\right)^{1/2p}$$

$$= \left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}\exp(-2p\int_0^t h(\tilde{X}_s)\,dY_s - \frac{4p^2}{2}\int_0^t h(\tilde{X}_s)^2\,ds)\right)^{1/2p}$$

$$\times\left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}\exp(2r\int_0^t h(\tilde X_s)^2\,ds)\right)^{1/2p}$$
$$\le (e^{Ct})^{1/2p} = e^{Ct}.$$

Similarly, – in fact, even easier, – we estimate the term

$$\left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-\int_0^t h(X_s)\,dY_s) \mid Y))^p\right)^{1/p}$$
$$\le (\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(-p\int_0^t h(X_s)\,dY_s)))^{1/p} \le e^{Ct}.$$

3 (8). A bound for the numerator in (47). We are to estimate it from above. We have,

$$\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\gamma^{-2} \mid Y, \tilde Y)\mid_{\tilde Y=Y}$$

$$= \mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(2\int_0^t h(X_s)\,dY_s - \int h^2(X_s)\,ds)$$

$$\times\exp(2\int_0^t h(\tilde X_s)\,d\tilde Y_s - \int_0^t h(\tilde X_s)^2\,ds) \mid Y, \tilde Y)\mid_{\tilde Y=Y}$$

$$\le e^{+Ct}\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(2\int_0^t h(X_s)\,dY_s + 2\int_0^t h(\tilde X_s)\,d\tilde Y_s \mid Y, \tilde Y)\mid_{\tilde Y=Y}$$

$$= e^{+Ct}\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(2\int_0^t h(X_s)\,dY_s) \mid Y)\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(\int_0^t h(\tilde X_s)\,d\tilde Y_s) \mid \tilde Y)$$

$$\le e^{+Ct}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(2\int_0^t h(X_s)\,dY_s) \mid Y))(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(2\int_0^t h(\tilde X_s)\,d\tilde Y_s) \mid \tilde Y))\mid_{\tilde Y=Y}.$$

In other words,

$$\left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\gamma^{-2} \mid Y, \tilde Y)\mid_{\tilde Y=Y}\right)$$

$$\le e^{+Ct}(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(2\int_0^t h(X_s)\,dY_s) \mid Y))(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\exp(2\int_0^t h(\tilde X_s)\,d\tilde Y_s) \mid \tilde Y))\mid_{\tilde Y=Y}.$$

The rest is standard. Whence, we get the following estimate,

$$\mathbb{E}^{\gamma}_{\mu_0}\gamma^{-1}\,\frac{\left(\mathbb{E}^{\gamma}_{\mu_0,\nu_0}\left(\gamma^{-2} \mid Y, \tilde Y\right)\mid_{\tilde Y=Y}\right)^{1/2}}{\mathbb{E}^{\gamma}_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde Y)\mid_{\tilde Y=Y}} \le \exp(+Ct), \tag{48}$$

with some finite non-random $C > 0$.

4 (9). The term $\mathbb{E}_{\mu_0,\nu_0} 1_\delta(X, \tilde X) = \mathbb{E}_{\mu_0,\nu_0} 1_\delta(Z)$. Emphasize that we are looking for a bound of this *non-conditional* probability, which should be less than any exponential, if the value R is chosen large enough. Intuitively, this looks reasonable under (A1), given that this assumption is sufficient for an exponential beta-mixing. The reasoning is similar to establishing the bounds (44). However, since we need an exponential bound with *any* constant, we will provide more details here. Here we do not separate X and $\tilde X$ as above, because it does not really simplify the calculus.

Given the recurrent Markov diffusion $Z_t \in \mathbb{R}^{2d}$, let us firstly construct a one-dimensional Itô process with reflection which dominates the process $|Z_t|$, – which is, generally speaking, non-Markov, – and secondly, construct a one-dimensional Markov diffusion, $\zeta_t,\ t \ge 0$, which dominates this Itô process, and which will help us estimate from above the value $\mathbb{E}_{\mu_0,\nu_0} 1_\delta(Z)$. Notice that we could use the same Dirichlet principle idea as in the previous Theorem; however, this does not simplify the calculus, so we work with the couple $(X, \tilde X)$.

The process $\hat Z_t := |Z_t| \vee (R-2)$ is a one-dimensional reflected Itô process, – although, generally speaking, not Markov, – with some finite drift β_t, and with a unit diffusion coefficient on $\hat Z_t > R-2$. Indeed,

$$d|Z_t|^2 = dX_t^2 + d\tilde X_t^2 = 2\sum X_t^i dX_t^i + \sum (dX_t^i)^2 + 2\sum \tilde X_t^i d\tilde X_t^i + \sum (d\tilde X_t^i)^2$$

$$= (2d + 2\sum_i X_t^i b^i(X_t) + 2\sum_i \tilde X_t^i b^i(\tilde X_t))dt + 2\sum X_t^i dW_t^i + 2\sum \tilde X_t^i d\tilde W_t^i.$$

Notice that there is no local time here, which is correct, because $2d$-dimensional process $(X_t, \tilde X_t)$ with independent components whose distributions are equivalent to standard $2d$-dimensional Wiener measure cannot attain zero at any positive time. Hence, $\mathbb{P}(|Z_t| > 0,\, t > 0) = 1$, and, applying Itô's formula, we get, also without any local time,

$$d|Z_t| = d(|Z_t|^2)^{1/2} = \frac{1}{2|Z_t|} d|Z_t|^2 - \frac{1}{8|Z_t|^3}(d|Z_t|^2)^2$$

$$= |Z_t|^{-1}(d + \textstyle\sum_i X_t^i b^i(X_t) + \sum_i \tilde X_t^i b^i(\tilde X_t))dt$$

$$+ |Z_t|^{-1} \sum X_t^i dW_t^i + |Z_t|^{-1} \sum \tilde X_t^i d\tilde W_t^i - \frac{X_t^2 + \tilde X_t^2}{2|Z_t|^3}\, dt.$$

Let us notice that on the set $|Z_t| \ge R - 2 > 0$, the values of $|Z_t|$ are separated away from zero. So, applying Ito-Tanaka's formula, see [22], to $|Z_t| \vee (R-2)$, we get,

$$d\hat Z_t = \hat\beta_t dt + 1(\hat Z_t > R-2) d\hat W_t + d\hat\varphi_t,$$

with a standard Wiener process $\hat W_t$,

$$\hat W_t := \int_0^t \left(|Z_s|^{-1}\sum X^i_s dW^i_s + |Z_s|^{-1}\sum \tilde X^i_s d\tilde W^i_s\right),$$

e.g., due to the Lévy characterization theorem, cf. [18, Theorem 4.1], and

$$\hat\beta_t = \left(\frac{(d+\sum_i X^i_t b^i(X_t) + \sum_i \tilde X^i_t b^i(\tilde X_t))}{|Z_t|} - \frac{X_t^2+\tilde X_t^2}{2|Z_t|^3}\right) 1\left(|X_t|^2+|\tilde X_t|^2 > (R-2)^2\right).$$

Here $\hat\varphi_t$ is a local time of $\hat Z_t$ at $R-2$. Notice that if $|Z_t|$ is large enough, then $\hat\beta_t$ is a negative value with a large modulus, which follows easily from the assumption (A1).

Consider the process $\bar Z_t$ with values in $[R-2,\infty)$ satisfying the stochastic differential equation with non-sticky reflection at $R-2$, – hence, it is *strong Markov*, e.g., because of the strong uniqueness, –

$$d\bar Z_t = \bar\beta dt + d\hat W_t + d\bar\varphi_t, \quad \bar Z_t \ge R-2, \quad \bar Z_0 = \max(|X_0|, R-2),$$

with a unit diffusion, and a constant large negative drift,

$$\bar\beta := \sup_{|Z_t|>R-2} \hat\beta_t,$$

where $\bar\varphi$ is the local time of the process $\bar Z_t$ at $R-2$.

Notice that

$$1(|Z_t|\ge R-2)\,\bar\beta \ge 1(|Z_t|>R-2)\,\hat\beta_t.$$

A routine comparison theorem shows that with probability one for all $t\ge 0$,

$$\hat Z_t \le \bar Z_t, \quad t\ge 0.$$

For the reader's convenience and because for this particular comparison setting the authors do not know any good reference, we provide the proof, although, of course, it looks very standard. Denote

$$g(x) = x^4 1(x>0).$$

Then, $g\in C^2_b$, and due to the Itô formula,

$$dg(\hat Z_t - \bar Z_t) = 1(\hat Z_t > \bar Z_t) d(\hat Z_t - \bar Z_t)^4$$

$$= 1(\hat Z_t > \bar Z_t)\tfrac{12}{2}(\hat Z_t - \bar Z_t)^2 (1(\hat Z_t > R-2) - 1)^2 dt$$

$$+1(\hat Z_t > \bar Z_t)\times 4(\hat Z_t - \bar Z_t)^3\left((\hat\beta_t - \bar\beta)dt + (1(\hat Z_t > R-2)-1)d\tilde W_t + (d\hat\varphi_t - d\bar\varphi_t)\right)$$

$$\le 0,$$

because $1(\hat{Z}_t > R-2) - 1 = -1(\hat{Z}_t \le R-2)$, and since almost surely

$$1(\hat{Z}_t > \bar{Z}_t) \times (\hat{Z}_t - \bar{Z}_t)^3 (\hat{\beta}_t - \bar{\beta}) \le 0,$$

$$1(\hat{Z}_t > \bar{Z}_t) \times 1(\hat{Z}_t \le R-2) = 0,$$

and

$$1(\hat{Z}_t > \bar{Z}_t)(\hat{Z}_t - \bar{Z}_t)^3 (d\hat{\varphi}_t - d\bar{\varphi}_t) \le 0.$$

Hence,

$$Eg(\hat{Z}_t - \bar{Z}_t) = 0, \quad \forall t \ge 0,$$

which implies

$$\mathbb{P}(\hat{Z}_t \le \bar{Z}_t, \quad t \ge 0) = 1.$$

5 (10). Now the bound will be established for the process $\bar{Z}$. Let

$$\tau_0 := \inf(s \ge 0 : \ \bar{Z}_s \le R-1), \quad \inf(\varnothing) = n,$$

(remind that $n = [t]$), and

$$\tau_1' := \inf(s > [\tau_0]_+ : \ \bar{Z}_s \ge R),$$

where we denote by $[a]_+$ the minimal integer which is not less than a, or, in other words, $[a]_+ = -[-a]$;

$$\tau_1 := \inf(s \ge \tau_1' : \ \bar{Z}_s \le R-1),$$

etc. Further, let $\tau_k - \tau_k' =: \ell_k$. Denote $N = \sup(k : \tau_k' \le n)$. Notice that

$$(\#1(Z) \le \epsilon' n) \subset \left(\#1(\bar{Z}) \le \epsilon' n\right) = \left(\#0(\bar{Z}) > (1-\epsilon')n\right),$$

where

$$\#1(\bar{Z}) = \sum_{i=0}^{n-1} 1(Z_i \le R; \sup_{i \le s \le i+1} Z_s < R+1),$$

and, naturally, $\#0(\bar{Z}) = n - \#1(\bar{Z})$. Let us choose $\epsilon' < 1/8$.

Further, $\#0(\bar{Z}) > (1-\epsilon')n$ implies that at least $(1-\epsilon')n$ intervals of length one with integer endpoints from the set $[0,1], \ldots, [n-1,n]$ contain at least one point where $\bar{Z} \ge R$; and, of course, every time, with probability one this is not just one point, but also $\bar{Z} \ge R$ in some its neighbourhood. There are three possibilities for every such interval I: (1) either a period where $\bar{Z} \ge R$ starts in this interval; (2) or such a period ends in this interval (of course, (2) may occur along with (1) on the same interval); (3) or neither (1) nor (2), but $\bar{Z} \ge R$ on the whole I. The number of I's of the kind (1) and (2) together does not exceed $2N$ (because the total number of such intervals does not exceed N, by definition of N). Hence, on the set $N < n/8$, the number of the I's of the 3rd kind is at least $6n/8$, which implies that on this set $\sum \ell_k \ge 6n/8$. Therefore, we have shown that

$$\big(\#0(\bar Z) > 7n/8\big) \subset (N \ge n/8) \cup \left(\sum_{k=1}^{n/8} \ell_k(\bar Z) \ge 6n/8\right).$$

We estimate,

$$\mathbb{P}_{\mu_0,\nu_0}\left(\sum_{k=1}^{n/8} \ell_k(\bar Z) \ge \frac{6}{8}n\right) \le \exp(-6cn/8)(\mathbb{E}\prod_{k=1}^{n/8}\exp(c\ell_k(\bar Z))).$$

Notice that, for $c > 0$ fixed, by choosing R large enough, one can make the value $\mathbb{E}_{\mu_0,\nu_0}(\exp(c\ell_k(\bar Z)) \mid \bar Z_{\tau_{k-1}})$ be arbitrary close to one, uniformly with respect to $k \ge 1$ and ω. Here we use the second part of the assumption (A1) to tackle $k = 1$. Hence, we firstly fix c large, and secondly R large, so that, say,

$$\mathbb{P}_{\mu_0,\nu_0}\left(\sum_{k=1}^{n/8} \ell_k(\bar Z) \ge \frac{6}{8}n\right) \le \exp(-5cn/8).$$

Finally,

$$\mathbb{P}_{\mu_0,\nu_0}(N \ge n/8) = \mathbb{P}_{\mu_0,\nu_0}\left(\tau'_{n/8}(\bar Z) \le n\right).$$

Remind that $\tau'_{k+1} = \inf(t \ge [\tau_k]_+ : \bar Z_t \ge R)$, and let us show that for R large enough the event $\tau'_{n/8}(\bar Z) \le n$ is highly improbable. Indeed, denote by q the probability

$$q := \sup_{0\le s\le 1} \mathbb{P}(\sup_{[s]_+\le t\le [s]_+ +1} Z_t \ge R, \text{ or } Z_{[s]_+ +1} > R-1 \mid \bar Z_s = R-1);$$

this value is arbitrarily small if R is large enough; the idea is that if, given $\bar Z_s = R-1$, the event $0(\bar Z)_{[s]_+}$ occurs, then certainly either $\sup_{[s]_+\le t\le [s]_+ +1} Z_t \ge R$, or $Z_{[s]_+ +1} > R-1$ occurs, too; in fact, the first must occur, and the second is just complementary.

Then, using the strong Markov property of $\bar Z$, and the well-known combinatorial identity $\sum_{k_1+\ldots+k_m=r} 1 = C^m_{r+m}$ for any fixed m and r, we conclude that the probability $\mathbb{P}\left(\tau_{n/8}(\bar Z) \le n\right)$ does not exceed the following sum, with $m := [n/8]$ and every $k_i \ge 0$,

$$\mathbb{P}_{\mu_0,\nu_0}\left(\tau_{n/8}(\bar Z) \le n\right) \le \sum_{k_1+\ldots+k_m \le n} q^m 1^{k_1+\ldots+k_m}$$

$$\le q^m \sum_{r=0}^{n} \sum_{k_1+\ldots+k_m=r} 1 = q^m \sum_{r=0}^{n} C^m_{r+m} \le q^m \sum_{r=0}^{n} 2^{r+m} \le q^m 2^{n+m+1}$$

$$= \exp\left((9n/8)\ln 2 + (\ln q)[n/8] + \ln 2\right).$$

Indeed, let us comment on the first inequality here. If $\tau_{n/8}(\bar Z) \le n$, then we have at least $n/8$ intervals where $\bar Z \ge R$, and which start in a special manner (see the definition of τ and τ' above). After the next return to the level $R-1$, and starting from the next integer value of time, say, $t = m$ (before that no new exceeds of any level count), the process tries to exceed the level R during the interval $[m, m+1]$, *or* exceed the level $R-1$ only at $m+1$. Once either of these happens, we wait till the next τ', which might yet have not occurred, and repeat this procedure. The condition $\tau_{n/8}(\bar Z) \le n$ guarantees that this occurs sooner or later at least $n/8$ times before n. Denote by k_m the number of integer intervals until this event occurs for the m-th time after the previous occurrence. Then, the probability $\mathbb{P}_{\mu_0,\nu_0}(\tau_{n/8}(\bar Z) \le n)$ can be estimated from above by the value $\sum_{k_1+\ldots+k_m\le n} q^m 1^{k_1+\ldots+k_m}$, if we just estimate by 1 every (conditional) probability that the rare event does not occur, and by q the (also conditional) probability that it does occur.

More formally, let us consider another sequence of stopping times, where m is integer,

$$\tilde\tau_{k+1} := \inf(t > [\tau_k]_+ :\ \bar Z_t \ge R) \wedge \inf(m \ge [\tau_k]_+ :\ \bar Z_m \ge R-1)).$$

Notice that we use the "old" stopping times τ_k in this definition. Naturally, for every k, $\tilde\tau_k \le \tau'_k$; in particular, $\tilde\tau_{n/8} \le \tau'_{n/8}$. Hence,

$$\mathbb{P}_{\mu_0,\nu_0}(\tau_{n/8} \le n) \le \mathbb{P}_{\mu_0,\nu_0}(\tilde\tau_{n/8} \le n).$$

Now, the event $\tilde\tau_{n/8} \le n$ implies that the combined set of intervals $\cup_{k=0}^{-1+n/8}[\tau_k, [\tilde\tau_{k+1}]_-]$ has a total length not exceeding n. Hence, the total length of unit intervals with integer endpoints within this sum does not exceed n. We estimate from above by q every (conditional) probability that on the next interval there is a stop $\tilde\tau$, and simply by 1 the probability that there is no stop. The whole probability of the latter event is thence estimated from above by the sum,

$$\sum_{k_1+\ldots+k_m\le n} q^m 1^{k_1+\ldots+k_m},$$

and the rest of the calculus readily follows. Choosing q small enough, we can get here an exponential bound $\exp(-cn)$ for the first term in (39), with any $c > 0$ (remind that $n = [t]$),

$$\mathbb{E}_{\mu_0,\nu_0} \sum_{\delta:\, \#1(\delta)<\epsilon' n} \Big(\mathbb{E}_{\mu_0,\nu_0}(1_\delta(X,Y) \mid Y, \tilde Y) \mid_{\tilde Y = Y}\Big) \le C_0 \exp(-Ct). \qquad (49)$$

Combining (49) with earlier estimates (37), (38), (39), and (40), as well as (48), we get the final estimate (11). The Theorem 2 is proved.

Remark 4. The proof suggests that the second part of the assumption (A1) could be relaxed to some polynomial moments on both initial measures. This should probably imply the final polynomial bound for the difference of the two filters. In the Theorem 2 itself, most likely, exponential moments for both initial measures with "every $c > 0$" could be replaced by "some c large enough".

Acknowledgements

The first author thanks the Department of Mathematics of the University of Maine and the grant RFBR 05-01-00449 for support. The paper happened to be technically involved, and such kind of text is a good source of potential mistakes; we were lucky to be able, at least, to decrease their number with the help of the remarks of the anonymous referee, to whom the authors are also grateful.

References

[1] Atar, R., Zeitouni, O. Exponential stability for nonlinear filtering. Ann. Inst. H. Poincaré Probab. Statist. 33 (1997), 6, 697–725.

[2] Baxendale, P., Chigansky, P., Liptser, R. Asymptotic stability of the Wonham filter: ergodic and nonergodic signals. SIAM J. Control Optim. 43 (2004), 2, 643–669 (electronic).

[3] Budhiraja, A. S., Ocone, D. L. Exponential stability in discrete-time filtering for bounded observation noise, Systems Control Lettr., 30 (1997), 185-193.

[4] Budhiraja, A. S., Ocone, D. L. Exponential stability for discrete-time filtering for nonergodic signals, Stochastic Process. Appl., 82 (1999), 245-257.

[5] Chigansky, P., Liptser, R. Stability of nonlinear filters in nonmixing case. Ann. Appl. Probab. 14 (2004), 4, 2038–2056.

[6] Del Moral P., Guionnet A. On the stability of interacting processes with applications to filtering and genetic algorithms. Ann. Inst. H. Poincaré, 37(2001), 2, 155-194 (2001).

[7] Gulinsky, O. V., Veretennikov, A. Yu. Large deviations for discrete-time processes with averaging. VSP, Utrecht, 1993.

[8] Kallianpur, G. Stochastic filtering theory, Springer, New York, 1980.

[9] Kleptsyna, M. L., Veretennikov, A.Yu. On discrete time ergodic filters with wrong initial data, Laboratoire de statistique et processus, Université du Maine, prépublication 06-3, 2006.
http://www.univ-lemans.fr/sciences/statist/download/Kleptsyna/filt19fg.pdf

[10] Krasnosel'skii, M. A., Lifshits, E. A., Sobolev, A. V. Positive linear systems. The method of positive operators. Heldermann Verlag, Berlin, 1989.

[11] Krylov, N. V. Controlled diffusion processes, Springer-Verlag, 1980.

[12] Krylov, N. V. Nonlinear elliptic and parabolic equations of the second order. D. Reidel Publishing Co., Dordrecht, 1987.

[13] Krylov, N. V. The regularity of conditional probabilities for stochastic processes, Theory Probab. Appl. 18 (1973), 151-155.

[14] Krylov, N. V., Safonov, M. V. A property of the solutions of parabolic equations with measurable coefficients. (Russian) Izv. Akad. Nauk SSSR Ser. Mat. 44 (1980), 1, 161-175, 239.

[15] Kunita, H. Asymptotic behavior of the nonlinear filtering errors of Markov processes. J. Multivariate Anal. 1 (1971), 365–393.

[16] Ladyženskaja, O. A., Solonnikov, V. A., Ural'ceva, N. N. Linear and quasilinear equations of parabolic type. AMS, Providence, R.I. 1967.

[17] Le Gland, F., Oudjane, N. A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals, Stochastic Process. Appl., 106 (2003), 279-316.

[18] Liptser, R. Sh., Shiryaev, A. N. Statistics of random processes, I : general theory. Springer, New York, 1977.

[19] Ocone, D., Pardoux, E. Asymptotic stability of the optimal filter with respect to its initial condition. SIAM J. Control Optim. 34 (1996), 1, 226–243.

[20] Parthasarathy, K. R. Probability measures on metric spaces, Academic Press, New York and London, 1967.

[21] Stannat, W. Stability of the pathwise filter equations for a time dependent signal on $\mathbb{R}^d$, Appl. Math. Optim., 52 (2005), 39-71.

[22] Tanaka, H. Note on continuous additive functionals of the 1-dimensional Brownian path, Z. Wahrsch. Gebiete, 1 (1963), 251-257.

[23] Veretennikov, A. Yu. Parabolic equations and Ito stochastic equations with coefficients discontinuous in time variable.Math. Notes. (1982),31(3-4),278-283

[24] Veretennikov, A. Yu. Estimates of the mixing rate for stochastic equations. (Russian) Teor. Veroyatnost. Primenen. 32 (1987), no. 2, 299–308; English translation: Theory Probab. Appl. 32 (1987), no. 2, 273–281.

[25] Veretennikov, A. Yu. On polynomial mixing and the rate of convergence for stochastic differential and difference equations. Teor. Veroyatn. i Primenen. 44 (1999), 2, 312–327; Engl. transl. in Theory Probab. Appl. 44 (2000), 2, 361–374.