Probab. Theory Relat. Fields (2008) 141:411–444
DOI 10.1007/s00440-007-0089-7

On discrete time ergodic filters with wrong initial data

M. L. Kleptsyna · A. Yu. Veretennikov

Received: 13 October 2005 / Revised: 5 June 2007 / Published online: 24 August 2007

Abstract For a class of *non-uniformly* ergodic Markov chains (X_n) satisfying exponential or polynomial beta-mixing, under observations (Y_n) subject to an IID noise with a positive density, it is shown that wrong initial data is forgotten in the mean total variation topology, with a certain exponential or polynomial rate.

Mathematics Subject Classification (2000) 60G35 · 62M20 · 93E11 · 93E15

1 Introduction

We consider a discrete time filter for a Markov chain (X_n) with values in the Euclidean space R^d, with observations (Y_n) from R^ℓ,

$$X_{n+1} = X_n + b(X_n) + \sigma(X_n)\xi_{n+1}, \qquad (n \geq 0), \tag{1}$$

$$Y_n = h(X_n) + V_n \quad (n \geq 1). \tag{2}$$

Here (ξ_n, V_n) is a sequence of IID random vectors of dimension $d + \ell$ with densities denoted by $q_\xi(x)\, q_V(y)$, $b(\cdot)$ is a d-dimensional vector-function, $\sigma(\cdot)$ a $d \times d$ matrix-function, $h(\cdot)$ an ℓ-dimensional vector-function. Suppose the exact initial distribution

M. L. Kleptsyna
Université du Maine, Le Mans, France
e-mail: Marina.Kleptsyna@univ-lemans.fr

A. Yu. Veretennikov (✉)
School of Mathematics, University of Leeds, Leeds LS2 9JT, UK
e-mail: A.Veretennikov@leeds.ac.uk

A. Yu. Veretennikov
Institute of Information Transmission Problems, Moscow, Russia

of X_0 denoted by μ_0, is known with some error. The main problem addressed in this paper is whether or not this error is forgotten by the optimal filtering algorithm in the long run. More precisely the setting is explained in the Sect. 2.2 below, because it requires new definitions. This setting is called a memoryless observation channel. Notice that if observations were not conditionally independent given X's, say,

$$U_n = \sum_{i=1}^{n} (h(X_i) + V_i),$$

then one could easily reconstruct its memoryless equivalent version,

$$Y_n = U_n - U_{n-1}.$$

So, conditional independence of the observations is not restrictive here.

As it was noticed in [14], the representation of the filtering algorithm below (see (7)) often allows a good modelling, although it rarely allows a good computing of all integrals.

Under *uniform ergodicity* assumptions, this problem was discussed in [2,7,14], where the limiting independence of the optimal filter algorithm on a wrong initial data has been established, along with certain exponential bounds, in continuous and discrete cases. See also [12], and discussion and further examples and counterexamples in [3,6]; see also [15] concerning a linear non-ergodic case.

The non-compact case attracted attention, especially in the last decade, and several papers considered this problem in the following partial cases. In [1] a non-compact case with linear observations and sufficiently small noise in observations was tackled. Small noise conditions appeared also in [5,16]. In [4], observation noise is assumed to be bounded, along with some additional condition on the positiveness of the noise density. In [14] and [13], mixing or "pseudo-mixing" conditions on the conditional kernel are assumed; the former is a uniform ergodic case which is essentially a compact case even though in a possibly non-compact state space, while the latter assumes that the observation noise is, again, small enough in a certain sense. All these cases do not allow Gaussian noise of an arbitrary level of intensity. In [17], a variational approach was applied to the filtering systems with a gradient type drift and linear observation part under additional assumptions. Our model is more general; this is essentially a non-compact and non-uniformly ergodic case, although it may not include some of the models from the cited papers.

Among technical tools that we use, there are the Birkhoff or Hilbert metric, as in [2] and [14], and recurrence or ergodicity bounds. The idea of the approach is that under appropriate ergodicity assumptions, the signal process spends most of its time in some compact ball, where the techniques of the compact case is applicable. The rest allows a good exponential or polynomial bound. The implementation of this idea, however, is less straightforward, due to several technical reasons. Firstly, we double our Markov process (X, Y) using another independent version of it, $(\tilde{X}, \tilde{Y})$, and, thence, we wish not the original signal process, but the doubled one, $(X, \tilde{X})$, to spend most of its time in some compact. Secondly, the Birkhoff metric does not suit well to any splitting of integrals and measures, so that this should be arranged in the original total variation metric. Thirdly, one has to estimate some unusual conditional probabilities after a certain substitution, $\tilde{Y} \mapsto Y$; they should be tackled using some new hints. We suggest two methods for this. The first method exploits the idea that good contraction holds true except on some rare event, related to the doubled signal process; this event is a subset of two further rare events, each of which is determined by only one of the two independent signal processes. Here "rare" means that both events allow some good estimates. This method essentially uses the condition (A3) below, and exponential or polynomial recurrence assumptions. The second method is based on a completely different idea how to tackle the same rare event, namely, using some analogue of Girsanov's change of measure, which makes signals and observations independent. This method requires more restrictive recurrence assumptions, but does not use the assumption (A3). Hence, the two main results are essentially different. At the same time, the "first part" of their proofs is the same, and this is the reason to combine them in one paper.

Technically, we use a simple splitting of the space of trajectories of our doubled signal process into 2^{n+1} parts, depending on whether it is in some appropriate compact or outside this compact at every fixed time from 0 to n. The common part of the proofs of the two theorems is an estimate on the "good" event that the doubled signal process visits some compact frequently.

Surprisingly, stopping times do not help here, possibly because they do not agree well enough with the Birkhoff metric.

The paper is arranged as follows: the Sect. 2 contains the assumptions, the main results, and some auxiliaries; the Sect. 3 is devoted to the proof of the first main result, and the Sect. 4 to the proof of the second one. We consider only Gaussian noises in the Theorem 2, although some results clearly can be obtained by this approach for more general models; see the Remark 3 after the statement of the Theorem 2 below.

2 Assumptions, main results, auxiliaries

2.1 Assumptions

The first group of assumptions serves the case of initial measures which satisfy the assumption of absolute continuity, see (A3) below. The measure equivalence is not assumed. The assumption (A3) is used directly in the end of the proof of the Theorem 1.

(A1) We assume that

$$0 < \inf_x \inf_{|\lambda|=1} \lambda^* \sigma\sigma^*(x)\lambda \le \sup_x \sup_{|\lambda|=1} \lambda^* \sigma\sigma^*(x)\lambda < \infty,$$

where $\lambda \in R^d$, the function b is locally bounded, and there exist $p = 0, 1$, $M > 0$ and $r \in (0, +\infty]$ such that

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

if $p = 1$ then we understand this as a limit with $r = +\infty$, that is,

$$\limsup_{|x|\to\infty} \left(\frac{|x + b(x)|}{|x|} - 1 \right) |x|^2 = -\infty. \tag{4}$$

(A2) The noise (ξ_n, V_n) is a sequence of IID random vectors with $E\xi_k = 0$; the component ξ_n and V_n are independent; if $p = 0$, we assume $E \exp(c|\xi|) < \infty$; if $p = 1$, we assume $E|\xi|^m < \infty$ for every $m > 0$; in all cases, for any $R > 0$ we assume,

$$C_R := \sup_{|x|,|v|\le R} \frac{q_\xi(x)}{q_\xi(v)} < \infty,$$

the density q_V is assumed to be positive everywhere; the function h is locally bounded.

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

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

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 *some* exponential moment for initial measures condition can be relaxed, depending on p. Gaussian noise condition is not used in this group of assumptions.

In the condition (A3) we do not require that the two measures are equivalent,but just an absolute continuity of μ_0 with respect to ν_0. The uniform norm L_∞ can be easily relaxed to $d\mu_0/d\nu_0 \in L_a$ with any $a > 1$; this would just change the constants in the final estimate in the Theorem 1. If $a = 1$—which means simply an absolute continuity—we can show just convergence in probability, without any useful bound. However, in the Theorem 2 we give certain estimate for convergence rate in this case under additional assumptions on other components of the system.

One may ask, what might happen if there is no even absolute continuity at all. In this case, if 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, and there is no absolute continuity of measures, 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.

The second group of assumptions serves the case when (A3) may fail. The recurrence and moment assumptions are more restrictive here, and the function h is bounded; the proof of the Theorem 2 is more involved.

(A′1) (*Recurrence*) We assume that

$$0 < \inf_x \inf_{|\lambda|=1} \lambda^* \sigma\sigma^*(x)\lambda \le \sup_x \sup_{|\lambda|=1} \lambda^* \sigma\sigma^*(x)\lambda < \infty,$$

the function b is locally bounded, and

$$\lim_{|x|\to\infty} (|x + b(x)| - |x|) = -\infty.$$

(A′2) (*Gaussian noises*) The noise (ξ_n, V_n) is an IID standard Gaussian random sequence of random variables of dimension $d + \ell$. The function h is bounded.

(A′3) (*Exponential moments*) Both initial measures μ_0 and ν_0 possess some exponential moment, that is, there exists $c > 0$ such that

$$\int e^{c|x|}(\mu_0(dx) + \nu_0(dx)) < \infty. \tag{6}$$

See the Remark 3 below concerning some possible extensions for (A′2).

2.2 Setting and main results

The setting is based on *the algorithm* that solves the exact filtering problem, which, as any Bayesian algorithm, depends on the initial data. Hence, we are going to plug in a new initial measure instead of the exact one into the algorithm. The filtered sequence is constructed via the observations (Y_n), as a sequence of conditional probabilities, $P_{\mu_0}(X_n \in \cdot \mid \mathcal{F}_n^Y)$—where $\mathcal{F}_n^Y = \sigma(Y_k : 1 \le k \le n)$—with the initial measure μ_0. Via the Bayes formula, this conditional measure can be represented as a *probability measure for any* Y, via the following non-linear operator $\bar{S}_n^{Y_1,\dots,Y_n,\mu_0}$, which we will denote for simplicity by $\bar{S}_n^{Y,\mu_0}$, applied to the measure μ_0,

$$P_{\mu_0}(X_n \in dx_n \mid Y_1, \dots, Y_n) = \int \prod_{i=1}^n Q(x_{i-1}, dx_i) c_i^{\mu_0} \Psi(x_i, Y_i) \mu_0(dx_0)$$

$$= d_n^{\mu_0} \int \prod_{i=1}^n Q(x_{i-1}, dx_i) \Psi(x_i, Y_i) \mu_0(dx_0) =: \mu_0 \bar{S}_n^{Y,\mu_0}(dx_n). \tag{7}$$

Here $\Psi(x_i, y_i)$ is a conditional density of Y_i at y_i, given $X_i = x_i$,

$$\Psi(x_i, y_i) = q_V(y_i - h(x_i))$$

(remind that q_V denotes the density of V_1), or

$$\Psi(x_i, Y_i) = q_V(Y_i - h(x_i)),$$

and $Q(x, dx')$ is a transition kernel for the Markov chain $X_n,\ n \geq 0$, that is,

$$Q(x, dx') = \frac{1}{\sqrt{\det(\sigma^*\sigma)(x)}}\, q_\xi(\sigma(x)^{-1}(x' - x - b(x)))\, dx'$$

(remind that q_ξ denotes the density of ξ_1). The random normalization constant $d_i^{\mu_0}$ is defined as follows,

$$d_i^{\mu_0} = \left(E_{\mu_0} \left(\prod_{j=1}^{i} \Psi(X_j, y_j) \right) \Bigg|_{y_1=Y_1,\dots y_i=Y_i} \right)^{-1},$$

and, correspondingly,

$$c_i^{\mu_0} = \frac{d_i^{\mu_0}}{d_{i-1}^{\mu_0}} = \frac{E_{\mu_0}\left(\prod_{j=1}^{i-1} \Psi(X_j, y_j)\right)}{E_{\mu_0}\left(\prod_{j=1}^{i} \Psi(X_j, y_j)\right)} \Bigg|_{y_1=Y_1,\dots y_i=Y_i}.$$

For the reader's convenience let us show the formula for $d_n^{\mu_0}$. Using standard notations for conditional and joint densities, dropping the index μ_0 for probability and expectation, and denoting just for this short explanation, $\bar{X}_n = (X_1, \dots, X_n)$, $\bar{Y}_n = (Y_1, \dots, Y_n)$, we have,

$$\begin{aligned} P(\bar{X}_n \in A \mid \bar{Y}_n) &= \int_A f(\bar{x}_n \mid \bar{Y}_n)\, d\bar{x}_n \\ &= \int_A \frac{f(\bar{x}_n, \bar{Y}_n)}{f(\bar{Y}_n)} \frac{f(\bar{x}_n)}{f(\bar{x}_n)}\, d\bar{x}_n = \frac{1}{f(\bar{Y}_n)} \int_A f(\bar{Y}_n \mid \bar{x}_n) f(\bar{x}_n)\, d\bar{x}_n \\ &= \frac{1}{f(\bar{Y}_n)} \int_A \prod \Psi(x_i, Y_i) f(\bar{x}_n)\, d\bar{x}_n = d_n \int_A \prod \Psi(x_i, Y_i) f(\bar{x}_n)\, d\bar{x}_n, \end{aligned}$$

due to the assumption of conditional independence of (Y_i)'s given (X_i). Hence, $d_n = f(\bar{Y}_n)^{-1}$, and for $A = (R^d)^n$, we get,

$$1 = d_n \int \prod \Psi(x_i, Y_i) f(\bar{x}_n)\, d\bar{x}_n = d_n E\left(\prod \Psi(X_i, y_i)\right)\Big|_{\bar{y}_n = \bar{Y}_n}.$$

Now the "wrong initialization" problem can be formulated more precisely as follows. One does not know the measure μ_0 exactly, but only some its approximation ν_0. Hence, one plugs in the observed values Y's and this new measure ν_0 into the formula (7). The problem is whether this algorithm forgets its wrong initial data in the long run, that is, whether the difference between the conditional measures provided by the algorithms with the exact and wrong initial data converges to zero in some suitable topology. However, even before we pose this question about convergence, we shall decide whether this operation of using ν_0 instead of μ_0 is well-defined. In which case it is well-defined and in which it is not? The answer is that it is not well-defined if and only if our actually observed vector $\bar{Y}_n$ is impossible under ν_0 for some n, or, equivalently, if the vector-value $(X_0, \ldots, X_n)$ starting from the distribution ν_0 is impossible under the observed $\bar{Y}_n$ for some n. Since clearly any value of $(X_0, \ldots, X_n)$ with $X_0 \in \text{supp}(\mu_0)$ is possible, we have a sufficient condition for our operation to be well-defined, $\text{supp}(\nu_0) \subset \text{supp}(\mu_0)$, or, equivalently,

$$\nu_0 \ll \mu_0. \tag{8}$$

This condition is sufficient whatever all other distributions are. Notice that in many papers on the subject this is, indeed, assumed. However, it is not necessary if we impose some other additional requirements, e.g., if the density of V_1 is positive everywhere, which we have assumed in both groups of our assumptions, see (A2) and (A$'$2).

Another issue is that while using the Birkhoff metric and induction we will need *equivalent* measures with bounded derivatives, see (12) below. So it looks as if (8) should have been assumed, at least. However, recall that induction can be started not necessarily from zero. On the other hand, after the first application of the “mixing inequality” (13) we will get comparable measures, that is, equivalent measures with bounded derivatives, as required. Thus, we can start our induction (12) from $n = 1$. This is why the condition (8), which seems so natural and nearly indispensable, in fact, is not required here.

Now we shall explain how one can interpret this setting in a probabilistic way, using again some Markov dynamics and conditioning. This is important for our presentation, although logically it might not necessarily follow from the previous paragraphs. In fact, for the initial distribution ν_0, we have another sequences of measures and observations,

$$d_n^{\nu_0} \int \prod_{i=1}^{n} Q(x_{i-1}, dx_i) \Psi(x_i, \tilde{Y}_i) \nu_0(dx_0) = \nu_0 \bar{S}_n^{\tilde{Y}, \nu_0}(dx_n) \equiv \nu_n(dx_n).$$

This can be, indeed, regarded as another conditional expectation, for the same Markov process starting from another initial distribution ν_0, given some new observations $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$. Without losing a generality, we can and will assume that this pair, $(\tilde{X}, \tilde{Y})$, is defined on some *independent probability space;* we will not change our notation for the probability measure, nor for expectation, though, both now apply to the process $(X, Y, \tilde{X}, \tilde{Y})$. However, due to the setting, only original observations Y are available, so that we are obliged to identify $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ with $(Y_1, \ldots, Y_n)$, that is, we *keep the original observations* that have risen from the original initial data μ_0, as though they were initialized by its substitution ν_0. The result, $\nu_0 \bar{S}_n^{Y, \nu_0}$, is still some conditional probability, namely, the conditional distribution of ν_n given $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$, after the values $(\tilde{Y}_1, \ldots, \tilde{Y}_n)$ have been replaced by $(Y_1, \ldots, Y_n)$. This operation is well defined almost surely with respect to the measure P_{μ_0}, due to our assumptions on the density q_V.

The *main question* here is about a discrepancy of the filter with a wrong measure ν_0 instead of μ_0 and the exact one, or, in other words, about the difference of the two measures,

$$(\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0})(dx_n),$$

whether it is reasonably small for large values of n. We will be interested in the distance in the mean total variation norm with respect to the original initial measure μ_0.

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

$$E_{\mu_0} \|\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0}\|_{TV} \le \begin{cases} C_m n^{-m}, & p = 1 \quad \forall m > 0, \\ C \exp(-cn), & p = 0. \end{cases} \tag{9}$$

2. *In addition, the following pathwise inequalities hold true*:

(i) *If*

$$E_{\mu_0} \|\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0}\|_{TV} \le C n^{-m},$$

then, for every $m' < m$, not necessarily integer, there exists a (*random*) n_0 *such that*

$$\|\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0}\|_{TV} \le n^{-m'+1}, \quad n \ge n_0.$$

(ii) *If*

$$E_{\mu_0} \|\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0}\|_{TV} \le C \exp(-cn),$$

then for any $c' < c$, there exists a (*random*) n_0 *such that*

$$\|\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0}\|_{TV} \le C \exp(-c'n), \quad n \ge n_0.$$

Theorem 2 1. *Under the assumptions* (A$'$1)–(A$'$3) *above, the following holds true*: *there exists $c_0 > 0$ such that*

$$E_{\mu_0} \|\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0}\|_{TV} \le C \exp(-c_0 n). \tag{10}$$

2. *In addition, the following pathwise inequalities hold true: for any $c < c_0$, there is a random time n_0 which is finite almost surely, such that*

$$\|\mu_0 \bar{S}_n^{Y,\mu_0} - \nu_0 \bar{S}_n^{Y,\nu_0}\|_{TV} \le C \exp(-cn), \quad n \ge n_0.$$

Remark 2 Notice that in the Theorem 1 the constants in the right hand side of the main inequality can be chosen uniformly over every class of problems with uniformly bounded values of the integrals in the assumption on the initial measures, all coefficients, non-degeneracy constant of the matrix $\sigma\sigma^*$, and all other constants (A2)–(A3). Similarly, one can state uniform bounds for appropriate classes of problems in the Theorem 2.

Remark 3 In the Theorem 2, non-Gaussian noises in both components of the system could be considered by this approach, too, under some general assumptions on the noise likelihood ratio functions. A simple sufficient condition for that is

$$\sup_{|x|,|\tilde{x}|,|x'|\leq R} \frac{Q(x,dx')}{Q(\tilde{x},dx')} < \infty \quad \forall R > 0,$$

for Q, and

$$0 < C^{-1} < \frac{q_V(y-h(x))}{q_V(y)} \leq C < \infty \quad \forall x, y,$$

for q_V (or Ψ), which can be checked for densities like $q_V(x) = \frac{1}{2}\exp(-|x|)$, and $q_V(x) = C_k(1+|x|)^{-k}$, $k \geq 2$. For densities with lighter tails our *hypothesis* is that the latter condition may be replaced by

$$0 < C^{-1} \leq \int \left(\frac{q_V(y-h(x))}{q_V(y)}\right)^4 dy \leq C < \infty,$$

or some further extensions of this condition. In this case, the likelihood ratios are the analogues of the Girsanov type exponentials in the proof of the Theorem 2 below. Of course, the noise centering condition is not essential under (A$'$1). We postpone a more detailed discussion about other cases till further papers.

Remark 4 One relevant and important counterexample can be found in [3], although it is not a counterexample to any particular theorem. A simpler counterexample *to theorems* could be constructed as follows. Suppose there is no mixing in the sense that the process X has two separated ergodic classes, and support of μ_0 lies in one of them, while support of ν_0 lies in the other one. Then, one just may not be able to use Y's in the filter algorithm, at least, if the support of the observation noise distribution is bounded. In this case the filter may never be able to start working with a wrong initial measure, and the principle that the filter forgets wrong initial data fails.

The continuous time case can be treated similarly, although with some essential technical changes; see [9,10].

3 The proof of Theorem 1

1. Let us introduce some indicators. First of all, in this proof, x stands for the whole sequence $(x_1,\ldots,x_n)$, and likewise for $\tilde{x}$, and the same for the random sequences

X, $\tilde X$, and Y. As suggested above in the setting, we consider independent couples (X, Y) and $(\tilde X, \tilde Y)$, with initial distributions of the first components, $\mathcal{L}(X_0) = \mu_0$ and $\mathcal{L}(\tilde X_0) = \nu_0$. For every $i \ge 0$, let $M_i := \max(|X_i|, |\tilde X_i|)$. For fixed R and n, we denote by δ a (non-random) vector of dimension $n+1$ with coordinates 1 or 0 at every place, and the following indicators, with a convention $0^0 = 1$,

$$1_\delta(X, \tilde X) := \prod_{i=0}^{n} (1(M_i \le R))^{\delta_i} \, (1 - 1(M_i \le R))^{1-\delta_i}$$
$$\equiv \prod_{i=0}^{n} (1(\delta_i = 1)1(M_i \le R) + 1(\delta_i = 0)1(M_i > R)).$$

Remind that this indicator function depends on R and n as parameters, which are dropped from the notation. In some cases it will be useful to present the latter indicator as

$$1_\delta(X, \tilde X) = \prod_{i=0}^{n} 1_\delta(i, M_i),$$

where for any $M > 0$,

$$1_\delta(i, M) = (1(\delta_i = 1)1(M \le R) + 1(\delta_i = 0)1(M > R)).$$

For every δ let us define

$$J = J(\delta) := \{i : 0 \le i \le n, \ \delta_i = 1\}.$$

Denote by Δ the set of all possible values of the vector δ.

Let us define new operators on the spaces of normalized and non-normalized measures on $R^{2d} = R^d \times R^d$, or, rather, on the space of pairs of measures, each on R^d, as follows,

$$(\mu, \nu)\bar S_n^{Y;\mu_0,\nu_0}(A \times B)$$
$$= \int\int 1(x_n \in A, \tilde x_n \in B)$$
$$\times \left(\prod_{i=1}^{n} c_i^{\mu_0} c_i^{\nu_0} \Psi(x_i, Y_i)\Psi(\tilde x_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde x_{i-1}, d\tilde x_i)\right) \mu(dx_0)\nu(d\tilde x_0),$$

and

$$(\mu_i, \nu_i) S_{i:i+1}^{Y;R;\delta}(A \times B)$$
$$= \int\int 1(x_{i+1} \in A, \tilde x_{i+1} \in B) 1_\delta(x_i, \tilde x_i) \, 1_\delta(x_{i+1}, \tilde x_{i+1})$$
$$\times \Psi(x_{i+1}, Y_{i+1})\Psi(\tilde x_{i+1}, Y_{i+1}) Q(x_i, dx_{i+1}) Q(\tilde x_i, d\tilde x_{i+1}) \mu_i(dx_i)\nu_i(d\tilde x_i),$$

and (we use a double integral notation just to emphasize that we integrate with respect to the variables x and $\tilde{x}$)

$$
\begin{aligned}
&(\mu, \nu)\bar{S}_n^{Y;R;\delta;\mu_0,\nu_0}(A \times B) \\
&\quad = \int\int 1(x_n \in A, \tilde{x}_n \in B) 1_\delta(x, \tilde{x}) \\
&\qquad \times \left(\prod_{i=1}^{n} c_i^{\mu_0} c_i^{\nu_0} \Psi(x_i, Y_i) \Psi(\tilde{x}_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde{x}_{i-1}, d\tilde{x}_i) \right) \mu(dx_0)\nu(d\tilde{x}_0),
\end{aligned}
$$

and

$$
\begin{aligned}
&(\mu, \nu) S_n^{Y;R;\delta}(A \times B) \\
&\quad = \int\int 1(x_n \in A, \tilde{x}_n \in B) 1_\delta(x, \tilde{x}) \\
&\qquad \times \left(\prod_{i=1}^{n} \Psi(x_i, Y_i) \Psi(\tilde{x}_i, Y_i) Q(x_{i-1}, dx_i) Q(\tilde{x}_{i-1}, d\tilde{x}_i) \right) \mu(dx_0)\nu(d\tilde{x}_0),
\end{aligned}
$$

which can be equivalently presented as

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

We remind that all the operators above correspond to the independent pairs of (X, Y) and $(\tilde{X}, \tilde{Y})$, due to the direct product of the integrals with respect to the variables x_i's and $\tilde{x}_i$'s.

Now we define a notion which will play a crucial role in the sequel. For every δ, let

$$
e_n^{Y;\delta;\mu_0,\nu_0} := (\mu_0, \nu_0)\bar{S}_n^{Y;R;\delta;\mu_0,\nu_0}(R^{2d}) \equiv E_{\mu_0,\nu_0}(1_\delta(Z) \mid Y, \tilde{Y})\Big|_{\tilde{Y}=Y},
$$

where $Z = (X, \tilde{X})$. Due to the assumption on the density q_V, these random variables are well-defined. Notice that the symmetry in the definition of $\bar{S}$ implies a very important identity,

$$
e_n^{Y;\delta;\mu_0,\nu_0} = e_n^{Y;\delta;\nu_0,\mu_0}.
$$

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

$$
\begin{aligned}
e_n^{Y;\delta;\mu_0,\nu_0} &= \iint 1_\delta(x,\tilde{x})\left(\prod_{i=1}^n c_i^{\mu_0} c_i^{\nu_0} \Psi(x_i,Y_i)\Psi(\tilde{x}_i,Y_i) Q(x_{i-1},dx_i) Q(\tilde{x}_{i-1},d\tilde{x}_i)\right) \\
&\quad \times \mu_0(dx_0)\nu_0(d\tilde{x}_0) = \int\int 1_\delta(\tilde{x},x) \\
&\quad \times \left(\prod_{i=1}^n c_i^{\mu_0} c_i^{\nu_0} \Psi(\tilde{x}_i,Y_i)\Psi(x_i,Y_i) Q(\tilde{x}_{i-1},d\tilde{x}_i) Q(x_{i-1},dx_i)\right) \\
&\quad \times \mu_0(dx_0)\nu_0(d\tilde{x}_0) (\text{by change of variables, } x_i \longleftrightarrow \tilde{x}_i, \text{ for all } i\text{'s}) \\
&= \iint 1_\delta(x,\tilde{x})\left(\prod_{i=1}^n c_i^{\nu_0} c_i^{\mu_0} \Psi(x_i,Y_i)\Psi(\tilde{x}_i,Y_i) Q(x_{i-1},dx_i) Q(\tilde{x}_{i-1},d\tilde{x}_i)\right) \\
&\quad \times \nu_0(dx_0)\mu_0(d\tilde{x}_0) = e_n^{Y;\delta;\nu_0,\mu_0}.
\end{aligned}
$$

Notice that the replacement of $\tilde{Y}$ by Y is well defined, because under the integrals we only have likelihood functions that depend on $\tilde{Y}$, but not on the couple Y, $\tilde{Y}$ together. Next, denote

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

The sense of the last notation is that the result of this action is a *normalized* measure restricted to the event $1_\delta(X,\tilde{X}) = 1$.

Next important step is due to the fact that the distance in total variation for the measures in R^d can be estimated from above via the correspondingly duplicated measures, and the latter can be split into 2^{n+1} terms as follows,

$$
\begin{aligned}
&\|\mu_0 \bar{S}_n^{Y;\mu_0,\nu_0} - \nu_0 \bar{S}_n^{Y;\nu_0,\mu_0}\|_{TV} \\
&\quad \le \|(\mu_0,\nu_0)\bar{S}_n^{Y;\mu_0,\nu_0} - (\nu_0,\mu_0)\bar{S}_n^{Y;\nu_0,\mu_0}\|_{TV} \\
&\quad \le \sum_{\delta\in\Delta} \|(\mu_0,\nu_0)\bar{S}_n^{Y;R;\delta;\mu_0,\nu_0} - (\nu_0,\mu_0)\bar{S}_n^{Y;R;\delta;\mu_0,\nu_0}\|_{TV} \\
&\qquad (\text{because } \|\textstyle\sum \cdot\| \le \sum \|\cdot\|) \\
&\quad = 2\sum_{\delta\in\Delta} \sup_D \Big(e_n^{Y;\delta;\mu_0,\nu_0}(\mu_0,\nu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D) - e_n^{Y;\delta;\nu_0,\mu_0}(\nu_0,\mu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D)\Big). \\
&\quad = 2\sum_{\delta\in\Delta} e_n^{Y;\delta;\mu_0,\nu_0} \sup_D \Big((\mu_0,\nu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D) - (\nu_0,\mu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D)\Big),
\end{aligned}
$$

where D runs all Borel sets $\mathcal{B}(R^{2d})$. The fact that the (random) normalization constant is the same for the two measures has been used essentially here. The first inequality in the chain of expressions above can be explained as follows,

$$
\begin{aligned}
&(1/2)\,\|(\mu_0,\nu_0)\bar S_n^{Y;\mu_0,\nu_0}-(\nu_0,\mu_0)\bar S_n^{Y;\nu_0,\mu_0}\|_{TV}\\
&=\sup_D\Bigg(\int 1((x_n,\tilde x_n)\in D)\mu_0(dx_0)\nu_0(d\tilde x_0)\\
&\quad\times\left(\prod_{i=1}^n c_i^{\mu_0}c_i^{\nu_0}\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)Q(x_{i-1},dx_i)Q(\tilde x_{i-1},d\tilde x_i)\right)\\
&\quad-\int 1((x_n,\tilde x_n)\in D)\nu_0(dx_0)\mu_0(d\tilde x_0)\Bigg)\\
&\quad\times\left(\prod_{i=1}^n c_i^{\mu_0}c_i^{\nu_0}\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)Q(x_{i-1},dx_i)Q(\tilde x_{i-1},d\tilde x_i)\right)\\
&\ge\sup_{A\in\mathcal B(R^d)}\int 1((x_n,\tilde x_n)\in A\times R^d)\mu_0(dx_0)\nu_0(d\tilde x_0)\\
&\quad\times\left(\prod_{i=1}^n c_i^{\mu_0}c_i^{\nu_0}\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)Q(x_{i-1},dx_i)Q(\tilde x_{i-1},d\tilde x_i)\right)\\
&\quad-\int 1((x_n,\tilde x_n)\in A\times R^d)\nu_0(dx_0)\mu_0(d\tilde x_0)\\
&\quad\times\left(\prod_{i=1}^n c_i^{\mu_0}c_i^{\nu_0}\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)Q(x_{i-1},dx_i)Q(\tilde x_{i-1},d\tilde x_i)\right)\Bigg)
\end{aligned}
$$

(here we will change variables in the second integral, $x\mapsto\tilde x$ and vice versa)

$$
\begin{aligned}
&=\sup_{A\in\mathcal B(R^d)}\Bigg(\int 1((x_n,\tilde x_n)\in A\times R^d)\mu_0(dx_0)\nu_0(d\tilde x_0)\\
&\quad\times\left(\prod_{i=1}^n c_i^{\mu_0}c_i^{\nu_0}\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)Q(x_{i-1},dx_i)Q(\tilde x_{i-1},d\tilde x_i)\right)\\
&\quad-\int 1((\tilde x_n,x_n)\in A\times R^d)\nu_0(d\tilde x_0)\mu_0(dx_0)\\
&\quad\times\left(\prod_{i=1}^n c_i^{\mu_0}c_i^{\nu_0}\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)Q(x_{i-1},dx_i)Q(\tilde x_{i-1},d\tilde x_i)\right)\Bigg)\\
&=\sup_{A\in\mathcal B(R^d)}\Bigg(\int 1(x_n\in A)\left(\prod_{i=1}^n c_i^{\mu_0}\Psi(x_i,Y_i)Q(x_{i-1},dx_i)\right)\mu_0(dx_0)\\
&\quad-\int 1(\tilde x_n\in A)\left(\prod_{i=1}^n c_i^{\nu_0}\Psi(\tilde x_i,Y_i)Q(\tilde x_{i-1},d\tilde x_i)\right)\nu_0(d\tilde x_0)\Bigg)\\
&=(1/2)\,\|\mu_0\bar S_n^{Y;\mu_0}-\nu_0\bar S_n^{Y;\nu_0}\|_{TV}.
\end{aligned}
$$

We will use the Birkhoff metric for positive measures, see [11], and also [2,14] (where it is called Hilbert metric; one more synonym is the projective metric),

$$\rho(\mu,\nu) = \begin{cases} \ln \dfrac{(\inf s: \ \mu \le s\nu)}{(\sup t: \ \mu \ge t\nu)}, & \text{if finite,} \\ +\infty, & \text{otherwise.} \end{cases}$$

Another equivalent definition reads,

$$\rho(\mu,\nu) = \begin{cases} \ln \sup(d\mu/d\nu) + \ln \sup(d\nu/d\mu), & \text{if finite,} \\ +\infty, & \text{otherwise.} \end{cases}$$

Due to the inequality for the total variation norm and the Birkhoff metric (see [2,14]), and since both measures below—that is, $(\mu_0,\nu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}$ and $(\nu_0,\mu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}$—are normalized, we have,

$$\begin{aligned} &2\sup_D \left((\mu_0,\nu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D) - (\nu_0,\mu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D)\right) \\ &\quad \le \frac{2}{\ln 3}\,\rho\left((\mu_0,\nu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}, (\nu_0,\mu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}\right). \end{aligned} \tag{11}$$

We claim that there exists $\pi_R < 1$ such that if $k \ge 1$, then

$$\begin{aligned} &\rho((\mu_0,\nu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}, (\nu_0,\mu_0)\hat{S}_n^{R;\delta;\mu_0,\nu_0}) \\ &\quad \equiv \rho\left((\mu_0,\nu_0)S_n^{Y;R;\delta}, (\nu_0,\mu_0)S_n^{Y;R;\delta}\right) \le C_R \pi_R^{k-1}, \end{aligned} \tag{12}$$

where

$$k = \#1(\delta) := \sum_{j=1}^{n} 1(j-1 \in J,\ j \in J),$$

or, in words, $\#1(\delta)$ is the total number of consequent *pairs of ones* in δ. This follows by induction from the following two inequalities, see, e.g., [14]; we use here short notations $(\mu_i,\nu_i) = (\mu_0\nu_0)S_i^{Y;R;\delta}$.

(1°) For every i,

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

(2°) There exists $\pi_R < 1$ such that if $i \in J$ and $i+1 \in J$,

$$\rho\left((\mu_i,\nu_i)S_{i:i+1}^{Y;R;\delta}, (\nu_i,\mu_i)S_{i:i+1}^{Y;R;\delta}\right) \le \pi_R\,\rho\left((\mu_i,\nu_i),(\nu_i,\mu_i)\right).$$

The latter follows from the Proposition 3.9 from [14], with the contraction constant, $\pi_R \le (1-\tilde C_R^{-2})/(1+\tilde C_R^{-2})$, due to the "mixing condition"

$$\sup_{D_R} \frac{Q_{i:i+1}(x_0,\tilde x_0,dx',d\tilde x')}{Q_{i:i+1}(v_0,\tilde v_0,dx',d\tilde x')} \le \sup_{D_R} \frac{\sup_x \det\sigma^*\sigma(x)}{\inf_x \det\sigma^*\sigma(x)} \sup_{D_R} \left(\frac{q_\xi(\sigma^{-1}(x'-v_0-b(v_0))}{q_\xi(\sigma^{-1}(x'-x_0-b(x_0))}\right)^2 =: \tilde C_R < \infty, \tag{13}$$

with $D_R := \{(x_0,\tilde x_0,v_0,\tilde v_0,x',\tilde x') : |x_0|,|\tilde x_0|,|v_0|,|\tilde v_0|,|x'|,|\tilde x'| \le R\}$. and Then, the meaning of the inequality (2°) is that the replacement of non-random kernels Q by random ones $Q\Psi$ does not change the supremum of the derivative of one measure with respect to another.

For the completeness and reader's convenience, we remind the proof of the inequality (1°). If using the second version of the definition for the Birkhoff metric, it suffices to check, for every i, that firstly,

$$f^1_{i+1} := \sup \frac{d(\mu_{i+1},\nu_{i+1})}{d(\nu_{i+1},\mu_{i+1})} \le \sup \frac{d(\mu_i,\nu_i)}{d(\nu_i,\mu_i)} =: f^1_i,$$

and secondly,

$$f^2_{i+1} := \sup \frac{d(\nu_{i+1},\mu_{i+1})}{d(\mu_{i+1},\nu_{i+1})} \le \sup \frac{d(\nu_i,\mu_i)}{d(\mu_i,\nu_i)} =: f^2_i,$$

where $d(\mu,\nu)/d(\nu,\mu)$ means a Radon–Nikodym derivative. The two inequalities can be proved similarly, so we only check the first one. We estimate,

$$\begin{aligned} f^1_{i+1} &= \sup(\mu_{i+1},\nu_{i+1})(dx_{i+1},d\tilde x_{i+1})/(\nu_{i+1},\mu_{i+1})(dx_{i+1},d\tilde x_{i+1}) \\ &= \sup \frac{\int \mu_i(dx_i)\nu_i(d\tilde x_i)Q(x_i,dx_{i+1})Q(\tilde x_i,d\tilde x_{i+1})\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)}{\int \nu_i(dx_i)\mu_i(d\tilde x_i)Q(x_i,dx_{i+1})Q(\tilde x_i,d\tilde x_{i+1})\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)} \\ &\quad \text{(remind that both integrals are with respect to } dx_i\, d\tilde x_i \text{ only)} \\ &\le \sup f^1_i \frac{\int \nu_i(dx_i)\mu_i(d\tilde x_i)Q(x_i,dx_{i+1})Q(\tilde x_i,d\tilde x_{i+1})\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)}{\int \nu_i(dx_i)\mu_i(d\tilde x_i)Q(x_i,dx_{i+1})Q(\tilde x_i,d\tilde x_{i+1})\Psi(x_i,Y_i)\Psi(\tilde x_i,Y_i)} \\ &= f^1_i \times 1. \end{aligned}$$

In the last inequality, we have replaced $\mu_i(dx_i)\nu_i(d\tilde x_i)$ by $\nu_i(dx_i)\mu_i(d\tilde x_i)$ in the numerator, with the help of the bound on the derivative, f^1_i.

The induction base $k = 1$ (*not* $k = 0$) in (12) is valid due to the fact that after the first pair of ones, the measures become comparable, by virtue of (13); and the induction step follows from (2°) directly.

Now we can estimate as follows,

$$
\begin{aligned}
&E_{\mu_0,\nu_0}\|\mu_0\bar{S}_n^{Y,\mu_0} - \nu_0\bar{S}_n^{Y,\nu_0}\|_{TV} \\
&\quad \le \sum_{\delta\in\Delta;\, \#1(\delta)\ge 1} C_R\pi_R^{\#1(\delta)-1}E_{\mu_0,\nu_0}e_n^{Y;\delta;\mu_0,\nu_0} \\
&\qquad + 2\sum_{\delta\in\Delta;\, \#1(\delta)=0} Ee_n^{Y;\delta;\mu_0,\nu_0}\sup_D((\mu_0,\nu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D) \\
&\qquad -(\nu_0,\mu_0)\hat{S}_n^{Y;R;\delta;\mu_0,\nu_0}(D)) \\
&\quad \le \sum_{\delta\in\Delta} C_R(\pi_R^{\#1(\delta)-1}\wedge 1)E_{\mu_0,\nu_0}e_n^{Y;\delta;\mu_0,\nu_0}.
\end{aligned}
\tag{14}
$$

Hence, our main goal here is to estimate the expectation $E_{\mu_0,\nu_0}e_n^{Y;\delta;\mu_0,\nu_0}$. Notice that the constant C_R in (12) and (14) could be chosen independently of R and induction could start from $k=0$, if we assume in addition to (A3) that μ_0 and ν_0 are equivalent with a bounded derivative $\|d\nu_0/d\mu_0\|_{L_\infty(\mu_0)} < \infty$.

2. Let us split the sum $\sum_{\delta\in\Delta}$ into two parts, with $\#1(\delta)\ge \epsilon n$, and with $\#1(\delta) < \epsilon n$, where $\epsilon > 0$ is to be chosen. Whatever $0<\epsilon<1$, we have, for n large enough,

$$
\begin{aligned}
&\sum_{\delta:\#1(\delta)\ge\epsilon n} (\pi_R^{\#1(\delta)-1}\wedge 1)E_{\mu_0}e_n^{Y;\delta;\mu_0,\nu_0} \\
&\quad = \sum_{\delta:\#1(\delta)\ge\epsilon n} (\pi_R^{\#1(\delta)-1}\wedge 1)E_{\mu_0}E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y,\tilde{Y})\Big|_{\tilde{Y}=Y} \\
&\quad \le \pi_R^{\epsilon n-1}\sum_{\delta:\#1(\delta)\ge\epsilon n} E_{\mu_0}E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y,\tilde{Y})\Big|_{\tilde{Y}=Y} \\
&\quad = \pi_R^{\epsilon n-1}E_{\mu_0}\sum_{\delta:\#1(\delta)\ge\epsilon n} P_{\mu_0,\nu_0}\left(1_\delta(X,\tilde{X}) \mid Y,\tilde{Y}\right)\Big|_{\tilde{Y}=Y} \\
&\quad = \pi_R^{\epsilon n-1}E_{\mu_0}P_{\mu_0,\nu_0}\left(\bigcup_{\delta:\#1(\delta)\ge\epsilon n} 1_\delta(X,\tilde{X}) \mid Y,\tilde{Y}\right)\Big|_{\tilde{Y}=Y} \le \pi_R^{\epsilon n-1}.
\end{aligned}
\tag{15}
$$

The equality in the last line here is because the indicators $1_\delta(X,\tilde{X})$ with different δ's correspond to disjoint events, and due to linearity.

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

$$
\begin{aligned}
&\sum_{\delta:\#1(\delta)<\epsilon n} (\pi_R^{\#1(\delta)-1}\wedge 1)E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y,\tilde{Y})\Big|_{\tilde{Y}=Y}\right) \\
&\quad \le \sum_{\delta:\#1(\delta)<\epsilon n} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde{X}) \mid Y,\tilde{Y})\Big|_{\tilde{Y}=Y}\right).
\end{aligned}
\tag{16}
$$

We have,

$$
\begin{aligned}
&\sum_{\delta:\,\#1(\delta)<\epsilon n} E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X)\mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right)\\
&= E_{\mu_0}\left(\sum_{\delta:\,\#1(\delta)<\epsilon n} E_{\mu_0,\nu_0}(1_\delta(X,\tilde X)\mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right)\\
&= E_{\mu_0}\left(E_{\mu_0,\nu_0}(\sum_{\delta:\,\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X)\mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right).
\end{aligned}
\tag{17}
$$

Let us introduce some new indicators: let $0 < R$ be large enough, and

$$
\#1(X)_R := \sum_{k=0}^{n} 1(|X_k| \le R), \qquad \#0(X)_R = \sum_{k=0}^{n} 1(|X_k| > R).
$$

Let us show that by the Dirichlet principle,

$$
1\left(\#1(X)_R \ge 1 + \frac{3+\epsilon}{4}n, \#1(\tilde X)_R \ge 1 + \frac{3+\epsilon}{4}n\right) \sum_{\delta:\,\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X) = 0. \tag{18}
$$

Indeed, first of all, notice that

$$
\sum_{\delta:\,\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X) = 1\left(\sum_{i=1}^{n} 1(|X_{i-1}| \vee |\tilde X_{i-1}| \vee |X_i| \vee |\tilde X_i| \le R) < \epsilon n\right).
$$

If all $|X_k| \vee |\tilde X_k| \le R$, then $\sum_{i=1}^n 1(|X_{i-1}| \vee |\tilde X_{i-1}| \vee |X_i| \vee |\tilde X_i| \le R) = n$. Every "large" coordinate $|X_k| > R$ or $|\tilde X_k| > R$, $0 \le k \le n$, can reduce the sum $\sum_{i=1}^n 1(|X_{i-1}| \vee |\tilde X_{i-1}| \vee |X_i| \vee |\tilde X_i| \le R)$ at most by two. So, if this sum has a value less than ϵn, this means that at least $(1-\epsilon)n/2$ coordinates $|X_k|$ or $|\tilde X_k|$ are greater than $|R|$, i.e., $\#0(X)_R + \#0(\tilde X)_R > \frac{1-\epsilon}{2}n$; hence, either $\#1(X)_R < 1 + \frac{3+\epsilon}{4}n$, or $\#1(\tilde X)_R < 1 + \frac{3+\epsilon}{4}n$,—since otherwise $\#0(X)_R \le \frac{1-\epsilon}{4}n$ and $\#0(\tilde X)_R \le \frac{1-\epsilon}{4}n$. Thus, $\sum_{i=1}^n 1(|X_{i-1}| \vee |\tilde X_{i-1}| \vee |X_i| \vee |\tilde X_i| \le R) < \epsilon n$, implies $\#1(X) < 1 + (3+\epsilon)n/4$

or $\#1(\tilde X) < 1 + (3+\epsilon)n/4$, so, (18) holds true. Hence, we have,

$$
\begin{aligned}
& E_{\mu_0}\left(E_{\mu_0,\nu_0}\left(\sum_{\delta:\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&\le E_{\mu_0}\left(E_{\mu_0,\nu_0}\left(1\left(\#1(X)_R<\frac{3+\epsilon}{4}n\right)\sum_{\delta:\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&\quad +E_{\mu_0}\left(E_{\mu_0,\nu_0}\left(1\left(\#1(\tilde X)_R<\frac{3+\epsilon}{4}n\right)\sum_{\delta:\#1(\delta)<\epsilon n} 1_\delta(X,\tilde X)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&\le E_{\mu_0}\left(E_{\mu_0,\nu_0}\left(1\left(\#1(X)_R<\frac{3+\epsilon}{4}n\right)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&\quad +E_{\mu_0}\left(E_{\mu_0,\nu_0}\left(1\left(\#1(\tilde X)_R<\frac{3+\epsilon}{4}n\right)\mid Y,\tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&= E_{\mu_0}\left(E_{\mu_0}\left(1\left(\#1(X)_R<\frac{3+\epsilon}{4}n\right)\mid Y\right)\right)\\
&\quad +E_{\mu_0}\left(E_{\nu_0}\left(1\left(\#1(\tilde X)_R<\frac{3+\epsilon}{4}n\right)\mid \tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&\quad \text{(because } X \text{ does not depend on } \tilde Y\text{, nor } \tilde X \text{ depends on } Y).
\end{aligned}
$$

We estimate,

$$
E_{\mu_0}\left(E_{\mu_0}\left(1\left(\#1(X)_R<\frac{3+\epsilon}{4}n\right)\mid Y\right)\right) = E_{\mu_0}\left(1\left(\#1(X)_R<\frac{3+\epsilon}{4}n\right)\right). \qquad (19)
$$

Next, we estimate the other term, using the assumption (A3),

$$
\begin{aligned}
& E_{\mu_0}\left(E_{\nu_0}\left(1\left(\#1(\tilde X)_R<\frac{3+\epsilon}{4}n\right)\mid \tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&\le C_2\, E_{\nu_0}\left(E_{\nu_0}\left(1\left(\#1(\tilde X)_R<\frac{3+\epsilon}{4}n\right)\mid \tilde Y\right)\Big|_{\tilde Y=Y}\right)\\
&\text{(because } E_{\mu_0}F(X,Y)\le C_2E_{\nu_0}F(X,Y))\\
&= C_2\, E_{\nu_0}\left(E_{\nu_0}\left(1\left(\#1(\tilde X)_R<\frac{3+\epsilon}{4}n\right)\mid \tilde Y\right)\right)\\
&\text{(because } E_{\nu_0}F(X,Y)=E_{\nu_0}F(\tilde X,\tilde Y))\\
&= C_2\, E_{\nu_0}\left(1\left(\#1(\tilde X)_R<\frac{3+\epsilon}{4}n\right)\right),
\end{aligned}
$$

similarly to (19) where we had the equality.

3. Let $n\ge n_0$, n_0 be large enough (at least $n_0>4/(1-\epsilon)$), and $1>\epsilon'>(1/n_0)+(3+\epsilon)/4$. Due to the bounds which easily follow from [18] and [19], the latter expectation possesses an appropriate bound, exponential or polynomial, depending on the value p, if R is chosen large enough, namely,

$$E_{\mu_0} 1(\#1(X)_R < \epsilon' n) \le \begin{cases} C_m n^{-m}, & p = 1, \quad \forall\, m > 0, \\ C \exp(-cn), & p = 0. \end{cases} \tag{20}$$

This follows readily from the hitting time estimates for $\hat\tau = \inf(t \ge 0 : |X_t| \le R)$, see [18,19],

$$\begin{cases} E_x \hat\tau^k \le C_m(1 + |x|^m) & (\forall m > 2k) \quad (p = 1, \ \ \forall k > 0), \\ E_x \exp(\alpha\hat\tau) \le C \exp(c|x|) & (\exists C, c, \alpha > 0) \quad (p = 0), \end{cases} \tag{21}$$

due to the inequality $P_{\mu_0}(\#1(X)_R < \epsilon' n) \le P_{\mu_0}(\hat\tau_{\epsilon' n} > n)$,—where $\hat\tau_1 = \hat\tau$, and by induction $\hat\tau_{n+1} := \inf(t \ge \hat\tau_n + 1 : |X_t| \le R), \ n \ge 1$,—and due to exponential Chebyshev's inequality in the case $p = 0$, and by standard inequalities for semi-martingales in the case $p = 1$. By virtue of (21), it can be shown by induction, in each case:

- [$p = 0$] We have,

$$P_x(\hat\tau_{\epsilon' n} > n) \le \exp(-\alpha n + \epsilon' n \ln C + c|x|), \tag{22}$$

  where the value $C := \sup_{|x| \le R} E \exp(\alpha\hat\tau)$ can be done arbitrarily close to 1 by choosing large R. Hence, we get an exponential upper bound for $P_x(\hat\tau_{\epsilon' n} > n)$.
- [$p = 1$] Let $\epsilon' < \epsilon'' < 1$. We have,

$$P_x(\hat\tau_{\epsilon' n} > n) = P_x\left(\sum_{i=1}^{\epsilon' n} (\hat\tau_k - \hat\tau_{k-1}) > n\right) \le C_m(1 + |x|^m)((1 - \epsilon'')n)^{-k} n^{k/2}, \tag{23}$$

  with any $m > 2k$, if we choose R so large that $\epsilon' \sup_{|x'| \le R} E_{x'}((\hat\tau_k - \hat\tau_{k-1}) \mid X_{\hat\tau_{k-1}}) < \epsilon'' < 1$.

Both (22) and (23) have been proved in [18,19], correspondingly, so we do not repeat the calculus in either case. For convenience of reading, however, recall the idea of the proofs. In both cases due to the choice of ϵ', the mean value of $\hat\tau_k - \hat\tau_{k-1}$ is less that $(\epsilon')^{-1}$. So, in the case $p = 0$ one can apply exponential martingale inequalities, which lead to (22). In the case $p = 1$ one applies polynomial martingale inequalities that generalize the fact that the k-th moment of the sum of independent identically distributed random variables growth as $n^{k/2}$, and this is why we get the multiple $n^{k/2}$ in the right hand side of (23).

Hence, in both cases (20) holds true. Some more details concerning similar bounds can be found in the proof of the Theorem 2 below, see (27)–(28).

4. Hence, we get the estimate for the expression in (17),

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

for all $m>0$ in the case $p=1$. Combining this with the earlier inequalities (14) and (15), we obtain in both cases the final estimate (9), which in the case $p=0$ may have a new constant c in the exponential.

5. The non-averaged bounds follow from Chebyshev's inequality and the Borel–Cantelli lemmae. The Theorem 1 is proved.

4 The proof of Theorem 2

1. We continue the proof from the estimates (14)–(16). Remind that our main task remains to estimate the second term of the sum (14),

$$\sum_{\delta:\,\#1(\delta)<\epsilon n}(\pi_R^{\#1(\delta)-1}\wedge 1)E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X)\mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right)$$
$$\le\sum_{\delta:\,\#1(\delta)<\epsilon n}E_{\mu_0}\left(E_{\mu_0,\nu_0}(1_\delta(X,\tilde X)\mid Y,\tilde Y)\Big|_{\tilde Y=Y}\right).$$

Remind that now we cannot use the assumption (A3), and the idea to tackle this term is quite different. Remind that now the noise (ξ, V) is Gaussian. Let

$$\gamma=\exp\left(-\sum_{i\le n}h(X_i)Y_i+\frac12\sum_{i\le n}h^2(X_i)-\sum_{i\le n}h(\tilde X_i)\tilde Y_i+\frac12\sum_{i\le n}h^2(\tilde X_i)\right).$$

We will use one more version of the Bayes formula,

$$E_{\mu_0,\nu_0}(1_\delta(X,\tilde X)\mid Y,\tilde Y)=\frac{E^\gamma_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)\gamma^{-1}\mid Y,\tilde Y\right)}{E^\gamma_{\mu_0,\nu_0}(\gamma^{-1}\mid Y,\tilde Y)},\tag{24}$$

where for any measure P, we denote by P^γ the measure with a density $dP^\gamma/dP=\gamma$, and E^γ then denotes expectation with respect to this measure P^γ. For the reader's convenience, remind the proof. For any bounded measurable function $g=g(Y,\tilde Y)$ we can check the definition of the conditional probability as follows,

$$
\begin{aligned}
&E_{\mu_0,\nu_0} g \frac{E^\gamma_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)\gamma^{-1} \mid Y,\tilde Y\right)}{\left(E^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y,\tilde Y)\right)} \\
&= E^\gamma \gamma^{-1} g \frac{E^\gamma_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)\gamma^{-1} \mid Y,\tilde Y\right)}{\left(E^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y,\tilde Y)\right)} \\
&= E^\gamma_{\mu_0,\nu_0}\left(g \frac{E^\gamma_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)\gamma^{-1} \mid Y,\tilde Y\right)}{\left(E^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y,\tilde Y)\right)} \left(E^\gamma_{\mu_0,\nu_0}\left(\gamma^{-1}\right) \mid Y,\tilde Y\right)\right) \\
&= E^\gamma_{\mu_0,\nu_0}\left(g\, E^\gamma_{\mu_0,\nu_0}\left(1_\delta(X,\tilde X)\gamma^{-1} \mid Y,\tilde Y\right)\right) \\
&= E^\gamma_{\mu_0,\nu_0}\left(g\, 1_\delta(X,\tilde X)\gamma^{-1}\right) = E_{\mu_0,\nu_0}\left(g\, 1_\delta(X,\tilde X)\right).
\end{aligned}
$$

Let us emphasize that here we do not need to substitute $\tilde Y$ by Y, because the matter was simply to check the definition of conditional expectation.

Next, due to the Cauchy–Bouniakovsky–Schwarz inequality—known in older monographs either as the Cauchy–Bouniakovsky, or Cauchy–Schwarz', we follow here some recent suggestions of unification—for the conditional expectation we estimate the numerator in (24),

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

The denominator will be treated separately.

The further plan uses the following idea. Firstly, under $P^\gamma_{\mu_0,\nu_0}$, the couples of processes $(X,\tilde X)$ and $(Y,\tilde Y)$ are independent—a discrete version of Girsanov's Theorem—so that

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

which is a non-random value. The last equality is also due to Girsanov's Theorem; its sense is that the density γ does not change the distribution of the non-observable process. Secondly, we will show that the *expectation* of the second factor divided by the denominator, with respect to P_{μ_0,ν_0}, or, equivalently, with respect to P_{μ_0},

$$
E_{\mu_0} \frac{\left(E^\gamma_{\mu_0,\nu_0}\left(\gamma^{-2} \mid Y,\tilde Y\right)\Big|_{\tilde Y = Y}\right)^{1/2}}{E^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y,\tilde Y)\Big|_{\tilde Y = Y}} = E^\gamma_{\mu_0}\gamma^{-1} \frac{\left(E^\gamma_{\mu_0,\nu_0}\left(\gamma^{-2} \mid Y,\tilde Y\right)\Big|_{\tilde Y = Y}\right)^{1/2}}{E^\gamma_{\mu_0,\nu_0}(\gamma^{-1} \mid Y,\tilde Y)\Big|_{\tilde Y = Y}}, \quad (25)
$$

does not exceed *some* exponential $\exp(Cn)$, with a constant C that depends only on $\|h\|_{L_\infty}$. Finally, as we will see shortly, the expression $E_{\mu_0,\nu_0} 1_\delta(X,\tilde X)$ can be made smaller that $\exp(-Cn)$ with *any* C under an appropriate choice of R, if $\#1(\delta) < \epsilon n$. Hence, we will get an exponential bound for the second part of the sum (14), too. Let us start this programme.

2. Denominator in (25). We are going to estimate it from below. We have,

$$
\begin{aligned}
&E^{\gamma}_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde{Y})\Big|_{\tilde{Y}=Y} = E^{\gamma}_{\mu_0,\nu_0} \\
&\quad \times \left(\exp\left(\sum_{i\le n} h(X_i)Y_i - \frac{1}{2}\sum_{i\le n} h(X_i)^2 + \sum_{i\le n} \tilde{h}(\tilde{X}_i)\tilde{Y}_i - \frac{1}{2}\sum_{i\le n} h(\tilde{X}_i)^2\right) \mid Y, \tilde{Y}\right)\Big|_{\tilde{Y}=Y} \\
&\ge e^{-Cn} E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(\sum_{i\le n} h(X_i)Y_i + \sum_{i\le n} h(\tilde{X}_i)\tilde{Y}_i \mid Y, \tilde{Y}\right)\right)\Big|_{\tilde{Y}=Y} \\
&= e^{-Cn} E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(\sum_{i\le n} h(X_i)Y_i\right) \mid Y\right) E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(\sum_{i\le n} h(\tilde{X}_i)\tilde{Y}_i\right) \mid \tilde{Y}\right) \\
&\ge e^{-Cn}\left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-\sum_{i\le n} h(X_i)Y_i\right) \mid Y\right)\right)^{-1} \\
&\quad \times \left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-\sum_{i\le n} h(\tilde{X}_i)\tilde{Y}_i\right) \mid \tilde{Y}\right)\right)^{-1}\Big|_{\tilde{Y}=Y}.
\end{aligned}
$$

In other words,

$$
\begin{aligned}
\left(E^{\gamma}_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde{Y})\Big|_{\tilde{Y}=Y}\right)^{-1} &\le e^{+Cn}\left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-\sum_{i\le n} h(X_i)Y_i\right) \mid Y\right)\right) \\
&\quad \times \left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-\sum_{i\le n} h(\tilde{X}_i)\tilde{Y}_i\right) \mid \tilde{Y}\right)\right)\Big|_{\tilde{Y}=Y}.
\end{aligned}
$$

The first conditional expectation in the latter bound suits well our further applications of the Bouniakovsky–Cauchy–Schwarz inequality. The second one suits it as well, because it can be re-written as

$$
\begin{aligned}
&\left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-\sum_{i\le n} h(\tilde{X}_i)\tilde{Y}_i\right) \mid \tilde{Y}\right)\right)\Big|_{\tilde{Y}=Y} \\
&\quad = \left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-\sum_{i\le n} h(\tilde{X}_i)Y_i\right) \mid Y\right)\right),
\end{aligned}
$$

since $\tilde{X}$ and $\tilde{Y}$ are independent under P^{γ}.

Further, with $p > 1,\ r = p^2$—we will use $p = 4$—

$$\left(E^{\gamma}_{\mu_0,\nu_0}\left(E^{\gamma}_{\mu_0,\nu_0}(\exp\left(-\sum_{i\le n} h(\tilde X_i)Y_i\right) \mid Y)\right)^p\right)^{1/p}$$

$$\le \left(E^{\gamma}_{\mu_0,\nu_0}\left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-p\sum_{i\le n} h(\tilde X_i)\,Y_i\right) \mid Y\right)\right)\right)^{1/p}$$

$$= \left(E^{\gamma}_{\mu_0,\nu_0}\exp\left(-p\sum_{i\le n} h(\tilde X_i)Y_i\right)\right)^{1/p}$$

$$= \left(E^{\gamma}_{\mu_0,\nu_0}\exp\left(-p\sum_{i\le n} h(\tilde X_i)Y_i - r\sum_{i\le n} h(\tilde X_i)^2 + r\sum_{i\le n} h(\tilde X_i)^2\right)\right)^{1/p}$$

$$\le \left(E^{\gamma}_{\mu_0,\nu_0}\exp\left(-2p\sum_{i\le n} h(\tilde X_i)Y_i - 2r\sum_{i\le n} h(\tilde X_i)^2_i\right)\right)^{1/2p}$$

$$\times\left(E^{\gamma}\exp(2r\sum_{i\le n} h(\tilde X_i)^2_i)\right)^{1/2p} = \left(E^{\gamma}_{\mu_0,\nu_0}\exp\left(-2p\sum_{i\le n} h(\tilde X_i)Y_i - \frac{4p^2}{2}\right.\right.$$

$$\left.\left.\times\sum_{i\le n} h(\tilde X_i)^2_i\right)\right)^{1/2p}\left(E^{\gamma}\exp(2r\sum_{i\le n} h(\tilde X_i)^2_i)\right)^{1/2p} \le \left(e^{Cn}\right)^{1/2p} = e^{Cn},$$

with a generic constant $C > 0$, which depends only on $\|h\|_{L_\infty}$

Similarly, and, in fact, even simpler, we estimate the term

$$\left(E^{\gamma}\left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-\sum_{i\le n} h(X_i)Y_i\right) \mid Y\right)\right)^p\right)^{1/p}$$

$$\le \left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(-p\sum_{i\le n} h(X_i)Y_i\right)\right)\right)^{1/p} \le e^{Cn}.$$

3. Numerator in (25). Similarly—and also simpler—we treat the numerator. We are to estimate it from above. We have,

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

$$= E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(2\sum_{i\le n} h(X_i)Y_i - \sum_{i\le n} h(X_i)^2 + 2\sum_{i\le n} h(\tilde X_i)\tilde Y_i - \sum_{i\le n} h(\tilde X_i)^2\right) \mid Y, \tilde Y\right)\Big|_{\tilde Y = Y}$$

$$
\begin{aligned}
&\le e^{+Cn} E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(2\sum_{i\le n} h(X_i)Y_i + 2\sum_{i\le n} h(\tilde X_i)\tilde Y_i\right) \mid Y, \tilde Y\right)\Big|_{\tilde Y = Y}\\
&= e^{+Cn} E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(2\sum_{i\le n} h_i Y_i\right) \mid Y\right) E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(\sum_{i\le n} h(\tilde X_i)\tilde Y_i\right) \mid \tilde Y\right)\\
&\le e^{+Cn}\left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(2\sum_{i\le n} h(X_i)Y_i\right) \mid Y\right)\right)\\
&\quad \times \left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(2\sum_{i\le n} h(\tilde X_i)\tilde Y_i\right) \mid \tilde Y\right)\right)\Big|_{\tilde Y = Y}.
\end{aligned}
$$

In other words,

$$
\begin{aligned}
\left(E^{\gamma}_{\mu_0,\nu_0}(\gamma^{-2} \mid Y, \tilde Y)\Big|_{\tilde Y = Y}\right)^{1/2} &\le e^{+Cn}\left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(2\sum_{i\le n} h(X_i)Y_i\right) \mid Y\right)\right)^{1/2}\\
&\quad \times \left(E^{\gamma}_{\mu_0,\nu_0}\left(\exp\left(2\sum_{i\le n} h(\tilde X_i)\tilde Y_i\right) \mid \tilde Y\right)\right)^{1/2}\Bigg|_{\tilde Y = Y}.
\end{aligned}
$$

The rest is similar to the calculus made for the denominator. Whence, due to Hölder's inequality,

$$
\begin{aligned}
&E^{\gamma}_{\mu_0}\gamma^{-1}\,\frac{\left(E^{\gamma}_{\mu_0,\nu_0}\left(\gamma^{-2} \mid Y, \tilde Y\right)\Big|_{\tilde Y = Y}\right)^{1/2}}{E^{\gamma}_{\mu_0,\nu_0}(\gamma^{-1} \mid Y, \tilde Y)\Big|_{\tilde Y = Y}}\\
&\le \left(E^{\gamma}\gamma^{-4}\right)^{1/4} e^{Cn/2}\left(E^{\gamma}\exp\left(4\sum_{i\le n} h(X_i)Y_i\right)\right)^{1/4}\left(E^{\gamma}\exp\left(4\sum_{i\le n} h(\tilde X_i)Y_i\right)\right)^{1/4}\\
&\quad \times\left(E^{\gamma}\exp\left(-2\sum_{i\le n} h(X_i)Y_i\right)\right)^{1/4}\left(E^{\gamma}\exp\left(-2\sum_{i\le n} h(\tilde X_i)Y_i\right)\right)^{1/4} \le \exp(Cn),
\end{aligned}
\tag{26}
$$

with $C > 0$ which depends only on $\|h\|_{L_\infty}$, because all factors allow this upper bound.

4. The term $\sum_{\delta:\,\#1(\delta)<\epsilon n} E_{\mu_0,\nu_0} 1_\delta(X, \tilde X) = E_{\mu_0,\nu_0} \sum_{\delta:\,\#1(\delta)<\epsilon n} 1_\delta(Z)$. This is now the last and the main issue; emphasize that we are looking for a bound of this *non-conditional* probability, which, for every $0 < \epsilon < 1$, should be less than any exponential, if the value R is chosen large enough. Intuitively, this looks reasonable under our assumption (A1), which is sufficient for an exponential mixing; however, we will not use any mixing bounds directly, and we suppose that it would not be easy to use them anyway. We will use the hint from the proof of the previous theorem, reducing the problem to estimates for X and $\tilde{X}$ separately.

We have,

$$E_{\mu_0,\nu_0} \sum_{\delta:\#1(\delta)<\epsilon n} 1_\delta(Z)$$
$$= P_{\mu_0,\nu_0}\left(\sum_{i=1}^{n} 1(|X_{i-1}| \le R, |\tilde{X}_{i-1}| \le R, |X_i| \le R, |\tilde{X}_i| \le R) < \epsilon n\right).$$

Hence, it suffices to show—see (18)—that for every $c > 0$, $0 < \epsilon < 1$,

$$P_{\mu_0}\left(\sum_{i=0}^{n} 1(|X_i| \le R) < n\epsilon\right) \le e^{-Cn}, \tag{27}$$

and

$$P_{\nu_0}\left(\sum_{i=1}^{n} 1(|\tilde{X}_i| \le R) < n\epsilon\right) \le e^{-Cn}, \tag{28}$$

all if R is large enough. Both inequalities can be proved similarly, we will establish the first one. Notice that the bounds are similar to (20) in the proof of the Theorem 1 above. The reason why we give more details here is that there no appropriate reference to exponential inequalities with an *arbitrary* constant.

Given the recurrent Markov process X_k, $k \ge 0$, let us construct a dominating one-dimensional random walk, ζ_k, on $Z_+ = \{0, 1, \ldots\}$. The idea is that for every *trajectory* of X, this random walk ζ will dominate X at all times, more precisely, $\zeta_k \ge (|X_k| - R)_+$, and because of its simple structure, certain exponential bounds can be easily computed for ζ. An idea of enlargement of the probability space will be used, although we will not change notation for the probability measure.

For any time k, let

$$x_k := -[-(|X_k| - R)_+],$$

or, by words, x_k is the least integer which is not less than $(|X_k| - R)_+$; of course, $x_k \ge 0$. Then, by induction, define

$$\tilde{x}_0 := x_0; \quad \tilde{x}_{k+1} - \tilde{x}_k := \begin{cases} x_{k+1} - x_k, & x_{k+1} - x_k \ge -1, \\ x_k - 1, & x_{k+1} - x_k < -1. \end{cases}$$

Generally speaking, the process $\tilde{x}_k$ is not Markov, although the pair $(X_k, \tilde{x}_k)$ is; the jumps of $\tilde{x}$ are $-1, 0, 1, \ldots$; and its jumps are not less than the jumps of the process x_k, therefore, $\tilde{x}_k \ge x_k$. Since $x_k \ge (|X_k| - R)_+$, we have also,

$$\tilde{x}_k \ge (|X_k| - R)_+.$$

In particular, $\tilde{x}_k = 0$ implies $(|X_k| - R)_+ = 0$, that is, $|X_k| \le R$.

Since the process $(\tilde{x}_k)$ is not Markov, we will "spoil" it further to get a simpler Markov dominating random walk. Define the non-random values,

$$q_{R,j} := \sup_{X_0} P(\tilde{x}_1 = \tilde{x}_0 + j \mid X_0), \qquad j \ge 1,$$

and

$$q^{(+)}_{R,0} := \sup_{X_0 : x_0 > 0} P(\tilde{x}_1 = \tilde{x}_0 \mid X_0),$$

and

$$q^{(0)}_{R,0} := 1 - \sum_{j \ge 1} q_{R,j},$$

and, finally,

$$q^{(+)}_{R,-1} := 1 - \sum_{j \ge 1} q_{R,j} - q^{(+)}_{R,0}.$$

Due to the assumptions, there exist $C > 0$ and $\bar{q}_R \to 0, \ \ R \to \infty$, such that for every $j \ge 1$,

$$q_{R,j} \le C \bar{q}_R^j, \tag{29}$$

with

$$\lim_{R \to \infty} \bar{q}_R = 0. \tag{30}$$

Moreover,

$$\lim_{R \to \infty} \sup_{X_0 : x_0 > 0} q^{(+)}_{R,0}(X_0) = 0. \tag{31}$$

Indeed, consider firstly the case $\underline{j \ge 1}$. Notice that $\tilde{x}_{k+1} - \tilde{x}_k = j$ is equivalent to $x_{k+1} - x_k = j$, which, by definition, is equivalent to

$$[-(|X_k| - R)_+] - [-(|X_{k+1}| - R)_+] = j.$$

It follows,

$$0 \le j - 1 < (|X_{k+1}| - R)_+ - (|X_k| - R)_+ < j + 1.$$

This implies,

$$|X_{k+1}| > |X_k| + j - 1.$$

Indeed, if $|X_k| \le R$, then $0 \le j-1 < (|X_{k+1}| - R)_+ - (|X_k| - R)_+$ means simply $0 \le j-1 < (|X_{k+1}| - R)_+$, or, equivalently, $0 \le j-1 < |X_{k+1}| - R$, or $|X_{k+1}| > R+j-1 \ge |X_k|+j-1$. If $|X_k| > R$ and $(|X_{k+1}|-R)_+-(|X_k|-R)_+ \ge 0$, we have, $(|X_{k+1}|-R)-(|X_k|-R) > j-1$, or $|X_{k+1}|-|X_k| > j-1$, as required. Thus,

$$\tilde{x}_{k+1} - \tilde{x}_k = j \ge 1 \quad \text{implies} \quad |X_{k+1}| \ge R, \quad \& \quad |X_{k+1}| - |X_k| > \tilde{x}_{k+1} - \tilde{x}_k - 1.$$

Therefore,

$$\begin{aligned}\sup_{\tilde{x}_0} P_{\tilde{x}_0}(\tilde{x}_1 = \tilde{x}_0 + j) &\le \sup_{\tilde{x}_0} P_{\tilde{x}_0}(\tilde{x}_1 - \tilde{x}_0 \ge j) \\ &\le \sup_{X_0} P(|X_1| - |X_0| > j-1;\ |X_1| \ge R \mid X_0).\end{aligned}$$

We will consider separately

$$\sup_{X_0 \colon |X_0| \le R/4} P(|X_1| - |X_0| > j-1;\ |X_1| \ge R \mid X_0)$$

and

$$\sup_{X_0 \colon |X_0| > R/4} P(|X_1| - |X_0| > j-1;\ |X_1| \ge R \mid X_0).$$

We estimate,

$$\begin{aligned}\sup_{|X_0| \le R/4} P(|X_1| - |X_0| > j-1;\ |X_1| \ge R \mid X_0) &\le \sup_{|X_0| \le R/4} P\,(|X_1| \\ \ge \frac{j-1}{2} + \frac{R}{2} \mid X_0\Big) &\le \sup_{|X_0| \le R/4} P\left(|X_0 + b(X_0)| + c|\xi_1| \ge \frac{j-1}{2} + \frac{R}{2} \mid X_0\right).\end{aligned}$$

Here we split the supremum into two,

$$\sup_{|X_0| \le R/4} \le \sup_{|X_0| \le R_0} + \sup_{R_0 < |X_0| \le R/4},$$

where R_0 is chosen so that

$$\sup_{|x| \ge R_0} (|x + b(x)| - |x|) \le 0.$$

Moreover, denote

$$C_0 := \sup_{|x|\le R_0} (|x+b(x)| - |x|) < \infty.$$

Remind that b is locally bounded. We have, with $R_0 < R/4$ and any $\lambda > 0$,

$$\begin{aligned}
&\sup_{|X_0|\le R_0} P\left(|X_0+b(X_0)| + c|\xi_1| \ge \frac{j-1}{2} + \frac{R}{2} \mid X_0\right)\\
&\quad\le \sup_{|X_0|\le R_0} P\left(|X_0+b(X_0)| - |X_0| + c|\xi_1| \ge \frac{j-1}{2} + \frac{R}{2} - R_0 \mid X_0\right)\\
&\quad\le \sup_{|X_0|\le R_0} P\left(|X_0+b(X_0)| - |X_0| + c|\xi_1| \ge \frac{j-1}{2} + \frac{R}{2} - \frac{R}{4} \mid X_0\right)\\
&\quad\le \sup_{|X_0|\le R_0} P\left(|X_0+b(X_0)| - |X_0| + c|\xi_1| \ge \frac{R}{4} + \frac{j-1}{2} \mid X_0\right)\\
&\quad\le P\left(c|\xi_1| \ge \frac{R}{4} + \frac{j-1}{2} - C_0\right) \le \exp\left(-\lambda\left(\frac{R}{4} + \frac{j-1}{2} - C_0\right)\right) Ee^{\lambda c|\xi_1|}.
\end{aligned}$$

Here we can take, for example, $\lambda = 1$. Further, estimate

$$\begin{aligned}
&\sup_{R_0<|X_0|\le R/4} P\left(|X_0+b(X_0)| + c|\xi_1| \ge \frac{j-1}{2} + \frac{R}{2} \mid X_0\right)\\
&\quad\le \sup_{R_0<|X_0|\le R/4} P\left(|X_0+b(X_0)| - |X_0| + c|\xi_1| \ge \frac{j-1}{2} + \frac{R}{2} - \frac{R}{4} \mid X_0\right)\\
&\quad\le P\left(c|\xi_1| \ge \frac{j-1}{2} + \frac{R}{4} - C_0\right) \le \exp\left(-\lambda(\frac{j-1}{2} + \frac{R}{4} - C_0)\right) Ee^{\lambda c|\xi_1|}.
\end{aligned}$$

Finally, due to (A1), with $\alpha(R) := -\sup_{|x|\ge R}(|x+b(x)| - |x|)$, we have, $0 \le \alpha(R) \to \infty,\ R \to \infty$, and

$$\begin{aligned}
&\sup_{X_0:\,|X_0|>R/4} P(|X_1| - |X_0| > j-1;\ |X_1| \ge R \mid X_0)\\
&\quad\le \sup_{|X_0|>R/4} P(|X_0+b(X_0)| - |X_0| + c|\xi_1| > j-1 \mid X_0)\\
&\quad\le P(c|\xi_1| > j-1+\alpha(R/4) \mid X_0) \le \exp(-\lambda(j-1+\alpha(R/4)))Ee^{c|\xi_1|}.
\end{aligned}$$

E.g., one can take $\lambda = 1$, to satisfy (29).

Consider the case $\underline{j = 0\ \&\ x_k > 0}$. The same calculus as above, gives a similar estimate. Indeed, $x_k > 0$ implies $0 < |X_k| - R \le x_k$. Hence, we have, with any positive λ in the last line,

$$\begin{aligned}
&\sup_{X_k:\,|X_k|>R} P(|X_{k+1}-|X_k|>-1;\,|X_k|\ge R\mid X_k)\\
&\le \sup_{|X_0|>R} P(|X_0+b(X_0)|+c|\xi_1|-|X_0|>-1\mid X_0)\\
&\le P(c|\xi_1|>-1+\alpha(R)\mid X_0)\le \exp(-\lambda(-1+\alpha(R)))\,Ee^{\lambda c|\xi_1|}.
\end{aligned}$$

Again, one can take $\lambda = 1$ here.

5. Our next issue is to construct a dominating *Markov* process for $\tilde{x}_k$, at last. All we need for this is to adjust the transition probabilities, so that they would not depend on X_k. For this aim, we will increase slightly all probabilities of jumps up, as well as zero if the state is positive, still keeping them small enough. For convenience of reading, we provide the construction below. Let

$$\zeta_0 = \tilde{x}_0.$$

Further, for $j \ge 0$ define

$$\zeta_{k+1} = \zeta_k + j, \quad \text{on the set} \quad (\tilde{x}_{k+1} = \tilde{x}_k + j) \cap (\tilde{x}_k > 0).$$

On some part of the set $(x_k > 0) \cap (\tilde{x}_{k+1} = \tilde{x}_k - 1)$, we will allow $\zeta_{k+1} = \zeta_k - 1$, but on some other part of this set the increment $\zeta_{k+1} - \zeta_k$ may be non-negative: 0, 1, 2, ... The following paragraph explains the details.

Consider random variables

$$\Delta q_{R,j}(X_k) := q_{R,j} - P(\tilde{x}_{k+1} = \tilde{x}_k + j \mid X_k),$$

and, in the case $\tilde{x}_k > 0$, define

$$\zeta_{k+1} = \zeta_k + j,\ \ j \ge 0, \quad \text{on the set} \quad (\tilde{x}_{k+1} = \tilde{x}_k - 1) \cap (a_{j-1} < \chi_{k+1} \le a_j),$$

where χ_{k+1} is a new independent random variable which is uniformly distributed on [0, 1], and the non-random values a_j are defined by the equalities,

$$a_j = \sum_{0\le i\le j} \Delta q_{R,i}(X_k), \quad a_{-1} = 0.$$

On the complementary set,

$$(\tilde{x}_{k+1} = \tilde{x}_k - 1) \cap \left(\chi_{k+1} \ge \lim_{j\to\infty} a_j \equiv \sum_{0\le i<\infty} \Delta q_{R,i}(X_k)\right),$$

we define

$$\zeta_{k+1} = \zeta_k - 1.$$

If $\tilde{x}_k = 0$, we define, in a similar manner,

$$\zeta_{k+1} = \zeta_k + j, \quad \text{if} \quad \tilde{x}_{k+1} = \tilde{x}_k + j, \; j > 0,$$

and,

$$\zeta_{k+1} = \zeta_k + j, \; j > 0, \quad \text{on the set} \quad (\tilde{x}_{k+1} = \tilde{x}_k) \cap (a_{j-1} < \chi_{k+1} \le a_j).$$

Further, let

$$\zeta_{k+1} = \zeta_k - 1,$$

on the set

$$(\zeta_k > 0) \cap (x_{k+1} - x_k = -1) \cap \left(\chi_{k+1} \ge \sum_{0 \le i < \infty} \Delta q_{R,i}(X_k) \right),$$

and

$$\zeta_{k+1} = \zeta_k,$$

on the set

$$(\zeta_k = 0) \cap (x_{k+1} - x_k = 0) \cap \left(\chi_{k+1} \ge \sum_{0 \le i < \infty} \Delta q_{R,i}(X_k) \right).$$

Notice that all transition probabilities for the process ζ_k are non-random, given the current state ζ_k at time k; moreover, they clearly do not depend on time k either. Indeed, we have,

$$\begin{aligned}
&P(\zeta_{k+1} - \zeta_k = j \mid \zeta_k) = q_{R,j}, \quad j \ge 1, \; \zeta_k > 0;\\
&P(\zeta_{k+1} - \zeta_k = 0 \mid \zeta_k) = q^{(+)}_{R,0}, \quad \zeta_k > 0;\\
&P(\zeta_{k+1} - \zeta_k = -1 \mid \zeta_k) = q^{(+)}_{R,-1}, \quad \zeta_k > 0;\\
&P(\zeta_{k+1} - \zeta_k = j \mid \zeta_k = 0) = q_{R,j}, \quad j \ge 1;\\
&P(\zeta_{k+1} - \zeta_k = 0 \mid \zeta_k = 0) = q^{(0)}_{R,0}.
\end{aligned}$$

All these equalities follow directly from the construction. For example, for $j \ge 1$, $\zeta > 0$,

$$P(\zeta_{k+1} - \zeta_k = j \mid \zeta_k) = q_{R,j},$$

as required.

Hence, $(\zeta_k, k \geq 0)$ is a Markov chain with the state space $\{0, 1, 2, \ldots\}$, and the following matrix of transition probabilities, with the usual notation $p_{ij} = P(\zeta_{k+1} = j \mid \zeta_k = i)$,

$$
\left(p_{ij}\right)_{0\leq i,j<\infty} = \begin{pmatrix}
q^{(0)}_{R,0} & q_{R,1} & q_{R,2} & q_{R,3} & q_{R,4} & \ldots \\
q^{(+)}_{R,-1} & q^{(+)}_{R,0} & q_{R,1} & q_{R,2} & q_{R,4} & \ldots \\
0 & q^{(+)}_{R,-1} & q^{(+)}_{R,0} & q_{R,1} & q_{R,2} & \ldots \\
0 & 0 & q^{(+)}_{R,-1} & q^{(+)}_{R,0} & q_{R,1} & \ldots \\
0 & 0 & 0 & q^{(+)}_{R,-1} & q^{(+)}_{R,0} & \ldots \\
0 & 0 & 0 & 0 & q^{(+)}_{R,-1} & \ldots \\
\ldots\ldots & & & & &
\end{pmatrix}
$$

Moreover, remind that the inequalities (29)–(31) hold true.

Next, by the construction,

$$\zeta_k \geq \tilde{x}_k,$$

because all jumps of ζ_k are greater than or equal to the jumps of $\tilde{x}_k$ *at any time* k, and, of course, $\zeta_0 = \tilde{x}_0$. The reason why we need this process is that for this simple random walk one can use explicit formulae for some asymptotics, as shown below.

Notice that $\zeta_k = 0$ implies $|X_k| \leq R$. Therefore, it is sufficient to show that for any $\epsilon > 0$,

$$P\left(\sum_{k=0}^{n} 1(\zeta_k > 0) \geq \epsilon n\right) \leq e^{-Cn}, \tag{32}$$

In turn, (32) follows from the inequality,

$$P(\bar{\tau} \geq \epsilon n) \leq e^{-Cn}, \tag{33}$$

with any $C > 0$ if $R > 0$ is large enough, where

$$\bar{\tau} = \sum_{j=0}^{n} \tau_j,$$

and

$$\tau_0 := \inf(k \geq 1 :\ \zeta_{k-1} > 0,\ \zeta_k = 0) - k_0,$$

where

$$k_0 := \inf(k \geq 0 :\ \zeta_k > 0),$$

and $\tau_j,\ j > 0$, are defined similarly by induction,

$$\tau_j := \inf(k > \tau_{j-1} : \zeta_{k-1} > 0,\ \zeta_k = 0) - k_j, \text{ where } k_j := \inf(\ell \geq \tau_{j-1} : \zeta_\ell > 0).$$

Naturally, all τ_j are independent and identically distributed, except, perhaps, τ_1 if $\zeta_0 > 0$; the latter is independent from the others, too, but may have a different distribution. We have,

$$P(\bar\tau \geq \epsilon n) \leq e^{-\lambda\epsilon n}\, Ee^{\lambda\tau_1} \left(Ee^{\lambda\tau_2}\right)^n = \exp\left(-n\epsilon\lambda + \ln Ee^{\lambda\tau_1} + n\ln Ee^{\lambda\tau_2})\right).$$

Due to (29)–(31), for every fixed $\lambda > 0$, we can choose R so large that both values $\ln Ee^{\lambda\tau_1}$ and $\ln Ee^{\lambda\tau_2}$ are arbitrarily close to 1. Indeed, let

$$T := \inf(k \geq 0 : \zeta_k = 0).$$

The standard Lyapunov function approach—e.g., as presented in [8]—applied to the process $\exp(\lambda\zeta_k + \lambda_1 k)$ with *any* $\lambda > \lambda_1 > 0$, readily shows the following:

$$v(\zeta_0) := E_{\zeta_0}\exp(\lambda_1 T) \leq C\exp(\lambda\zeta_0), \quad \text{as} \quad \bar q_R \to 0,$$

with any fixed $C > 1$. So,

$$\begin{aligned} Ee^{\lambda\tau_2} &\leq 1 + \sum_{j\geq 1} C\bar q^j v(j) \leq 1 + C\sum_{j\geq 1}\bar q^j\exp(\lambda j) \\ &\leq 1 + \bar q_R\exp(\lambda)\frac{1}{1-\bar q_R\exp(\lambda)} \approx 1. \end{aligned}$$

Similarly,

$$\begin{aligned} Ee^{\lambda\tau_1} &\leq 1 + \sum_{j\geq 1} P(\zeta_0 = j)v(j) \leq 1 + C\sum_{j\geq 1}\tilde q^j\exp(\lambda j) \\ &\leq 1 + \tilde q e^{\lambda}\frac{1}{1-\tilde q\exp(\lambda)} \approx 1, \end{aligned}$$

because due to the moment assumption from (A1),

$$\begin{aligned} P(\zeta_0 = j) &\leq P_{\mu_0}(|X_0| > R + j - 1) \\ &\leq e^{-(R-1+j)}E_{\mu_0}e^{|X_0|} = \left(e^{-R+1}\right)^j E_{\mu_0}e^{|X_0|} \leq C\tilde q^j \end{aligned}$$

with an *arbitrary* small $\tilde q > 0$, if R is large enough. Notice that it would be enough to use just the fact that the value $Ee^{\lambda\tau_1}$ is finite; however, since λ may be arbitrarily large, anyway we need the moment assumption which implies $Ee^{\lambda\tau_1} \approx 1$ for R large enough. Hence, one can choose any large $\lambda > 0$, and then for this fixed λ we choose R large enough, so that, eventually, with some non-random n_0,

$$P(\bar{\tau} \geq \epsilon n) \leq \exp(-\lambda n/2), \quad n \geq n_0.$$

Thus, the inequality (33) is satisfied, which finally implies,

$$\sum_{\#0(\delta)\geq\epsilon n} E_{\mu_0,\nu_0} 1_\delta(X) \leq \exp(-\lambda n/4), \quad n \geq n_0,$$

with any large λ, if R is large enough. So, we get (27) and similarly (28). Now, combining these bounds with (14) and (15), we get the desired inequality (10).
6. The pointwise bound follows immediately from Chebyshev's inequality and the Borel–Cantelli lemmae. The Theorem 2 is proved.

Acknowledgements The authors are grateful to the anonymous referees for useful comments and remarks. The second author thanks the Department of Mathematics of the University of Maine, Le Mans, and the grant RFBR 05-01-00449 for support.

References

1. Atar, R.: Exponential stability for nonlinear filtering of diffusion processes in a noncompact domain. Ann. Probab. **26**(4), 1552–1574 (1998)
2. Atar, R., Zeitouni, O.: Exponential stability for nonlinear filtering. Ann. Inst. H. Poincaré Probab. Statist. **33**(6), 697–725 (1997)
3. Baxendale, P., Chigansky, P., Liptser, R.: Asymptotic stability of the Wonham filter: ergodic and nonergodic signals. SIAM J. Control Optim. **43**(2), 643–669 (2004) (electronic)
4. Budhiraja, A.S., Ocone, D.L.: Exponential stability in discrete-time filtering for bounded observation noise. Syst. Control Lett. **30**, 185–193 (1997)
5. Budhiraja, A.S., Ocone, D.L.: Exponential stability for discrete-time filtering for nonergodic signals. Stoch. Process. Appl. **82**, 245–257 (1999)
6. Chigansky, P., Liptser, R.: Stability of nonlinear filters in nonmixing case. Ann. Appl. Probab. **14**(4), 2038–2056 (2004)
7. Del Moral, P., Guionnet, A.: On the stability of interacting processes with applications to filtering and genetic algorithms. Ann. Inst. H. Poincaré **37**(2), 155–194 (2001)
8. Gulinsky, O.V., Veretennikov, A.Yu.: Large deviations for discrete-time processes with averaging. VSP, Utrecht (1993)
9. Kleptsyna, M.L., Veretennikov, A.Yu.: On ergodic filters with wrong initial data. Comptes Rendus Acad. Sci. Paris Ser. Math. **344**(11), 727–731 (2007) http://dx.doi.org/10.1016/j.crma.2007.04.015
10. Kleptsyna, M.L., Veretennikov, A.Yu.: On continuous time ergodic filters with wrong initial data, submitted; the preprint version at http://www.univ-lemans.fr/sciences/statist/download/Kleptsyna/filt36c.pdf
11. Krasnosel'skii, M.A., Lifshits, E.A., Sobolev, A.V.: Positive Linear Systems. The Method of Positive Operators. Heldermann Verlag, Berlin (1989)
12. Kunita, H.: Asymptotic behavior of the nonlinear filtering errors of Markov processes. J. Multivar. Anal. **1**, 365–393 (1971)
13. Le Gland, F., Oudjane, N.: A robustification approach to stability and to uniform particle approximation of nonlinear filters: the example of pseudo-mixing signals. Stoch. Process. Appl. **106**, 279–316 (2003)
14. Le Gland, F., Oudjane, N.L.: Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters. Ann. Appl. Probab. **14**(1), 144–187 (2004)
15. Ocone, D., Pardoux, E.: Asymptotic stability of the optimal filter with respect to its initial condition. SIAM J. Control Optim. **34**(1), 226–243 (1996)

16. Oudjane, N., Rubenthaler, S.: Stability and uniform particle approximation of nonlinear filters in case of non ergodic signals. Stoch. Anal. Appl. **31**(3), 421–448 (2005)
17. Stannat, W.: Stability of the pathwise filter equations for a time dependent signal on R^d. Appl. Math. Optim. **52**, 39–71 (2005)
18. Veretennikov, A.Yu.: Estimates of the mixing rate for stochastic equations. (Russian) Teor. Veroyatnost. Primenen. **32**(2), 299–308 (1987) Engl. transl.: Theory Probab. Appl. **32**(2), 273–281 (1987)
19. Veretennikov, A.Yu.: On polynomial mixing and the rate of convergence for stochastic differential and difference equations. Teor. Veroyatn. Primenen. **44**(2), 312–327 (1999); Engl. transl.: Theory Probab. Appl. **44**(2), 361–374 (2000)