



UNIVERSITY OF LEEDS

This is a repository copy of *On Robustness of Discrete Time Optimal Filters*.

White Rose Research Online URL for this paper:

<http://eprints.whiterose.ac.uk/104262/>

Version: Accepted Version

Article:

Veretennikov, A and Kleptsyna, ML (2016) On Robustness of Discrete Time Optimal Filters. *Mathematical Methods of Statistics*, 25 (3). pp. 207-218. ISSN 1066-5307

<https://doi.org/10.3103/S1066530716030042>

© Allerton Press, Inc., 2016. This is an author produced version of a paper published in *Mathematical Methods of Statistics*. Uploaded in accordance with the publisher's self-archiving policy. The final publication is available at Springer via <https://doi.org/10.3103/S1066530716030042>.

Reuse

Unless indicated otherwise, fulltext items are protected by copyright with all rights reserved. The copyright exception in section 29 of the Copyright, Designs and Patents Act 1988 allows the making of a single copy solely for the purpose of non-commercial research or private study within the limits of fair dealing. The publisher or other rights-holder may allow further reproduction and re-use of this version - refer to the White Rose Research Online record for this item. Where records identify the publisher as the copyright holder, users can verify any specific terms of use on the publisher's website.

Takedown

If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.



eprints@whiterose.ac.uk
<https://eprints.whiterose.ac.uk/>

On robustness of discrete time optimal filters

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

August 28, 2016

Abstract

A new result on stability of an optimal nonlinear filter for a Markov chain with respect to small perturbations on every step is established. An exponential recurrence of the signal is assumed.

1 Introduction

Stability of optimal filters is a topical research area in the last three or even more decades. In this direction, a lot has been understood and achieved under the “uniform ergodicity” assumptions due to the method by Atar and Zeitouni (see [?]) based on the Birkhoff metric (also known as projective or Hilbert metric). This method under such assumptions guarantees an exponential rate, with which the optimal filter algorithms “forgets” wrong – or unspecified – initial conditions. The method has been extended to the “non-uniform ergodic” case (see [?]) by combining the application of Birkhoff metric with a modified version of the coupling method, which led to exponential and polynomial rates, with which the algorithm may “forget” wrong initial data. However, an unspecified initial distribution is not the only option for an unspecified model. Small errors on each step of the algorithm (in discrete time case) is one more possibility to “spoil” the model and it is quite a natural option. In the “uniformly ergodic” case this issue was also studied, see [?]. The “non-uniformly ergodic” case is still waiting for its investigation and our goal here is to attack this

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

†University of Leeds, UK; National Research University Higher School of Economics, & Institute of Information Transmission Problems, Moscow, Russia, e-mail: A.Veretennikov at leeds.ac.uk

problem. In our setting only “uniformly small” errors are allowed; the conditions on the densities of the noise both in the signal and in the observations look a bit too strict, so that new studies will be required to weaken conditions so as to include wider classes of processes.

The setting described earlier is not only interesting as such: it may also serve as a base for studying unspecified models with an unknown parameter. In such models, observations should allow to estimate the parameter. Once the estimator is, at least, consistent, there is a hope that the filtering algorithm for a model with an estimate instead of the “true parameter” may be close enough to the exact model. Hence, the previous studies hopefully could be applied. This programme – again in the “uniform” case – was realised in [?, ?], where it was *assumed* that the estimator satisfies certain large deviation conditions. However, in many examples “non-uniform” conditions are more than natural. Hence, a large part of the problem with unknown parameters remains open and requires further investigations. We restrict ourselves to the case of exponential recurrence and exponential moments for the signal and postpone other cases – like a polynomial one – till further research.

The paper consists of four sections: I – introduction; II – the main result, remarks and examples, auxiliary results; III – proof of the Lemma 1; IV – proof of the main result.

2 Main Result, Examples, Auxiliary results

We consider the following model, with a non-observed (Markov) state process $\{X_n, n \geq 0\}$ and an observation process $\{Y_n, n \geq 1\}$, taking value in \mathbb{R}^d and \mathbb{R}^ℓ respectively. We assume that the state sequence $\{X_n, n \geq 0\}$ is defined as a homogeneous Markov chain with transition probability kernel $Q(x, dx')$, *i.e.*:

$$\mathbb{P}[X_n \in dx' | X_{0:n-1}] |_{X_{n-1}=x} = \mathbb{P}[X_n \in dx' | X_{n-1}] |_{X_{n-1}=x} = Q(x, dx'), \quad (1)$$

for all $n \geq 1$, and with initial distribution μ_0 .

We also assume (cf. [?]) that given the state sequence $\{X_n, n \geq 0\}$, the observations $\{Y_n, n \geq 1\}$ are independent, that the conditional distribution of Y_n depends only on X_n , and that the conditional probability distribution $\mathbb{P}[Y_n \in dy | X_n = x]$ is absolutely continuous with respect to the Lebesgue measure, *i.e.*:

$$\mathbb{P}[Y_n \in dy | X_n = x] = \Psi(x, y) dy, \quad (2)$$

for some Borel measurable with respect to the couple (x, y) function Ψ . The basic example which is to be covered will be the following:

The basic example covered by the model in (??)–(??) is a discrete time filter for a Hidden Markov chain (X_n) with values in the Euclidean space \mathbb{R}^d , with conditionally Markov observations (Y_n) also from \mathbb{R}^ℓ satisfying the system

$$X_{n+1} = X_n + b(X_n) + \xi_{n+1}, \quad n \geq 0, \quad (3)$$

$$Y_n = h(X_n) + V_n, \quad n \geq 1, \quad (4)$$

where (ξ_n, V_n) is a sequence of IID random vectors of dimension $d + \ell$ with densities $q_\xi(x)$, $q_V(y)$, $b(\cdot)$ is a d -dimensional vector-function, $h(\cdot)$ an ℓ -dimensional vector-function, that is,

$$Q(x, dx') = q_\xi((x' - x - b(x))) dx', \quad (5)$$

and

$$\Psi(x, y) = q_V(y - h(x)) \quad (6)$$

(recall that q_ξ and q_V stand for the densities of ξ_1 and V_1 respectively).

The problem addressed in this paper is as follows. Assume that the exact parameters of the model (??)–(??) — i.e., the initial distribution μ_0 , the transition kernel $Q(x, dx')$ and the conditional density of the observations $\Psi(x, y)$ — are known with some errors or that we know only an approximations to the exact characteristics of this model. Hence, the statistician is unable to use the exact optimal filtering algorithm for estimation of X_n at each time n , and he is left to apply a filtering algorithm with wrong parameters and with additional errors in the algorithm itself.

Under such conditions, the goal is to investigate the asymptotic behaviour of this error in the available algorithm in the long run. It follows from the earlier results on the subject – see [?] – that it is sufficient to work with errors in the kernels assuming that initial distribution μ_0 is known exactly. (If not, it may be tackled by using the methods and results from [?].) More precisely the setting will be explained in the section ?? below.

Throughout the paper, we denote the wrong transition kernel and conditional density of the observations by $P(x, dx')$ and by $\Xi(x, y)$ respectively. Assumptions will be stated in the form of Q, P and Ψ, Ξ ; also, examples will be provided in terms of the coefficients and properties of the original system (??)–(??).

In most cases in the sequel we will be using one single integration sign for single or for mutiple integrals (except in certain cases for double integrals where a single sign my be confusing). Also, we will often drop the area of integration if it is the whole space, i.e., \mathbb{R}^d , or \mathbb{R}^{dn} , or likewise, except in cases where it may be confusing or where we want to emphasize the dimension.

2.1 Assumptions and main result

To explain the main problem addressed in this paper in detail, we should formulate what exact and especially wrong filtering algorithms are. Let us start with the exact one. The problem of nonlinear filtering is to compute at each time n the conditional probability distribution μ_n of the state X_n given the observation sequence $Y_{1:n} = \{Y_1, Y_2, \dots, Y_n\}$, *i.e.*:

$$\mu_n(A) = \mu_n^Y(A) = \mathbb{P}[X_n \in A \mid Y_1, \dots, Y_n].$$

Using Bayes' formula, the exact posterior filtering conditional measure can be represented as a probability measure for any Y via the following random non-linear operator \bar{S}_n^{Y, μ_0} , applied to the initial measure μ_0 (as usual, we write the operator on the right from the measure, see the end of the formula below),

$$\begin{aligned} \mu_n(dx_n) &= \mu_n^Y(dx_n) = \mathbb{P}_{\mu_0}(X_n \in dx_n \mid Y_1, \dots, Y_n) \\ &= \int_{\mathbb{R}^{nd}} Q(x_{n-1}, dx_n) \prod_{i=1}^{n-1} Q(x_{i-1}, dx_i) d_i^{\mu_0} \Psi(x_i, Y_i) \mu_0(dx_0) \\ &= \frac{1}{c_n^{Y, \mu_0}} \int_{\mathbb{R}^{nd}} \prod_{i=1}^n Q(x_{i-1}, dx_i) \Psi(x_i, Y_i) \mu_0(dx_0) =: \mu_0 \bar{S}_n^{Y, \mu_0}(dx_n), \end{aligned} \quad (7)$$

where n -variate integration is taken over $dx_0 \dots dx_{n-1}$. Recall that here $\Psi(x_i, y_i)$ is a conditional density of Y_i at y_i given $X_i = x_i$ (see (??)), and $Q(x, dx')$ is a transition kernel for the Markov chain X_n , $n \geq 0$. The random normalization constant $c_n^{\mu_0}$ is defined as follows,

$$c_i^{Y, \mu_0} = E_{\mu_0} \prod_{j=1}^i \Psi(X_j, y_j) \Big|_{y_1=Y_1, \dots, y_i=Y_i}, \quad (8)$$

and, correspondingly,

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

Hence, the definition (??) maybe rewritten as

$$c_n^{\mu_0} = c_n^{Y, \mu_0} = \int_{\mathbb{R}^{(n+1)d}} \prod_{i=1}^n Q(x_{i-1}, dx_i) \Psi(x_i, Y_i) \mu_0(dx_0). \quad (9)$$

It is also convenient to let

$$c_0^Y = 1.$$

Now, the “wrong filtering algorithm” can be formulated more precisely as follows. Recall that it is assumed that we do not know the transition kernel $Q(x, dx')$ and the conditional density of the observations $\Psi(x, y)$ exactly, but only some approximations $P(x, dx')$ and $\Xi(x, y)$ respectively. Hence we can define another sequence of measures $(\mu'_n(A))_{n \geq 1}$ as follows:

$$\begin{aligned} \mu'_n(dx_n) &= \mu_n^Y(dx_n) = \int_{\mathbb{R}^{nd}} \prod_{i=1}^n P(x_{i-1}, dx_i) \tilde{d}_i^{\mu_0} \tilde{\Xi}(x_i, Y_i) \mu_0(dx_0) \\ &= \frac{1}{\tilde{c}_n^{Y, \mu_0}} \int_{\mathbb{R}^{nd}} \prod_{i=1}^n P(x_{i-1}, dx_i) \Xi(x_i, Y_i) \mu_0(dx_0) =: \mu_0 \tilde{S}_n^{Y, \mu_0}(dx_n), \end{aligned} \tag{10}$$

where the “wrong” normalizing constant \tilde{c}_n^{Y, μ_0} can be defined as follows:

$$\tilde{c}_n^{Y, \mu_0} = \int_{\mathbb{R}^{(n+1)d}} \prod_{i=1}^n P(x_{i-1}, dx_i) \Xi(x_i, Y_i) \mu_0(dx_0).$$

The measure μ'^Y is nothing else but the conditional measure of X' given the observation Y' (see (??)–(??)) with observations Y' replaced by the “real” observations Y . This replacement will turn out to be correctly defined thanks to the assumption (A3) – see below – due to which the conditional distributions given Y and given Y' are equivalent in the sense of equivalent measures. Note that without this condition the issue of well-posedness of this replacement would arise since, for example, it may be impossible to substitute the values taken from a “wrong distribution” if this distribution is singular with respect to the actual one. The property of equivalence of measures is standard in the area – cf. [?] – and it removes this “obstacle”. We do not discuss here what kind of consequences could be met without this assumption.

The problem under consideration is to estimate the behaviour of the difference,

$$E_{\mu_0} \|\mu'_n{}^Y - \mu_n^Y\|_{TV}, \quad n \geq 0,$$

in the long run. We may not hope that this discrepancy goes to zero as $n \rightarrow \infty$, but just that under certain conditions it may remain small for all values of n , if the

actual and “wrong” models are close enough in some particular quantitative sense. Hence, we aim to establish an upper bound to the quantity

$$\sup_{n \geq 0} E_{\mu_0} \|\mu_n^{X,Y} - \mu_n^Y\|_{TV} \leq ?$$

with such a right hand side that it remains small if the “wrong model” is close enough to the exact one.

Assumptions

Naturally, we will need certain assumptions.

(A1) — bounded (small) perturbations of the kernels —

We assume that

$$\ln\left(\sup_{x, x', z, y} \frac{Q(x, dx')\Psi(z, y)}{P(x, dx')\Xi(z, y)}\right) + \ln\left(\sup_{x, x', z, y} \frac{P(x, dx')\Xi(z, y)}{Q(x, dx')\Psi(z, y)}\right) = q < \infty.$$

(A2) — local “mixing” —

We assume that for any $R > 0$

$$C_R =: \sup_{D_R} \left(\frac{Q(x_0, dx')}{Q(v_0, dx')}, \frac{P(x_0, dx')}{P(v_0, dx')} \right) < \infty, \quad (11)$$

with $D_R := \{(x_0, v_0, x') : |x_0|, |v_0|, |x'| \leq R\}$.

(A3) — positiveness of conditional densities —

$$\Psi(x, y) > 0, \quad \Xi(x, y) > 0, \quad \forall x, y.$$

(A4) — condition of exponential recurrence in terms of the transition kernels:

there exist $\rho \in (0, 1)$, $R, K, c > 0$ such that for $|x| > R$,

$$\left(\int \exp(c|x'|)Q(x, dx') \right) \vee \left(\int \exp(c|x'|)P(x, dx') \right) \leq \rho \exp(c|x|), \quad (12)$$

and for $|x| \leq R$,

$$\left(\int \exp(c|x'|)Q(x, dx') \right) \vee \left(\int \exp(c|x'|)P(x, dx') \right) \leq K. \quad (13)$$

This condition may be also re-written as follows in terms of the mean values $a(x) := \mathbb{E}_x^Q X_1$ and $|a'(x)| := |\mathbb{E}_x^P X_1|$ (generally speaking, with different $\rho < 1$ and $K < \infty$):

$$|a(x)| \leq \rho|x| + K, \quad \& \quad |a'(x)| \leq \rho|x| + K.$$

Here as usual, \mathbb{E}^μ stands for the expectation with respect to a probability measure μ .

(A5) — uniformly small influence of observations: there exists $\delta > 0$ such that

$$\sup_{x,y,x'} \frac{\Psi(x,y)}{\Psi(x',y)} \vee \sup_{x,y,x'} \frac{\Xi(x,y)}{\Xi(x',y)} \leq 1 + \delta,$$

and for ρ from (A4)

$$(1 + \delta)\rho < 1. \quad (14)$$

Although the assumption (A4) requires explicitly exponential or lighter tails of the noise in the signal, we give examples with both exponential and polynomial tails showing in what situations the assumption (A5) may be verified, with a hope that polynomial examples may be useful in the future. Also note that the assumption (A2) is used essentially in the auxiliary results, which are briefly reminded below with some small changes; hence, the role this assumption will be practically invisible and to appreciate it the reader should have a look at earlier papers such as [?]. However, very briefly, we could not hope to estimate the total variation distance between two measures if, say, transition measures of the signal process from different states were singular; hence, some condition of this sort is indispensable, and the particular form of this condition in terms of ratios suits the use of the Birkhoff metric technique.

In the sequel it will be helpful to define the “conditional kernels”

$$\bar{Q}^{y,\mu}(x, dx') := Q(x, dx') \frac{\Psi(x', y)}{\iint Q(x, dx_1) \Psi(x_1, y) \mu(dx)}, \quad (15)$$

and

$$\bar{P}^{y,\mu}(x, dx') := P(x, dx') \frac{\Xi(x', y)}{\iint P(x, dx_1) \Xi(x_1, y) \mu(dx)}. \quad (16)$$

In the calculus below the variable y will play the role of Y_i for some i . Also, slightly abusing notations where it does not look confusing, we will use y for the vector (y_1, \dots, y_n) , and likewise Y will be used as a short notation for (Y_1, \dots, Y_n) . We also emphasize that for multiple integrals the unique sign of integration is used except for the double integrals where it may be confusing. Also, if there is no domain of integration, then the (in all cases definite) integral is taken over the whole space.

Also, we will be using the notation $(X'_n{}^Y)$, which will stand for a Markov process with transition probabilities given by the (non-homogeneous) transition kernels

$$\bar{P}'_n := \bar{P}^{y,\mu}|_{(y=Y_n, \mu=\mu'_n{}^Y)}, \quad n = 0, 1, \dots, \quad (17)$$

where the sequence of measures $\mu'_n{}^Y$ was defined earlier in (??); note that the latter coincides with the sequence of marginal distributions of the process $(X'_n{}^Y)$ at time n ; also note that generally speaking, the processes (X'_n) and $(X'_n{}^Y)$ are different.

The following theorem is the main result of the paper.

Theorem 1. *Let $\int e^{\epsilon|x|} \mu_0(dx) < \infty$. Then, under the assumption (A2) – (A5) above, there exists a constant $C > 0$ such that for any model satisfying also (A1) the following bound holds true:*

$$\sup_n \mathbb{E}_{\mu_0} \|\mu'_n{}^Y - \mu_n^Y\|_{TV} \leq Cq. \quad (18)$$

Emphasize that here the constant C does not depend on q . Also, note that q may be an arbitrary value greater than zero; however, the problem is, of course, meaningful if q is small.

2.2 Remarks and examples

Remark 1. *Assumption (A1) is valid for the model (??)–(??) with $q_v = q_\xi = C \exp(-|x|)$ and with any Borel function \tilde{b} such that $\tilde{b}(x) = b(x)$ if $|x| > K$ for some $K \geq 0$. It would be also nice to localize this condition, see the Remark ?? in the sequel; we leave it till further studies.*

Remark 2. Let us show how the assumption (A5) may be checked in some examples.

Example 1. Let $q_v(y) \sim c(1 + |y|)^{-m}$, $|y| \rightarrow \infty$. Assume that $\sup_x |h(x)| \leq \delta'$ and this δ' is small. Then for $|y| \rightarrow \infty$,

$$\begin{aligned} \frac{\Psi(x, y)}{\Psi(x', y)} &= \frac{q_v(y - h(x))}{q_v(y - h(x'))} \\ &\sim \frac{(1 + |y - h(x)|)^{-m}}{(1 + |y - h(x')|)^{-m}} = \frac{(1 + |y - h(x')|)^m}{(1 + |y - h(x)|)^m} \\ &= \left(\frac{1 + |y - h(x')|}{1 + |y - h(x)|} \right)^m \leq \left(\frac{1 + |y| + \delta'}{1 + |y| - \delta'} \right)^m \\ &= \left(\frac{1 + \delta'/(1 + |y|)}{1 - \delta'/(1 + |y|)} \right)^m \approx 1 + 2m\delta'/(1 + |y|) \leq 1 + 2m\delta'. \end{aligned}$$

For y bounded, the ratio remains close to one if δ' is small enough.

Example 2. Let $g_v(y) \sim c \exp(-|y|)$ as $|y| \rightarrow \infty$ (equivalent, i.e., the ratio converges to one), and let g_v be continuous. Then, for $|y|$ large enough we have,

$$\begin{aligned} \frac{\Psi(x, y)}{\Psi(x', y)} &= \frac{q_v(y - h(x))}{q_v(y - h(x'))} \\ &= \frac{q_v(y - h(x))/c \exp(-|y - h(x)|)}{q_v(y - h(x'))/c \exp(-|y - h(x')|)} = \frac{c \exp(-|y - h(x)|)}{c \exp(-|y - h(x')|)} \\ &\sim \exp(-|y - h(x)| + |y - h(x')|) \leq \exp(2\delta') \approx 1 + 2\delta'. \end{aligned}$$

And for $|y|$ bounded the ratio remains close to one due to continuity.

Remark 3. It would be interesting to replace the Assumption (A1) by a **local** assumption of the type,

$$\ln\left(\sup_{x, x', z, y \in K} \frac{Q(x, dx')\Psi(z, y)}{P(x, dx')\Xi(z, y)} \right) + \ln\left(\sup_{x, x', z, y \in K} \frac{P(x, dx')\Xi(z, y)}{Q(x, dx')\Psi(z, y)} \right) = q_K < \infty$$

with q_K small, perhaps, in addition to

$$\ln\left(\sup_{x, x', z, y} \frac{Q(x, dx')\Psi(z, y)}{P(x, dx')\Xi(z, y)} \right) + \ln\left(\sup_{x, x', z, y} \frac{P(x, dx')\Xi(z, y)}{Q(x, dx')\Psi(z, y)} \right) = q < \infty,$$

(with q arbitrary and finite) and to change the current statement of the Theorem ?? to one of the following bounds,

$$E_{\mu_0} \|\mu_n'^Y - \mu_n^Y\|_{TV} < Cq_K + \ln \sup_{x \in K} \mathbb{E}_x \exp(\alpha \hat{\tau}), \quad (19)$$

or,

$$E_{\mu_0} \|\mu_n'^Y - \mu_n^Y\|_{TV} < Cq_K + q \ln \sup_{x \in K} \mathbb{E}_x \exp(\alpha \hat{\tau}). \quad (20)$$

At the moment it is a conjecture that one of the bounds (??-??) or something close to it may hold true under less rigorous conditions than those in the Theorem ??, including also Gaussian densities.

Remark 4. Sufficient conditions for the assumption (A4) in terms of the system (??)-(??) and its approximation

$$X'_{n+1} = X'_n + \tilde{b}(X'_n) + \xi'_{n+1}, \quad n \geq 0, \quad (21)$$

$$Y'_n = \tilde{h}(X'_n) + V'_n, \quad n \geq 1, \quad (22)$$

may be offered as follows.

(A4') — another condition of exponential recurrence: there exist $r > 0$, $N > 0$ such that for $|x| \geq N$,

$$|x + b(x)| \vee |x + \tilde{b}(x)| \leq |x| - r,$$

and for any $R > 0$

$$\sup_{|x| \leq R} |x + b(x)| \vee \sup_{|x| \leq R} |x + \tilde{b}(x)| < \infty,$$

and

$$\mathbb{E}\xi_1 = \mathbb{E}\xi'_1 = 0,$$

and finally, there exists $\epsilon > 0$ such that

$$\mathbb{E} \exp(\epsilon |\xi_1|) + \mathbb{E} \exp(\epsilon |\xi'_1|) < \infty.$$

Apparently, (A4') implies the following: there exists a constant $C < \infty$ such that for any probability measure μ ,

$$\sup_y \int \int e^{\epsilon |x'|} \mu \bar{Q}^{y, \mu}(x, dx') \mu(dx) \leq C \int e^{\epsilon |x|} \mu(dx).$$

Equivalently, we could say that for any probability measure μ we have,

$$\sup_{\omega} \mathbb{E}_{\mu} (e^{\epsilon|X_1|} | Y_1) \leq C \int e^{\epsilon|x|} \mu(dx). \quad (23)$$

The same holds true for the approximation kernel P . It follows from [?, ?] that under the assumption (A4'), the assumption (A4) holds true.

2.3 Auxiliary results

In [?] it was proved, in particular, that under the ‘‘exponential’’ assumptions equivalent to (A4) the following estimate holds true:

$$E_{\mu_0} \|\mu_0 \bar{S}_n^{Y, \mu_0} - \nu_0 \bar{S}_n^{Y, \nu_0}\|_{TV} \leq C \exp(-C'n). \quad (24)$$

Here $\mu_0 \bar{S}_n^{Y, \mu_0}$ is defined by (??).

We will need some modification of (??). Recall that the proof of this estimate was based on the inequalities (14) and (20) from [?]. In turn, (14)/[?] followed from (11) and (12)/[?], while (20)/[?] was a corollary from the results about mixing for the recurrent and ergodic signal process. What is important for the present paper, is that the basic inequality (12) [?] admits an improved version under the condition that the initial Birkhoff distance (13) between measures μ_0 and ν_0 is finite:

$$\rho(\mu_0, \nu_0) < \infty.$$

In [?] this was not assumed and there was no reason for using such an improved version; on the contrary, the absence of this assumption allowed to cover a wider class of processes. However, now this will be important and the new version we need is as follows:

$$\rho(\mu_n, \nu_n) \leq \rho((\mu_n, \nu_n), (\nu_n, \mu_n)) \leq C \pi_R^{k-1} \rho(\mu_0, \nu_0), \quad (25)$$

with some $C > 0$ and $\pi_R < 1$. We do not explain here what are exactly k, μ_n, ν_n , et al. because it would require to copy several pages from [?], but use the notations from [?] verbatim. The point is that as a result of these improvements, we now formulate a version of Theorem 1 from [?] as follows.

Theorem 2 *Let $\int e^{\epsilon|x|} \mu_0(dx) < \infty$. Then under the assumptions (A2) – (A4), the following bound holds true: there exist constants $C, C(\mu), \alpha, \epsilon > 0$ such that*

$$E_{\mu_0} \|\mu_0 \bar{S}_n^{Y, \mu_0} - \nu_0 \bar{S}_n^{Y, \nu_0}\|_{TV} \leq C(\mu_0) \exp(-\alpha n) \rho(\mu_0, \nu_0), \quad (26)$$

with

$$C(\mu) \leq \int \exp(\epsilon|x|) \mu(dx).$$

Note that the Theorem does not use the condition (A1) here because the statement relates only to the identical kernels: $P \equiv Q$, and also $\Xi \equiv \Psi$, in which case (A1) holds automatically with $q = 0$.

Also note that both versions – the Theorem 2 above and the Theorem ?? from [?] – could be combined with the help of the value $1 \wedge \rho(\mu_0, \nu_0)$ in the right hand side.

Lemma 1. *Assume (A4) and (A5). Then for the “conditional kernel” $\bar{P}^{y,\mu}$ defined by (??) the following bounds holds true: for $|x|$ large enough,*

$$\sup_y \left| \int e^{c|x'|} \bar{P}^{y,\mu}(x, dx') \right| \leq \rho' e^{c|x|}, \quad (27)$$

and with any $R > 0$ for $|x| \leq R$,

$$\sup_y \left| \int e^{c|x'|} \bar{P}^{y,\mu}(x, dx') \right| \leq K', \quad (28)$$

with $\rho' = \rho(1 + \delta)$ and $K' = (1 + \delta)K$. Moreover, since $\rho' < 1$ (see the Assumption (A5)), the following inequality holds true,

$$\sup_{t \geq 0} \mathbb{E}_x e^{c|X_t^{x,Y}|} \leq K'' + e^{c|x|}, \quad (29)$$

with $K'' = K'/(1 - \rho')$.

3 Proof of auxiliary Lemma

Proof of Lemma ??.

Let us check the inequality (??). We have due to (A5) and (A4), if $|x| > R$, then

$$\begin{aligned} \sup_y \int e^{c|x'|} \bar{P}^{y,\mu}(x, dx') &= \sup_y \int e^{c|x'|} P(x, dx') \frac{\Xi(x', y)}{\iint P(x, dx_1) \Xi(x_1, y) \mu(dx)} \\ &\stackrel{(A5)}{\leq} (1 + \delta) \int e^{c|x'|} P(x, dx') \stackrel{(A4)}{\leq} (1 + \delta) \rho e^{c|x|} = \rho' e^{c|x|}, \end{aligned}$$

indeed, with $\rho' = (1 + \delta)\rho < 1$. Now let us verify the inequality (??). If $|x| \leq R$, then we have,

$$\begin{aligned} \sup_y \int e^{c|x'|} \bar{P}^{y,\mu}(x, dx') &= \sup_y \int e^{c|x'|} P(x, dx') \frac{\Xi(x', y)}{\iint P(x, dx_1) \Xi(x_1, y) \mu(dx)} \\ &\stackrel{(A5)}{\leq} (1 + \delta) \int e^{c|x'|} P(x, dx') \stackrel{(??)}{\leq} (1 + \delta) K = K'. \end{aligned}$$

Further, denote $\hat{\mathcal{F}}_n = \sigma(Y_1, \dots, Y_n; X_0^{Y}, \dots, X_n^{Y})$. We estimate,

$$E \left(e^{c|X_n^{Y}|} | \hat{\mathcal{F}}_{n-1} \right) \leq \rho' e^{c|X_{n-1}^{Y}|} + K',$$

and in a similar manner by induction,

$$E_x e^{c|X_n^{Y}|} \leq \frac{K'}{1 - \rho'} + (\rho')^n e^{c|x|} \leq \frac{K'}{1 - \rho'} + e^{c|x|}.$$

The Lemma ?? is proved.

4 Proof of Theorem ??

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

$$\rho(\mu, \nu) = \begin{cases} \ln \frac{(\inf s : \mu \leq s\nu)}{(\sup t : \mu \geq t\nu)}, & \text{if finite,} \\ +\infty, & \text{otherwise.} \end{cases} \quad (30)$$

Another equivalent definition reads,

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

Similarly to (??), for any measure μ we can define the following nonlinear operators $\bar{S}_{k:n}^Y$: for any $k < n$,

$$\mu \bar{S}_{k:n}^Y(A) = \mu \bar{S}_{k:n}^Y(A) = \frac{1}{C_{k:n}^Y} \int_{\mathbb{R}^{(n-k)d}} \mathbf{1}(x_n \in A) \prod_{j=k+1}^n Q(x_{j-1}, dx_j) \Psi(x_j, y_j) \mu(dx_k) |_{y=Y},$$

with a normalizing constant $c_{k:n}$ (which depends on μ as well):

$$c_{k:n} = c_{k:n}^Y = \int_{\mathbb{R}^{(n-k)d}} \prod_{j=k+1}^n Q(x_{j-1}, dx_j) \Psi(x_j, y_j) \mu(dx_k) |_{y=Y}, \quad k < n,$$

and for $k = n$ we let

$$\mu \bar{S}_{n:n}^Y(A) = \mu^Y(A).$$

As usual, (conditional) Markov kernels $\bar{Q}^y(x, dx')$ determine operators on measures: we write,

$$\mu \bar{Q}^y(A) := \frac{\int_{\mathbb{R}^{2d}} 1(x' \in A) \Psi(x, y) Q(x, dx') \mu(dx)}{\int_{\mathbb{R}^{2d}} \Psi(x, y) Q(x, dx') \mu(dx)},$$

and similarly,

$$\mu \bar{P}^y(A) := \frac{\int_{\mathbb{R}^{2d}} 1(x' \in A) \Xi(x, y) P(x, dx') \mu(dx)}{\int_{\mathbb{R}^{2d}} \Xi(x, y) P(x, dx') \mu(dx)}.$$

Now we have a simple identity, which will play the key role in the proof:

$$\mu_n^{\prime,Y} - \mu_n^Y = \sum_{k=1}^n (\mu_k^{\prime,Y} \bar{S}_{k:n}^Y - \mu_{k-1}^{\prime,Y} \bar{S}_{k-1:n}^Y). \quad (31)$$

It was noted in [?] that the following important property holds true:

$$\mu_{k-1}^{\prime,Y} \bar{S}_{k-1:n}^Y = (\mu_{k-1}^{\prime,Y} \bar{Q}^{Y_k}) \bar{S}_{k:n}^Y; \quad (32)$$

recall that the operator \bar{Q}^{Y_k} reads,

$$\mu \bar{Q}^{Y_k}(dx') = \frac{\int Q(x, dx') \Psi(x', Y_k) \mu(dx)}{\iint Q(x, dx_1) \Psi(x_1, Y_k) \mu(dx)}.$$

For the reader's convenience we recall the reasoning. Indeed,

$$\mu\bar{S}_{k:n}^Y = \frac{1}{c_{k:n}^Y} \mu S_{k:n}^Y,$$

where the linear *non-normalized* operator $S_{k:n}^Y$ is defined as follows:

$$\mu S_{k:n}^Y(A) = \int_{\mathbb{R}^{(n-k)d}} \mathbf{1}(x_n \in A) \prod_{j=k+1}^n \Psi(x_j, Y_j) Q(x_{j-1}, dx_j)$$

Therefore, we have,

$$\begin{aligned} \mu S_{k:n}^Y(A) &= \int_{\mathbb{R}^{(n-k-1)d}} \mathbf{1}(x_n \in A) \prod_{j=k+2}^n \Psi(x_j, Y_j) Q(x_{j-1}, dx_j) \\ &\quad \cdot \int_{\mathbb{R}^d} Q(x_k, dx_{k+1}) \Psi(x_{k+1}, Y_{k+1}) \mu(dx_k) = \nu S_{k+1:n}^Y(A), \end{aligned} \quad (33)$$

with the non-normalized measure $\nu(dx_{k+1})$ defined by the formula

$$\nu(dx_{k+1}) = \int_{\mathbb{R}^d} Q(x_k, dx_{k+1}) \Psi(x_{k+1}, Y_{k+1}) \mu(dx_k)$$

Hence,

$$\mu\bar{S}_{k:n}^Y = \frac{\mu S_{k:n}^Y}{c_{k:n}^{\mu,Y}} = \frac{\nu S_{k+1:n}^Y}{c_{k:n}^{\mu,Y}} = \frac{\nu S_{k+1:n}^Y}{c_{k+1:n}^{\nu,Y}} \cdot \frac{c_{k+1:n}^{\nu,Y}}{c_{k:n}^{\mu,Y}}. \quad (34)$$

Also note (follows from the calculus with $A = \mathbb{R}^d$) that

$$c_{k:n}^{\mu,Y} = c_{k+1:n}^{\nu,Y},$$

because

$$\begin{aligned} \mu S_{k:n}^Y(\mathbb{R}^d) &= \int_{\mathbb{R}^{(n-k-1)d}} \mathbf{1}(x_n \in \mathbb{R}^d) \prod_{j=k+2}^n \Psi(x_j, y_j) Q(x_{j-1}, dx_j) \\ &\quad \cdot \int_{\mathbb{R}^d} Q(x_k, dx_{k+1}) \Psi(x_{k+1}, y_{k+1}) \mu(dx_k)|_{y=Y} = \nu S_{k+1:n}^Y(\mathbb{R}^d). \end{aligned} \quad (35)$$

The equation (??) implies that

$$\mu\bar{S}_{k:n} = \frac{\nu S_{k+1:n}}{c_{k+1:n}^{\nu}} = \frac{(\nu/\nu(\mathbb{R}^d)) S_{k+1:n}}{c_{k+1:n}^{\nu}/\nu(\mathbb{R}^d)} = \frac{\tilde{\nu} S_{k+1:n}}{c_{k+1:n}^{\tilde{\nu}}},$$

with

$$\tilde{\nu}(dx') = \tilde{\nu}^{Y_k}(dx') = \frac{\nu(dx')}{\nu(\mathbb{R}^d)} = \mu \bar{Q}^{Y_k}(dx') = \frac{\int Q(x, dx') \Psi(x', Y_k) \mu(dx)}{\iint Q(x, dx_1) \Psi(x_1, Y_k) \mu(dx)}.$$

So, indeed, the announced important property (??) holds true.

Further, since $\mu_n^{\prime Y} \bar{S}_{n:n} = \mu_n^{\prime Y}$ and $\mu_0 \bar{S}_{0:n} = \mu_n^Y$ and because $\mu_0^{\prime Y} = \mu_0$ and $\mu_0^{\prime Y} \bar{S}_{0:n} = \mu_0 \bar{S}_{0:n} = \mu_n^Y$, we obtain,

$$\mu_n^{\prime Y} - \mu_n^Y = \sum_{k=1}^n (\mu_{k-1}^{\prime Y} \bar{P} \bar{S}_{k:n} - (\mu_{k-1}^{\prime Y} \bar{Q}) \bar{S}_{k:n}), \quad (36)$$

where $\mu_{k-1}^{\prime Y} \bar{Q} = \mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}$, $\mu_{k-1}^{\prime Y} \bar{P} = \mu_{k-1}^{\prime Y} \bar{P}^{Y_k}$. So, it follows that

$$\|\mu_n^{\prime Y} - \mu_n^Y\|_{TV} \leq \sum_{k=1}^n \|\mu_{k-1}^{\prime Y} \bar{P} \bar{S}_{k:n} - (\mu_{k-1}^{\prime Y} \bar{Q}) \bar{S}_{k:n}\|_{TV},$$

and

$$E_{\mu_0} \|\mu_n^{\prime Y} - \mu_n^Y\|_{TV} \leq \sum_{k=1}^n E_{\mu_0} \|\mu_{k-1}^{\prime Y} \bar{P}^{Y_k} \bar{S}_{k:n} - (\mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}) \bar{S}_{k:n}\|_{TV}. \quad (37)$$

2. By virtue of the Theorem 2 under our assumptions we have,

$$E_{\mu, \nu} \|\mu \bar{S}_{0:n}^{Y, \mu} - \nu \bar{S}_{0:n}^{Y, \nu}\|_{TV} \leq C(\mu, \nu) e^{-\alpha n} \rho(\mu, \nu),$$

where α does not depend on the initial measures, while $C(\mu, \nu)$ admits a bound

$$C(\mu, \nu) \leq \int (e^{c|x|} \mu(dx) + e^{c|x'|} \nu(dx')).$$

with some $c > 0$.

Also recall that due to the Lemma ??,

$$\sup_{t \geq 0} E_{\nu_0} e^{c|X_t^{\prime Y}|} \leq K + \int e^{c|x'|} \nu_0(dx'), \quad (38)$$

with a non-random K , and a similar bound holds true for the exact conditional process X :

$$\sup_{t \geq 0} E_{\nu_0} e^{\epsilon |X_t^Y|} \leq K + \int e^{\epsilon |x'|} \mu_0(dx'). \quad (39)$$

3. Further, all of the above imply that

$$\begin{aligned} E_{\mu_k, \nu_k} \left(\|\mu_k \bar{S}_{k:n}^{Y_{k+1}, \dots, Y_n} - \nu_k \bar{S}_{k:n}^{Y_{k+1}, \dots, Y_n}\|_{TV} \mid Y_1, \dots, Y_k \right) \\ \leq C(\mu_k, \nu_k) e^{-\alpha(n-k)} \rho(\mu_k, \nu_k) \\ \leq C e^{-\alpha(n-k)} \rho(\mu_k, \nu_k) \int e^{\epsilon |x|} (\mu_k^{Y_1, \dots, Y_k}(dx) + \nu_k^{Y_1, \dots, Y_k}(dx)), \end{aligned} \quad (40)$$

where the constants C and α are non-random and do not depend on k . Denote

$$D(\mu, \nu) := \int e^{\epsilon |x|} (\mu(dx) + \nu(dx)).$$

By virtue of the inequalities (??) and (??) together with (??)–(??), we have,

$$\begin{aligned} \mathbb{E}_{\mu_0} \|\mu_n^{\prime Y} - \mu_n^Y\|_{TV} \\ \leq \sum_{k=1}^n C \mathbb{E}_{\mu_0} D(\mu_k^{\prime Y}, \mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}) e^{-\alpha(n-k)} \sup_{\omega} \rho(\mu_k^{\prime Y}, \mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}) \\ \leq Cq \sum_{k=1}^n \mathbb{E}_{\mu_0} D(\mu_k^{\prime Y}, \mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}) e^{-\alpha(n-k)}. \end{aligned}$$

But $\mu_k^{\prime Y} = \mu_{k-1}^{\prime Y} \bar{P}^{Y_k}$ (a similar operator but with the wrong kernel), so we obtain

$$\rho(\mu_k^{\prime Y}, \mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}) = \rho(\mu_{k-1}^{\prime Y} \bar{P}^{Y_k}, \mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}) \leq q,$$

by the assumption (A1). Thus,

$$E_{\mu_0} \|\mu_n^{\prime Y} - \mu_n^Y\|_{TV} \leq \sum_{k=1}^n Cq e^{-\alpha(n-k)} \mathbb{E}_{\mu_0} D(\mu_k^{\prime Y}, \mu_{k-1}^{\prime Y} \bar{Q}^{Y_k}). \quad (41)$$

4. It remains to estimate the term $\mathbb{E}_{\mu_0} D(\mu'_k{}^Y, \mu'_{k-1}{}^Y \bar{Q}^{Y_k})$ so as to show that it does not exceed some finite constant uniformly in k . We have,

$$\mathbb{E}_{\mu_0} D(\mu'_k{}^Y, \mu'_{k-1}{}^Y \bar{Q}^{Y_k}) = \mathbb{E}_{\mu_0} \int e^{c|x|} \mu'_k{}^Y(dx) + \mathbb{E}_{\mu_0} \int e^{c|x|} \mu'_{k-1}{}^Y \bar{Q}^{Y_k}(dx).$$

Here the first term in the right hand side satisfies the bound

$$(\sup_k) \mathbb{E}_{\mu_0} \int e^{c|x|} \mu'_k{}^Y(dx) \leq K'' + \int e^{c|x|} \mu_0(dx) < \infty,$$

according to the bound (??) of the Lemma ?? and to the assumption of the Theorem $\int e^{\epsilon|x|} \mu_0(dx) < \infty$.

Let us inspect the second part of the right hand side. We have,

$$\begin{aligned} E_{\mu_0} \int e^{\epsilon|x|} \mu'_{k-1}{}^Y \bar{Q}^{Y_k}(dx) &= E_{\mu_0} \int e^{c|x_k|} \mu'_{k-1}{}^Y \bar{Q}^{Y_k}(dx_k) \\ &\stackrel{(A5)}{\leq} (1 + \delta) E_{\mu_0} \int e^{c|x_k|} (\mu'_{k-1} Q)(dx_k) \\ &\leq K(1 + \delta) + (1 + \delta) E_{\mu_0} \int e^{c|x|} \mu'_{k-1}(dx) \\ &\stackrel{(??.)}{\leq} K' + K' + (1 + \delta) \int e^{c|x|} \mu_0(dx) < \infty, \end{aligned}$$

according again to (??) and to the assumption (A5). The Theorem ?? is proved.

Acknowledgements

The authors are grateful to the anonymous referee for useful remarks.

For the second author, the article was prepared within the framework of a subsidy granted to the HSE by the Government of the Russian Federation for the implementation of the Global Competitiveness Program.

References

- [1] Atar, R., Zeitouni, O. Exponential stability for nonlinear filtering. *Ann. Inst. H. Poincaré Probab. Statist.* 33(6), 1997, 697-725.
- [2] Gulinskii, O. V. , Veretennikov, A. Yu. Rate of mixing and the averaging principle for stochastic recursive procedures. *Aut. Remote Control.* 51(6), 1990, 779-788.
- [3] Gulinsky, O. V., Veretennikov, A. Yu. Large deviations for discrete - time processes with averaging. VSP, Utrecht, The Netherlands, 1993.
- [4] Kleptsyna, M. L., Veretennikov, A. Yu. On discrete time ergodic filters with wrong initial data. *Probability Theory and Related Fields*, 141, 2008, 411-444.
- [5] Krasnosel'skii, M. A., Lifshits, E. A., Sobolev, A. V. Positive linear systems. The method of positive operators. Heldermann, Berlin, 1989.
- [6] Le Gland, F., Oudjane, N. Stability and uniform approximation of nonlinear filters using the Hilbert metric and application to particle filters, *The Annals of Applied Probability*, 14(1), 2004, 144-187.
- [7] Papavasiliou, A., A uniformly convergent adaptive particle filter, *J. Appl. Prob.* 42, 2005, 1053-1068.
- [8] Papavasiliou, A., Parameter estimation and asymptotic stability in stochastic filtering, *Stochastic Processes and their Applications*, 116, 2006, 1048-1065.