ON FILTERING OF MARKOV CHAINS IN STRONG NOISE

P.CHIGANSKY

ABSTRACT. The filtering problem for a finite state Markov chain observed in white noise is addressed in continuous time. The low signal to noise asymptotic is derived for the performance indices of MAP and MMSE estimates of the signal.

1. INTRODUCTION

Consider the continuous time signal/observation pair $(X_t, Y_t)_{t\ge 0}$, where the *signal* X is a Markov chain with values in $\mathbb{S} = \{a_1, ..., a_d\}$, transition intensities matrix Λ and initial distribution ν. The observation process Y is generated by

$$Y_t = \int_0^t h(X_s)ds + \sigma B_t,$$

where h is an $\mathbb{S} \mapsto \mathbb{R}$ function, $\sigma > 0$ is the noise intensity and B is a Brownian motion, independent of X. All the random variables are assumed to be supported on a complete probability space $(\Omega, \mathscr{F}, \mathrm{P})$.

The conditional probabilities $\pi_t(i) = \mathrm{P}(X_t = a_i|\mathscr{F}_t^Y)$, $i = 1, ..., d$, where $\mathscr{F}_t^Y = \sigma\{Y_s, 0 \le s \le t\}$ is the σ-algebra of events generated by the observations past, are the main building blocks in the signal estimation problem. In particular the maximum a posteriori probability (MAP) and the minimal mean square error (MMSE) estimates are given by

$$\bar{X}_t = \operatorname*{argmax}_{a_i\in\mathbb{S}} \pi_t(i) \quad \text{and} \quad \widehat{X}_t = \sum_{i=1}^d a_i \pi_t(i), \quad t \ge 0, \tag{1.1}$$

respectively. The vector π_t of conditional probabilities satisfies the Wonham ([6], see also [5]) filtering Itô stochastic differential equation (SDE)

$$d\pi_t = \Lambda^* \pi_t dt + \sigma^{-2}\big(\mathrm{diag}(\pi_t) - \pi_t\pi_t^*\big)\big(dY_t - \pi_t(h)dt\big), \quad \pi_0 = \nu, \tag{1.2}$$

where $\mathrm{diag}(x)$, $x \in \mathbb{R}^d$ stands for the scalar matrix with x_i on the diagonal, h is a column vector with entries $h(a_i)$ and x^* is the transposed of x. Hereafter the functions on $\mathbb{S}$ are identified with the vectors in $\mathbb{R}^d$ and the space of probability measures on $\mathbb{S}$ is identified with the simplex $\mathcal{S}^{d-1} = \{x \in \mathbb{R}^d : x_i \ge 0, \sum_{i=1}^d x_i = 1\}$. For any $f : \mathbb{S} \mapsto \mathbb{R}$ and $\eta \in \mathcal{S}^{d-1}$ we denote $\eta(f) = \sum_{i=1}^d \eta_i f(a_i)$.

While the estimates can be efficiently calculated via (1.1) and (1.2), not much is known about the optimal performance they attain:

$$\mathcal{E}_{\mathrm{mse}}(t) = \min_{\zeta\in\mathbb{L}^2(\Omega,\mathscr{F}_t^Y,\mathrm{P})} \mathrm{E}\big(X_t - \zeta\big)^2 = \mathrm{E}\big(X_t - \widehat{X}_t\big)^2 = \mu(a^2) - \mathrm{E}\pi_t^2(a) \tag{1.3}$$

1991 *Mathematics Subject Classification.* 93E11, 62M05, 62M02.
Key words and phrases. nonlinear filtering, markov chains.
Research supported by a grant of the Israeli Science Foundation.

and

$$\mathcal{E}_{\mathrm{map}}(t) = \min_{\zeta\in\mathbb{L}^\infty(\Omega,\mathscr{F}^Y_t,\mathrm{P})} \mathrm{P}\big(X_t\ne\zeta\big) = 1-\mathrm{E}\max_{a_i\in\mathbb{S}}\pi_t(i). \tag{1.4}$$

Since the process $\bar{B}_t=\sigma^{-1}\Big(Y_t-\int_0^t\pi_s(h)ds\Big)$ is the *innovation* Brownian motion, the solution of (1.2) is a Markov process evolving in $\mathcal{S}^{d-1}$. Moreover as π_t takes values in a compact state space it is guaranteed to have at least one invariant measure $\mathcal{M}(d\eta)$ (on the Borel field of $\mathcal{S}^{d-1}$), which was recently shown to be unique (see [1], [2]) in the case of ergodic chain X. Hence both limits

$$\mathcal{E}_{\mathrm{mse}} := \lim_{t\to\infty}\mathcal{E}_{\mathrm{mse}}(t) \quad\text{and}\quad \mathcal{E}_{\mathrm{map}} := \lim_{t\to\infty}\mathcal{E}_{\mathrm{map}}(t)$$

exist and do not depend on ν.

In the case $d=2$, the exact expressions are available for both performance indices in terms of integrals with respect to the stationary probability density of π_t, which is explicitly computable by solving the corresponding Kolmogorov-Fokker-Plank equation (see [6], Chapter 15 §2 in [5]). The closed form solution for the KFP equation is unavailable in the multivariate case $d>2$, which makes the performance analysis of the filter (1.2) much more complicated.

The asymptotic expansion of $\mathcal{E}_{\mathrm{map}}$ is given in [4] for the slow switching signal, when the transition intensities matrix of the signal is $\Lambda^\varepsilon:=\Lambda\varepsilon$ and $\varepsilon\to0$. It is easy to see that this asymptotic is equivalent to the high signal-to-noise limit $\sigma\to0$:

$$\mathcal{E}^\sigma_{\mathrm{map}} = \Big(\sum_{i=1}^d\mu_i\sum_{j\ne i}\frac{2\lambda_{ij}}{(h_i-h_j)^2}\Big)\sigma^2\log\Big(\frac{1}{\sigma^2}\Big)\big(1+o(1)\big),\quad \sigma\to0$$

where μ is the stationary invariant distribution of X, i.e. $\mu_i=\lim_{t\to\infty}\mathrm{P}(X_t=a_i)>0$, $i=1,...,d$ and $h_i\ne h_j$ for all $i\ne j$ is assumed. Similar asymptotic is shown to hold for $\mathcal{E}^\sigma_{\mathrm{mse}}$ in[1] [3]:

$$\mathcal{E}^\sigma_{\mathrm{mse}} = \sum_{i=1}^d\sum_{j\ne i}\frac{2\mu_i\lambda_{ij}}{(h_i-h_j)^2}\big(a_i-a_j\big)^2\sigma^2\log\Big(\frac{1}{\sigma^2}\Big)\big(1+o(1)\big),\quad \sigma\to0$$

These limits reveal how the invariant measure $\mathcal{M}^\sigma(d\eta)$ of the filtering process π^σ_t concentrates as $\sigma\to0$ around $\mathcal{M}^0(d\eta)=\sum_{i=1}^d\mu_i\delta_{e_i}(d\eta)$, where e_i is a probability vector with the unit i-th entry.

In this note the low signal-to-noise asymptotic $\sigma\to\infty$ is considered. In this regime the optimal estimates converge to the corresponding trivial a priori estimates. The main result is the next theorem, which describes the concentration of $\mathcal{M}^\sigma(d\eta)$ around $\mathcal{M}^\infty(d\eta)=\delta_\mu(d\eta)$.

Theorem 1.1. *For any $t\ge0$*

$$\sigma\big(\pi^\sigma_t-\nu_t\big)\xrightarrow[\sigma\to\infty]{\mathbb{L}^p}\xi_t,\quad p\ge1 \tag{1.5}$$

where $\nu_t=e^{\Lambda^*t}\nu$ and ξ_t is a zero mean Gaussian process with the covariance matrix P_t, being solution of the Lyapunov equation

$$\dot{P}_t=\Lambda^*P+P\Lambda+\big(\mathrm{diag}(\nu_t)-\nu_t\nu_t^*\big)hh^*\big(\mathrm{diag}(\nu_t)-\nu_t\nu^*_t\big),\quad P_0=0. \tag{1.6}$$

[1]The discrete time case is treated in [3], but the result and all the arguments can be easily translated to the continuous-time

If X is ergodic the algebraic Lyapunov equation

$$0 = \Lambda^* P + P\Lambda + \big(\mathrm{diag}(\mu) - \mu\mu^*\big)hh^*\big(\mathrm{diag}(\mu) - \mu\mu^*\big) \tag{1.7}$$

has a unique solution P in the class of nonnegative definite matrices satisfying $\sum_{ij} P_{ij} = 0$. Let ξ be a zero mean Gaussian vector with covariance P, then for a continuous function $F : \mathcal{S}^{d-1} \mapsto \mathbb{R}$

$$\lim_{\sigma\to\infty} \int_{\mathcal{S}^{d-1}} F\big(\sigma(\eta - \mu)\big)\mathcal{M}^\sigma(d\eta) = \mathrm{E}F(\xi), \tag{1.8}$$

whenever the right hand side is well defined.

The immediate consequence of this theorem are asymptotic expressions for $\mathcal{E}^\sigma_{\mathrm{mse}}$ and $\mathcal{E}^\sigma_{\mathrm{map}}$

Corollary 1.2. *Assume that X is ergodic, then*

$$\lim_{\sigma\to\infty} \sigma^2\big(\mathcal{E}^\infty_{\mathrm{mse}} - \mathcal{E}^\sigma_{\mathrm{mse}}\big) = a^* P a,$$

where $\mathcal{E}^\infty_{\mathrm{mse}} := \mu(a^2) - \mu^2(a)$ is the a priori estimation error and P is defined by (1.7).

$\mathcal{E}^\sigma_{\mathrm{map}}$ may exhibit two different asymptotics, depending on μ.

Corollary 1.3. *If X is ergodic, then*

$$\lim_{\sigma\to\infty} \sigma\big(\mathcal{E}^\infty_{\mathrm{map}} - \mathcal{E}^\sigma_{\mathrm{map}}\big) = \mathrm{E}\max_{j\in\mathcal{J}} \xi_j, \tag{1.9}$$

where $\mathcal{E}^\infty_{\mathrm{map}} := 1 - \max_{a_i\in\mathbb{S}} \mu_i$ is the a priori error probability, ξ is a zero mean Gaussian random vector with covariance matrix P, defined in (1.7) and $\mathcal{J} = \{i : \mu_i = \max_j \mu_j\}$. If μ has a unique maximal atom, then for any integer $p \ge 1$

$$\lim_{\sigma\to\infty} \sigma^p\big(\mathcal{E}^\infty_{\mathrm{map}} - \mathcal{E}^\sigma_{\mathrm{map}}\big) = 0. \tag{1.10}$$

Note that when the maximal atom of μ is not unique, the right hand side of (1.9) remains positive in general as the following example demonstrates.

Example 1.4. For telegraphic signal $d = 2$ the Lyapunov equation (1.6) is one dimensional in the required class of matrices, since $-P_{12} = -P_{21} = P_{11} = P_{22} := P$, which satisfies

$$0 = -2(\lambda_{12} + \lambda_{21})P + (h_1 - h_2)^2\mu_1^2\mu_2^2.$$

For the binary chain $\mu_1 = \lambda_{21}/(\lambda_{12} + \lambda_{21})$ and $\mu_2 = \lambda_{12}/(\lambda_{12} + \lambda_{21})$ and hence

$$P = \frac{(h_1 - h_2)^2\lambda_{21}^2\lambda_{12}^2}{2(\lambda_{12} + \lambda_{21})^5}.$$

By Corollary 1.2

$$\lim_{\sigma\to\infty} \sigma^2 \lim_{t\to\infty} \Big(\mathrm{E}(X_t - \mu(a))^2 - \mathrm{E}\big(X_t - \pi_t(a)\big)^2\Big) = (a_1 - a_2)^2 \frac{(h_1 - h_2)^2\lambda_{21}^2\lambda_{12}^2}{2(\lambda_{12} + \lambda_{21})^5}.$$

By Corollary 1.3 if $\lambda_{12} \neq \lambda_{21}$,

$$\lim_{\sigma\to\infty} \sigma\Big(\lim_{t\to\infty} \mathrm{E}\max(\pi_t, 1 - \pi_t) - \max(\mu_1, 1 - \mu_1)\Big) = 0,$$

If $\lambda_{12} = \lambda_{21} := \lambda$, then

$$\lim_{\sigma\to\infty} \sigma\Big(\lim_{t\to\infty} \mathrm{E}\max(\pi_t, 1-\pi_t) - \frac{1}{2}\Big) = \max\big(\xi, -\xi\big) = \mathrm{E}|\xi| =$$
$$2\sqrt{P}\int_0^\infty x\frac{1}{\sqrt{2\pi}}e^{-x^2/2}dx = \frac{\sqrt{2}|h_1-h_2|}{\lambda}\cdot 0.3989....$$

2. The proofs

2.1. The proof of Theorem 1.1. Since $\nu_t = \exp(\Lambda^* t)\nu$ solves $\dot{\nu}_t = \Lambda^*\nu_t$, $\nu_0 = \nu$, the process $\delta^\sigma_t := \pi^\sigma_t - \nu_t$ satisfies

$$d\delta^\sigma_t = \Lambda^*\delta^\sigma_t dt + \sigma^{-1}\big(\mathrm{diag}(\pi^\sigma_t) - \pi^\sigma_t\pi^{\sigma *}_t\big)hd\bar{B}_t, \quad \delta^\sigma_0 = 0,$$

and hence

$$\delta^\sigma_t = \sigma^{-1}\int_0^t e^{\Lambda^*(t-s)}\big(\mathrm{diag}(\pi^\sigma_s) - \pi^\sigma_s\pi^{\sigma *}_s\big)hd\bar{B}_s.$$

Since the integrand is continuous and bounded for any $t \ge 0$,

$$\lim_{\sigma\to\infty}\delta^\sigma_t = 0, \quad \mathrm{P}-a.s.$$

which is valid in $\mathbb{L}^p$ for any fixed $p \ge 1$ as well by dominated convergence theorem. Let q^σ_t be solution of the linear SDE

$$dq^\sigma_t = \Lambda^* q^\sigma_t dt + \sigma^{-1}\big(\mathrm{diag}(\nu_t) - \nu_t\nu^*_t\big)hd\bar{B}_t, \quad q^\sigma_0 = \nu. \tag{2.1}$$

The process $\Delta^\sigma_t = \sigma(\pi^\sigma_t - q^\sigma_t)$ satisfies

$$d\Delta^\sigma_t = \Lambda^*\Delta^\sigma_t dt + \big(\Gamma(\pi^\sigma_t) - \Gamma(\nu_t)\big)hd\bar{B}_t, \quad \Delta^\sigma_0 = 0,$$

where $\Gamma(x) = \mathrm{diag}(x) - xx^*$ is set for brevity. Then

$$\Delta^\sigma_t = \int_0^t e^{\Lambda^*(t-s)}\big(\Gamma(\pi^\sigma_s) - \Gamma(\nu_s)\big)hd\bar{B}_s \xrightarrow{\sigma\to\infty} 0, \quad \mathrm{P}-a.s \text{ and in } \mathbb{L}^p,$$

since $\Gamma(\cdot)$ is continuous, π^σ_t and ν_t are bounded and $\pi^\sigma_t \to \nu_t$ as $\sigma \to \infty$.

Define $\xi_t = \sigma(q^\sigma_t - \nu_t)$, which satisfies

$$d\xi_t = \Lambda^*\xi_t dt + \big(\mathrm{diag}(\nu_t) - \nu_t\nu^*_t\big)hd\bar{B}_t, \quad \xi_0 = 0.$$

Clearly ξ_t is a zero mean diffusion with covariance matrix given by (1.6) and (1.5) follows

$$\sigma\big(\pi^\sigma_t - \nu_t\big) = \sigma\big(\pi^\sigma_t - q^\sigma_t\big) + \sigma\big(q^\sigma_t - \nu_t\big) \xrightarrow[\sigma\to\infty]{\mathbb{L}^p} \xi_t.$$

The stationary version (1.8) is verified by similar arguments, applied to (1.2) in its reduced form. Namely it is regarded now as a diffusion in $\mathbb{R}^{d-1}$, satisfying

$$d\pi^\sigma_t = \big(a + A^*\pi^\sigma_t\big)dt + \sigma^{-1}\big(\mathrm{diag}(\pi^\sigma_t) - \pi^\sigma_t\pi^{\sigma *}_t\big)bd\bar{B}_t, \tag{2.2}$$

where a is $(d-1)$ dimensional column vector with entries λ_{dj}, $j = 1, ..., d-1$, A is the square matrix with entries $\lambda_{ij} - \lambda_{dj}$ $1 \le i, j \le d-1$ and b is the vector with entries $h_i - h_d$, $i = 1, ..., d-1$. To simplify the notations we do not distinguish $d-1$ dimensional vectors (as e.g. vectors in $\mathcal{S}^{d-1}$), when embedded in $\mathbb{R}^d$ or $\mathbb{R}^{d-1}$. In particular the solutions of both (1.2) and (2.2) are identified and are denoted by π^σ_t, whose interpretation as a vector in $\mathbb{R}^d$ or in $\mathbb{R}^{d-1}$ should be clear from the context.

If (2.2) is solved subject to a random vector π_0 with distribution $\mathcal{M}^\sigma(d\eta)$ and independent of $\bar{B}$, the process π^σ_t is stationary. Note that the stationary probability vector μ, embedded in $\mathbb{R}^{d-1}$ as mentioned above, satisfies $0 = a + A^*\mu$ and thus the process $\delta^\sigma_t = \pi^\sigma_t - \mu$ satisfies

$$d\delta^\sigma_t = A^*\delta^\sigma_t dt + \sigma^{-1}\big(\mathrm{diag}(\pi^\sigma_t) - \pi^\sigma_t\pi^{\sigma*}_t\big)bd\bar{B}_t, \quad \delta^\sigma_0 = \pi_0 - \mu,$$

and hence

$$\delta^\sigma_t = e^{A^*t}(\pi_0 - \mu) + \sigma^{-1}\int_0^t e^{A^*(t-s)}\big(\mathrm{diag}(\pi^\sigma_s) - \pi^\sigma_s\pi^{\sigma*}_s\big)bd\bar{B}_s.$$

Since X is ergodic, A^* is a stability matrix, i.e. all its eigenvalues have negative real parts, and hence

$$\lim_{\sigma\to\infty}\varlimsup_{t\to\infty}\|\delta^\sigma_t\| = 0, \quad \mathrm{P}-a.s. \text{ and in } \mathbb{L}^p. \tag{2.3}$$

Now let q^σ_t be the solution of

$$dq^\sigma_t = (a + A^*q^\sigma_t)dt + \sigma^{-1}\big(\mathrm{diag}(\mu) - \mu\mu^*\big)bd\bar{B}_t, \quad q^\sigma_0 = \mu$$

Then the process $\Delta^\sigma_t = \sigma(\pi^\sigma_t - q^\sigma_t)$ satisfies

$$d\Delta^\sigma_t = A^*\Delta^\sigma_t dt + \big(\Gamma(\pi^\sigma_t) - \Gamma(\mu)\big)bd\bar{B}_t, \quad \Delta^\sigma_0 = \pi_0 - \mu$$

and

$$\Delta^\sigma_t = e^{A^*t}(\pi_0 - \mu) + \int_0^t e^{A^*(t-s)}\big(\Gamma(\pi^\sigma_s) - \Gamma(q^\sigma_s)\big)bd\bar{B}_s.$$

Since A is stable and $\Gamma(\cdot)$ is continuous, (2.3) implies

$$\lim_{\sigma\to\infty}\varlimsup_{t\to\infty}\|\Delta^\sigma_t\| = 0, \quad \mathrm{P}-a.s. \text{ and in } \mathbb{L}^p.$$

The process $\xi_t = \sigma(q^\sigma_t - \mu)$ satisfies

$$d\xi_t = A^*\xi_t + \big(\mathrm{diag}(\mu) - \mu\mu^*\big)bd\bar{B}_t, \quad \xi_0 = 0.$$

Since A is a stability matrix, the limit

$$\xi = \lim_{t\to\infty}\int_0^t e^{A^*(t-s)}\big(\mathrm{diag}(\mu) - \mu\mu^*\big)bd\bar{B}_s$$

exists and is a zero mean random Gaussian vector with the covariance matrix uniquely solving

$$0 = A^*P + PA + \big(\mathrm{diag}(\mu) - \mu\mu^*\big)bb^*\big(\mathrm{diag}(\mu) - \mu\mu^*\big).$$

Note that this equation is nothing but (1.7) with the imposed constrain. Since $\sigma(\pi^\sigma_t - \mu) = \sigma(\pi^\sigma_t - q^\sigma_t) + \xi_t$ we have

$$\lim_{\sigma\to\infty}\lim_{t\to\infty}\sigma(\pi^\sigma_t - \mu) = \xi,$$

and (1.8) follows since

$$\mathrm{E}F(\sigma(\pi^\sigma_t - \mu)) \equiv \int_{\mathcal{S}^{d-1}} F\big(\sigma(\eta - \mu)\big)\mathcal{M}^\sigma(d\eta), \quad \forall t \geq 0.$$

by stationarity of π^σ_t. $\square$

2.2. **The proof of Corollaries.** Let π_t^σ be the stationary solution of (1.2), then by (1.3) and (1.8) one gets the claim of Corollary 1.2

$$\sigma^2\big(\mu(a^2)-\mu^2(a)-\mathcal{E}^\sigma_{\mathrm{mse}}\big)=\sigma^2\big(\mathrm{E}\pi_t^2(a)-\mu^2(a)\big)=$$
$$\sigma^2\mathrm{E}\big(\pi_t^\sigma(a)-\mu(a)\big)^2=a^*\mathrm{E}\sigma^2\big(\pi_t^\sigma-\mu\big)\big(\pi_t^\sigma-\mu\big)^*a\xrightarrow{\sigma\to\infty}a^*Pa.$$

Let $\mathcal{J}=\{i:\mu_i=\max_j\mu_j\}$ and assume $\mu_1\in\mathcal{J}$ for definiteness. Let π_t^σ be the stationary solution of (1.2). Then

$$\sigma\big(\mathcal{E}^\infty_{\mathrm{map}}-\mathcal{E}^\sigma_{\mathrm{map}}\big)=\sigma\big(\mathrm{E}\max_{a_i\in\mathbb{S}}\pi_t^\sigma(i)-\max_i\mu_i\big)=\mathrm{E}\sigma\max_{a_i\in\mathbb{S}}(\pi_t^\sigma(i)-\mu_1)=$$
$$\mathrm{E}\max_{a_i\in\mathbb{S}}\big(\sigma(\pi_t^\sigma(i)-\mu_i)+\sigma(\mu_i-\mu_1)\big)\xrightarrow{\sigma\to\infty}\mathrm{E}\max_{j\in\mathcal{J}}\xi_j,$$

where ξ is a zero mean Gaussian random vector with covariance P given by (1.7) and the convergence holds by (1.8), since $\max_i(x_i)$, $x\in\mathbb{R}^d$ is a continuous function and $\mu_i-\mu_1<0$ for $i\notin\mathcal{J}$.

Suppose now that μ_1 is the unique maximal atom of μ and let $r=\max_{j\neq1}|\mu_1-\mu_j|>0$. Let $A_\sigma:=\{\|\pi_t^\sigma-\mu\|\le r/2\}$, where $\|\cdot\|$ is the usual Euclidian norm and let $I(A_\sigma)$ be the indicator function of A_σ and $A_\sigma^c=\Omega\backslash A_\sigma$. Then

$$\max_{a_i\in\mathbb{S}}\pi_t^\sigma(i)=I(A_\sigma)\pi_t^\sigma(1)+I(A^c_\sigma)\max_{a_i\in\mathbb{S}}\pi_t^\sigma(i)=\pi_t^\sigma(1)+I(A^c_\sigma)\big(\max_{a_i\in\mathbb{S}}\pi_t^\sigma(i)-\pi_t^\sigma(1)\big),$$

Hence for any two integers $q>p\ge1$,

$$\sigma^p\big|\mathrm{E}\max_{a_i\in\mathbb{S}}\pi_t^\sigma(i)-\max_i\mu_i\big|=$$
$$\sigma^p\big|\big(\mathrm{E}\pi_t^\sigma(1)-\mu_1\big)+\mathrm{E}I(A^c_\sigma)\big(\max_{a_i\in\mathbb{S}}\pi_t^\sigma(i)-\pi_t^\sigma(1)\big)\big|=$$
$$\sigma^p\mathrm{E}I(A^c_\sigma)\big|\max_{a_i\in\mathbb{S}}\pi_t^\sigma(i)-\pi_t^\sigma(1)\big|\le2\sigma^p\frac{\mathrm{E}\|\pi_t^\sigma-\mu\|^q}{r^q}\xrightarrow{\sigma\to\infty}0,$$

since by (1.8), the limit $\lim_{\sigma\to\infty}\sigma^q\mathrm{E}\|\pi_t^\sigma-\mu\|^q$ exists and is finite. □

References

[1] P. Baxendale, P. Chigansky, R.Liptser, Asymptotic stability of the Wonham filter: ergodic and nonergodic signals, SIAM J. Control Optim. 43 (2004), no. 2, 643–669.

[2] P. Chigansky, A formula for the top Lyapunov exponent of the Zakai equation, submitted (preprint `math.PR/0404515` at `http://www.arxiv.org/`)

[3] G. Golubev, On filtering for a Hidden Markov Chain under square perfromance critetion, Problems of Information Transmission Vol. 36, No. 3, 2000, pp. 213-219

[4] R. Khasminskii, O. Zeitouni, Asymptotic filtering for finite state Markov chains, Stoch. Processes and Appl., 1996, vol. 63, pp. 1-10.

[5] R.Lipster, A.Shiryaev, *Statistics of random processes: theory and applications*, I,II, Springer-Verlag, 2nd Ed., 2001

[6] W. M. Wonham, Some applications of stochastic differential equations to optimal nonlinear filtering. J. Soc. Indust. Appl. Math. Ser. A Control 2 347–369 (1965).

Department of Mathematics, The Weizmann Institute of Science, Rehovot 76100, Israel

E-mail address: `pavel.chigansky@weizmann.ac.il`