Annales de l'Institut Henri Poincaré - Probabilités et Statistiques
2014, Vol. 50, No. 1, 236–255
DOI: 10.1214/12-AIHP514

ANNALES DE L'INSTITUT HENRI POINCARÉ
PROBABILITÉS ET STATISTIQUES
www.imstat.org/aihp

Nonconventional limit theorems in averaging

Yuri Kifer

[a]*Institute of Mathematics, The Hebrew University, Jerusalem 91904, Israel. E-mail: kifer@math.huji.ac.il*

Received 9 September 2011; revised 4 June 2012; accepted 4 July 2012

Abstract. We consider "nonconventional" averaging setup in the form $\frac{\mathrm{d}X^\varepsilon(t)}{\mathrm{d}t} = \varepsilon B(X^\varepsilon(t), \Xi(q_1(t)), \Xi(q_2(t)), \dots, \Xi(q_\ell(t)))$ where $\Xi(t), t \ge 0$ is either a stochastic process or a dynamical system with sufficiently fast mixing while $q_j(t) = \alpha_j t, \alpha_1 < \alpha_2 < \cdots < \alpha_k$ and $q_j, j = k+1, \dots, \ell$ grow faster than linearly. We show that the properly normalized error term in the "nonconventional" averaging principle is asymptotically Gaussian.

Résumé. Nous considérons un cadre non conventionnel de moyenne de la forme $\frac{\mathrm{d}X^\varepsilon(t)}{\mathrm{d}t} = \varepsilon B(X^\varepsilon(t), \Xi(q_1(t)), \Xi(q_2(t)), \dots, \Xi(q_\ell(t)))$ où $\Xi(t), t \ge 0$ est un processus stochastique ou un système dynamique suffisamment mélangeant tandis que $q_j(t) = \alpha_j t, \alpha_1 < \alpha_2 < \cdots < \alpha_k$ et $q_j, j = k+1, \dots, \ell$ ont une croissance sur-linéaire. Nous montrons que le terme d'erreur après renormalisation est asymptotiquement gaussien.

MSC: Primary 34C29; secondary 60F17; 37D20

Keywords: Averaging; Limit theorems; Martingales; Hyperbolic dynamical systems

1. Introduction

Nonconventional ergodic theorems (see [12]) known also after [2] as polynomial ergodic theorems studied the limits of expressions having the form $1/N \sum_{n=1}^N F^{q_1(n)} f_1 \cdots F^{q_\ell(n)} f_\ell$ where F is a weakly mixing measure preserving transformation, f_i's are bounded measurable functions and q_i's are polynomials taking on integer values on the integers. Originally, these results were motivated by applications to multiple recurrence for dynamical systems taking functions f_i being indicators of some measurable sets and only convergence in the L^2-sense was dealt with but later [1] provided also almost sure convergence under additional conditions. Recently such results were extended in [6] to the continuous time dynamical systems, i.e. to expressions of the form

$$\frac{1}{\mathcal{T}} \int_0^{\mathcal{T}} F^{q_1(t)} f_1 \cdots F^{q_\ell(t)} f_\ell \,\mathrm{d}t,$$

where F^s is now an ergodic measure preserving flow.

In this paper we consider the averaging setup

$$X^\varepsilon(n+1) = X^\varepsilon(n) + \varepsilon B\bigl(X^\varepsilon(n), \Xi\bigl(q_1(n)\bigr), \dots, \Xi\bigl(q_\ell(n)\bigr)\bigr) \tag{1.1}$$

in the discrete time case and

$$\frac{\mathrm{d}X^\varepsilon(t)}{\mathrm{d}t} = \varepsilon B\bigl(X^\varepsilon(t), \Xi\bigl(q_1(t)\bigr), \dots, \Xi\bigl(q_\ell(t)\bigr)\bigr) \tag{1.2}$$

in the continuous time case with $\varXi$ being either a stochastic process or having the form $\varXi(s)=F^s f$ where F^s is a dynamical system and f is a function. Positive functions $q_1,\dots,q_\ell$ will satisfy certain conditions which will be specified in the next section, in particular, first k of them are linear while others grow faster than preceeding ones. An example where (1.2) emerges is obtained when we consider a time dependent small perturbation of the oscillator equation

$$\ddot{x}+\lambda^2 x=\varepsilon g(x,\dot{x},t), \tag{1.3}$$

where the force term g depends on time in a random way $g(x,y,t)=g(x,y,\varXi(q_1(t)),\dots,\varXi(q_\ell(t)))$. Then passing to the polar coordinates (r,ϕ) with $x=r\sin(\lambda(t-\phi))$ and $\dot{x}=\lambda r\cos(\lambda(t-\phi))$ the equation (1.3) will be transformed into (1.2). It seems reasonable that a random force may depend on versions of a same process or a dynamical system moving with different speeds which is what we have here.

As it is well known (see, for instance, [29]), if $B(x,y_1,\dots,y_\ell)$ is bounded and Lipschitz continuous in x and the limit

$$\bar{B}(x)=\lim_{\mathcal{T}\to\infty}\frac{1}{\mathcal{T}}\int_0^{\mathcal{T}} B\big(x,\varXi\big(q_1(t)\big),\dots,\varXi\big(q_\ell(t)\big)\big)\,\mathrm{d}t \tag{1.4}$$

exists then for any $S\ge 0$,

$$\lim_{\varepsilon\to 0}\sup_{0\le t\le S/\varepsilon}\big|X^\varepsilon(t)-\bar{X}^\varepsilon(t)\big|=\lim_{\varepsilon\to 0}\sup_{0\le t\le S}\big|Z^\varepsilon(t)-\bar{Z}(t)\big|=0, \tag{1.5}$$

where

$$\frac{\mathrm{d}\bar{X}^\varepsilon(t)}{\mathrm{d}t}=\varepsilon\bar{B}\big(\bar{X}^\varepsilon(t)\big)\quad\text{and}\quad Z^\varepsilon(t)=X^\varepsilon(t/\varepsilon),\qquad \bar{Z}(t)=\bar{X}^\varepsilon(t/\varepsilon). \tag{1.6}$$

In the discrete time case we have to take

$$\bar{B}(x)=\lim_{N\to\infty}\frac{1}{N}\sum_{n=0}^{N} B\big(x,\varXi\big(q_1(n)\big),\dots,\varXi\big(q_\ell(n)\big)\big) \tag{1.7}$$

and (1.5) remains true with $\bar{X}^\varepsilon$ given by (1.6) and (1.7). Almost everywhere limits in (1.4) and (1.7) follow from [22] under our (and even weaker) assumptions and in some relevant to our setup cases they could be derived by nonconventional pointwise ergodic theorems from [6] and [1].

After nonconventional ergodic theorems (or in the probabilistic language laws of large numbers) are established the next natural step is to obtain central limit theorem type results which was accomplished in [25]. The averaging principle (1.5) can be considered as an extension of the ergodic theorem since if $B(x,\xi_1,\dots,\xi_\ell)$ in (1.1) does not depend on x then $X^{1/N}(N)$ becomes the nonconventional average $\frac{1}{N}S_N$ where $S_N=\sum_{0\le n\le N}B(\varXi(q_1(n),\dots,\varXi(q_\ell(n)))$. Now if $\frac{1}{N}S_N$ converges to $\bar{B}$ as $N\to\infty$ then convergence in distribution of $\sqrt{N}(X^{1/N}(N)-\bar{B})$ to a normal random variable is, in fact, a nonconventional central limit theorem. The main goal of this paper is to extend the functional central limit theorem type results obtained in [25] for such sums S_N to the above nonconventional averaging setup in the spirit of what was done in the standard (conventional) averaging case in [18] and [20]. Central limit theorem type results turn in the averaging setup into assertions about Gaussian approximations of the slow motion X^ε given by (1.1) or by (1.2) where $\varXi$ is a fast mixing stochastic process or a dynamical system while unlike the standard (conventional) case we have the process $\varXi$ taken simultaneously at different times $q_i(t)$ in the right hand side of (1.1) and (1.2).

We prove, first, our limit theorems for stochastic processes under rather general conditions resembling the definition of mixingales (see [27] and [28]) and then check these conditions for more familiar classes of stochastic processes and dynamical systems. In [25] we imposed mixing assumptions in a standard way relying on two parameter families of σ-algebras (see [5]) while our assumptions here use only filtrations (i.e. nondecreasing families) of σ-algebras which are easier to construct for various classes of dynamical systems. As one of applications we check some form of our conditions for Anosov flows which serve as fast motions in our nonconventional averaging setup where we rely on the notion of Markov families from [8] and [9].

At the end of the paper we discuss a fully coupled averaging setup in our nonconventional situation where already an averaging principle itself becomes a problem.

2. Preliminaries and main results

Our setup consists of a $\wp$-dimensional stochastic process $\{\Xi(t), t \geq 0 \text{ or } t = 0, 1, \dots\}$ on a probability space $(\Omega, \mathcal{F}, Pr)$ together with a filtration of σ-algebras $\mathcal{F}_l \subset \mathcal{F}, 0 \leq l \leq \infty$ so that $\mathcal{F}_l \subset \mathcal{F}_{l'}$ if $l \leq l'$. For convenience we extend the definitions of $\mathcal{F}_l$ given only for $l \geq 0$ to negative l by defining $\mathcal{F}_l = \mathcal{F}_0$ for $l < 0$. In order to relax required stationarity assumptions to some kind of weak "limiting stationarity" our setup includes another probability measure P on the space $(\Omega, \mathcal{F})$. Namely, we assume that the distribution of $\Xi(t)$ with respect to P does not depend on t and the joint distribution of $\{\Xi(t), \Xi(t')\}$ for $t \geq t'$ depends only on $t - t'$ which can be written in the form

$$\Xi(t)P = \mu \quad \text{and} \quad \big(\Xi(t), \Xi\big(t'\big)\big)P = \mu_{t-t'} \quad \text{for all } t \geq t', \tag{2.1}$$

where μ is a probability measure on $\mathbb{R}^{\wp}$ and $\mu_s, s \geq 0$ is a probability measure on $\mathbb{R}^{\wp} \times \mathbb{R}^{\wp}$.

Our setup relies on two probability measures Pr and P in order to include, for instance, Markov processes $\Xi(t)$ satisfying the Doeblin condition (see [16] or [10]) starting at a fixed point or with another noninvariant distribution. Then Pr will be a corresponding probability in the path space while P will be the stationary probability constructed by the initial distribution being the invariant measure of $\Xi(t)$. Usual mixing conditions for stochastic processes are formulated in terms of a double parameter family of σ-algebras via a dependence coefficient between widely separated past and future σ-algebras (cf. [5] and [25]) but this approach often is not convenient for applications to dynamical systems where natural future σ-algebras do not seem to exist unless an appropriate symbolic representation is available. By this reason we formulate below a different set of mixing and approximation conditions for the process Ξ which seem to be new and will enable us to treat some of dynamical systems models within a class of stochastic processes satisfying our assumptions.

In order to avoid some of technicalities we restrict ourselves here mostly to bounded functions though our results can be obtained for more general classes of functions with polynomial growth supplemented by appropriate moment boundedness conditions similarly to [25]. For any function $g = g(\xi, \tilde{\xi})$ on $\mathbb{R}^{\wp} \times \mathbb{R}^{\wp}$ introduce its Hölder norm

$$|g|_\kappa = \sup\left\{\big|g(\xi, \tilde{\xi})\big| + \frac{|g(\xi, \tilde{\xi}) - g(\xi', \tilde{\xi}')|}{|\xi - \xi'|^\kappa + |\tilde{\xi} - \tilde{\xi}'|^\kappa} \colon \xi \neq \xi', \xi \neq \xi'\right\}. \tag{2.2}$$

Here and in what follows $|\psi - \tilde{\psi}|^\kappa$ for two vectors $\psi = (\psi_1, \dots, \psi_\varrho)$ and $\tilde{\psi} = (\tilde{\psi}_1, \dots, \tilde{\psi}_\varrho)$ denotes the sum $\sum_{i=1}^{\varrho} |\psi_i - \tilde{\psi}_i|^\kappa$. Next, for $p, q \geq 1$ and $s \geq 0$ we define a sort of a mixing coefficient

$$\begin{aligned}\eta_{p,\kappa,s}(n) = \sup_{t \geq 0}\big\{ \big\| & E\big(g\big(\Xi(n+t), \Xi(n+t+s)\big) | \mathcal{F}_{[t]}\big) \\ & - E_P g\big(\Xi(n+t), \Xi(n+t+s)\big)\big\|_p \colon g = g(\xi, \tilde{\xi}), |g|_\kappa \leq 1\big\}, \quad \eta_{p,\kappa}(n) = \eta_{p,\kappa,0}(n),\end{aligned} \tag{2.3}$$

where $\|\cdot\|_p$ is the L^p-norm on the space $(\Omega, \mathcal{F}, Pr)$, $[\cdot]$ denotes the integral part and throughout this paper we write E for the expectation with respect to Pr and E_P for the expectation with respect to P. We will need also an (one-sided) approximation coefficient

$$\zeta_q(n) = \sup_{t \geq 0}\big\| E\big(\Xi(t) | \mathcal{F}_{[t]+n}\big) - \Xi(t)\big\|_q. \tag{2.4}$$

Assumption 2.1. *Given $\kappa \in (0, 1]$ there exist $p, q \geq 1$ and $m, \delta > 0$ satisfying*

$$\gamma_m = E\big|\Xi(0)\big|^m < \infty, \qquad \frac{1}{2} \geq \frac{1}{p} + \frac{2}{m} + \frac{\delta}{q}, \qquad \delta < \kappa - \frac{\varrho}{p}, \qquad \kappa q > 1 \tag{2.5}$$

with $\varrho = (\ell - 1)\wp$ and such that

$$\sum_{n=0}^{\infty} n\big(\eta_{p,\kappa}^{1-\varrho/(p\theta)}(n) + \zeta_q^\delta(n)\big) < \infty \quad \text{and} \quad \lim_{n \to \infty} \eta_{p,\kappa,s}(n) = 0 \quad \text{for all } s \geq 0, \tag{2.6}$$

where $\frac{\varrho}{p} < \theta < \kappa$.

Next, let $B = B(x, \xi) = (B^{(1)}(x, \xi), \dots, B^{(d)}(x, \xi))$, $\xi = (\xi_1, \dots, \xi_\ell) \in \mathbb{R}^{\ell\wp}$ be a d-vector function on $\mathbb{R}^d \times \mathbb{R}^{\ell\wp}$ such that for some constant $K > 0$ and all $x, \tilde{x} \in \mathbb{R}^d$, $\xi, \tilde{\xi} \in \mathbb{R}^{\ell\wp}$, $i, j, l = 1, \dots, d$,

$$\left|B^{(i)}(x, \xi)\right| \le K, \qquad \left|B^{(i)}(x, \xi) - B^{(i)}(\tilde{x}, \tilde{\xi})\right| \le K\left(|x - \tilde{x}| + \sum_{j=1}^{\ell} |\xi_j - \tilde{\xi}_j|^\kappa\right)$$

$$\text{and} \quad \left|\frac{\partial B^{(i)}(x, \xi)}{\partial x_j}\right| \le K, \qquad \left|\frac{\partial^2 B^{(i)}(x, \xi)}{\partial x_j \partial x_l}\right| \le K. \tag{2.7}$$

We will be interested in the central limit theorem type results as $\varepsilon \to 0$ for the solution $X^\varepsilon(t) = X_x^\varepsilon(t)$ of the equation

$$\frac{\mathrm{d}X^\varepsilon(t)}{\mathrm{d}t} = \varepsilon B\big(X^\varepsilon(t), \Xi\big(q_1(t)\big), \Xi\big(q_2(t)\big), \dots, \Xi\big(q_\ell(t)\big)\big), \qquad X_x^\varepsilon(0) = x, \quad t \in [0, \mathcal{T}/\varepsilon], \tag{2.8}$$

where $q_1(t) < q_2(t) < \cdots < q_\ell(t), t > 0$ are increasing functions such that $q_j(t) = \alpha_j t$ for $j \le k < \ell$ with $\alpha_1 < \alpha_2 < \cdots < \alpha_k$ whereas the remaining $q_j' s$ grow faster in t. Namely, we assume similarly to [25] that for any $\gamma > 0$ and $k + 1 \le i \le \ell$,

$$\lim_{t\to\infty}\big(q_i(t + \gamma) - q_i(t)\big) = \infty \tag{2.9}$$

and

$$\lim_{t\to\infty}\big(q_i(\gamma t) - q_{i-1}(t)\big) = \infty. \tag{2.10}$$

Set

$$\bar{B}(x) = \int B(x, \xi_1, \dots, \xi_\ell)\,\mathrm{d}\mu(\xi_1)\cdots \mathrm{d}\mu(\xi_\ell). \tag{2.11}$$

We consider also the solution $\bar{X}^\varepsilon(t) = \bar{X}_x^\varepsilon(t)$ of the averaged equation

$$\frac{\mathrm{d}\bar{X}^\varepsilon(t)}{\mathrm{d}t} = \varepsilon \bar{B}\big(\bar{X}^\varepsilon(t)\big), \qquad \bar{X}_x^\varepsilon(0) = x. \tag{2.12}$$

It will be convenient to denote $Z^\varepsilon(t) = X^\varepsilon(t/\varepsilon)$, $\bar{Z}(t) = \bar{X}^\varepsilon(t/\varepsilon)$ and to introduce $Y^\varepsilon(t) = Y_y^\varepsilon(t)$ by

$$Y_y^\varepsilon(t) = y + \int_0^t B\big(\bar{Z}(s), \Xi\big(q_1(s/\varepsilon)\big), \Xi\big(q_2(s/\varepsilon)\big), \dots, \Xi\big(q_\ell(s/\varepsilon)\big)\big)\,\mathrm{d}s. \tag{2.13}$$

Theorem 2.2. *Suppose that* (2.7), (2.9), (2.10) *and Assumption* 2.1 *hold true. Then the family of processes* $G^\varepsilon(t) = \varepsilon^{-1/2}(Y_z^\varepsilon(t) - \bar{Z}_z(t)), t \in [0, \mathcal{T}]$ *converges weakly as* $\varepsilon \to 0$ *to a Gaussian process* $G^0(t), t \in [0, \mathcal{T}]$ *having not necessarily independent increments (see an example in* [25]*) with covariances of its components* $G^0(t) = (G^{0,1}(t), \dots, G^{0,d}(t))$ *having the form* $EG^{0,l}(s)G^{0,m}(t)) = \int_0^{\min(s,t)} A^{l,m}(u)\,\mathrm{d}u$ *with the matrix function* $\{A^{l,m}(u), 1 \le l, m \le d\}$ *computed in Section* 4. *Furthermore, the family of processes* $Q^\varepsilon(t) = \varepsilon^{-1/2}(Z^\varepsilon(t) - \bar{Z}(t)), t \in [0, \mathcal{T}]$ *converges weakly as* $\varepsilon \to 0$ *to a Gaussian process* $Q^0(t), t \in [0, \mathcal{T}]$ *which solves the equation*

$$Q^0(t) = G^0(t) + \int_0^t \nabla \bar{B}\big(\bar{Z}(s)\big) Q^0(s)\,\mathrm{d}s. \tag{2.14}$$

In the discrete time setup (1.1) *the similar results hold true assuming that* q_i*'s take on integer values on integers,* γ *in* (2.9) *is replaced by* 1, α_i *is replaced by* i *for* $i = 1, \dots, k$ *and defining* $Z^\varepsilon(t) = X^\varepsilon([t/\varepsilon])$ *together with* $Y^\varepsilon = Y_y^\varepsilon$ *given by*

$$Y_y^\varepsilon(t) = y + \int_0^t B\big(\bar{Z}(s), \Xi\big(q_1\big([s/\varepsilon]\big)\big), \Xi\big(q_2\big([s/\varepsilon]\big)\big), \dots, \Xi\big(q_\ell\big([s/\varepsilon]\big)\big)\big)\,\mathrm{d}s \tag{2.15}$$

while leaving all other definitions and assumptions the same as above.

Observe that we work with $\bar{B}$ defined by (2.11) but in our circumstances the law of large numbers from [22] yields $\bar{B}$ also as an almost sure limit in (1.4) and (1.7) even under weaker conditions than here. Note also that we need the full strength of (2.6) only for one argument in Section 4 borrowed from [18] but for a standard limit theorem not in the averaging setup, i.e. when $B(x,\xi_1,\dots,\xi_\ell)=B(\xi_1,\dots,\xi_\ell)$ does not depend on x, it suffices to require only summability of the expression in brackets in (2.6).

An important point in the proof of the first part of Theorem 2.2 is to introduce the representation

$$B(x,\xi)=\bar{B}(x)+B_1(x,\xi_1)+\cdots+B_\ell(x,\xi_1,\dots,\xi_\ell), \tag{2.16}$$

where $\xi=(\xi_1,\dots,\xi_\ell)$ and for $i<\ell$,

$$\begin{aligned}&B_i(x,\xi_1,\dots,\xi_\ell)\\&\quad=\int B(x,\xi_1,\dots,\xi_\ell)\,\mathrm{d}\mu(\xi_{i+1})\cdots\,\mathrm{d}\mu(\xi_\ell)-\int B(x,\xi_1,\dots,\xi_\ell)\,\mathrm{d}\mu(\xi_i)\cdots\,\mathrm{d}\mu(\xi_\ell)\end{aligned} \tag{2.17}$$

while

$$B_\ell(x,\xi_1,\dots,\xi_\ell)=B(x,\xi_1,\dots,\xi_\ell)-\int B(x,\xi_1,\dots,\xi_\ell)\,\mathrm{d}\mu(\xi_\ell). \tag{2.18}$$

Next, set

$$\begin{aligned}&Y_i^\varepsilon(t)=\int_0^{t/\alpha_i}B_i\big(\bar{Z}(s),\Xi\big(q_1(s/\varepsilon)\big),\Xi\big(q_2(s/\varepsilon)\big),\dots,\Xi\big(q_\ell(s/\varepsilon)\big)\big)\,\mathrm{d}s\quad\text{for } i=1,\dots,k\\&\quad\text{while for } i=k+1,\dots,\ell \text{ set}\quad Y_i^\varepsilon(t)=\int_0^t B_i\big(\bar{Z}(s),\Xi\big(q_1(s/\varepsilon)\big),\Xi\big(q_2(s/\varepsilon)\big),\dots,\Xi\big(q_\ell(s/\varepsilon)\big)\big)\,\mathrm{d}s\end{aligned} \tag{2.19}$$

with $Y_0^\varepsilon(t)=Y_{0,y}^\varepsilon(t)=y+\int_0^t B_i(\bar{Z}(s))\,\mathrm{d}s$. Thus Y_y^ε from (2.13) has the representation

$$Y_y^\varepsilon(t)=Y_0^\varepsilon(t)+\sum_{i=1}^k Y_i^\varepsilon(\alpha_i t)+\sum_{i=k+1}^\ell Y_i^\varepsilon(t). \tag{2.20}$$

We consider also $X_0^\varepsilon(t)=\bar{X}^\varepsilon(t)$, $X_i^\varepsilon(t)=X_{i,x}^\varepsilon(t)=x+\varepsilon\int_0^t B_i(X_i^\varepsilon(s),\Xi(q_1(s)),\dots,\Xi(q_\ell(s)))\,\mathrm{d}s$ and $Z_i^\varepsilon(t)=X_i^\varepsilon(t/\varepsilon)$ for all $i\ge 0$. For $i\ge 1$ set also

$$G_i^\varepsilon(t)=\varepsilon^{-1/2}Y_i^\varepsilon(t)\quad\text{and}\quad Q_i^\varepsilon(t)=\varepsilon^{-1/2}Z_i^\varepsilon(t). \tag{2.21}$$

Relying on martingale approximations (which also can be done employing mixingales from [27] and [28]) we will show that any linear combination $\sum_{i=1}^k\lambda_i G_i^\varepsilon$ converges weakly as $\varepsilon\to 0$ to a Gaussian process $\sum_{i=1}^k\lambda_i G_i^0$. It turns out that in the continuous time case each G_i^ε, $i=k+1,\dots,\ell$ converges weakly as $\varepsilon\to 0$ to zero, and so the processes Y_i^ε, $i>k$ do not play any role in the limit. It follows that G^ε converges weakly to a Gaussian process G^0 such that $G(t)=\sum_{i=1}^k\lambda_i G_i^0(\alpha_i t)$. On the other hand, in the discrete time case each G_i^ε, $i>k$ cannot be disregarded, in general, and it converges weakly as $\varepsilon\to 0$ to a Gaussian process G_i^0 which is independent of any other G_j^0. The above difference between discrete and continuous time cases is due to the different natural forms of the assumption (2.9) in these two cases. These arguments yield the first part of Theorem 2.2 while its second part concerning convergence of Q^ε as $\varepsilon\to 0$ is proved via some Taylor expansion and approximation arguments.

In order to clarify the role of the coefficients $\eta_{p,\kappa}$ and ζ_q we compare them with the more familiar mixing and approximation coefficients defined via a two parameter family of σ-algebras $\mathcal{G}_{s,t}\in\mathcal{F}$, $-\infty\le s\le t\le\infty$ by

$$\varpi_p(n)=\sup_{s\ge 0,g}\big\{\big\|E(g|\mathcal{G}_{-\infty,s})-E_P g\big\|_p\colon g\text{ is }\mathcal{G}_{s+n,\infty}\text{-measurable and }|g|\le 1\big\} \tag{2.22}$$

and

$$\beta_q(n) = \sup_{t\geq 0}\big\| E\big(\Xi(t)|\mathcal{G}_{t-n,t+n}\big) - \Xi(t)\big\|_q, \tag{2.23}$$

respectively, where $\mathcal{G}_{st} \subset \mathcal{G}_{s't'}$ if $s' \leq s$ and $t' \geq t$. Then setting $\mathcal{F}_l = \mathcal{G}_{-\infty,l}$ we obtain by the contraction property of conditional expectations that

$$\begin{aligned}\beta_q(n) &\geq \sup_{t\geq 0}\big\| E\big(\Xi(t)|\mathcal{G}_{t-n,t+n}\big) - \Xi(t) + \Xi(t) - E\big(\Xi(t)|\mathcal{G}_{-\infty,[t]+n+1}\big)\big\|_q \\ &\geq \zeta_q(n+1) - \beta_q(n) \quad \text{i.e.} \quad \beta_q(n) \geq \frac{1}{2}\zeta_q(n+1).\end{aligned} \tag{2.24}$$

Furthermore,

$$\begin{aligned}&\big\| g\big(\Xi(n+t), \Xi(n+t+s)\big) - g\big(E\big(\Xi(n+t)|\mathcal{G}_{n+t-[n/2],n+t+[n/2]}\big), \\ &\quad E\big(\Xi(n+t+s)|\mathcal{G}_{n+t+s-[n/2],n+t+s+[n/2]}\big)\big)\big\|_p \leq 2|g|_\kappa \beta^\kappa_{p\kappa}\big([n/2]\big),\end{aligned}$$

and so

$$\eta_{p,\kappa}(n) \leq \big(\varpi_p\big([n/2]\big) + 2\beta^\kappa_{p\kappa}\big([n/2]\big)\big)|g|_\kappa. \tag{2.25}$$

Thus, appropriate conditions on decay of coefficients ϖ_p and β_q as in [25] yield corresponding conditions on $\eta_{p,\kappa}$ and ζ_q. The other direction does not hold true but still it turns out that most of the technique from [25] can be employed in our circumstances, as well.

The conditions of Theorem 2.2 hold true for many important stochastic processes. In the continuous time case they are satisfied when, for instance, $\Xi(t) = f(\Upsilon(t))$ where $\Upsilon(t)$ is either an irreducible continuous time finite state Markov chain or a nondegenerate diffusion process on a compact manifold while f is a Hölder continuous vector function. In the discrete time case we can take, for instance, $\Xi(n) = f(\Upsilon(n))$ with $\Upsilon(n)$ being a Markov chain satisfying the Doeblin condition (see, for instance, [16], pp. 367–368). In all these examples $\eta_{p,\kappa}(n)$ and $\zeta_q(n)$ decay in n exponentially fast while (2.6) requires much less. In fact, in both cases $\Xi(t)$ may depend on whole paths of a Markov process Υ assuming only certain weak dependence on their tails.

Important classes of processes satisfying our conditions come from dynamical systems. In Section 6 we take $\Xi(t) = \Xi(t,z) = g(F^t z)$ where F^t is a C^2 Anosov flow (see [24]) on a compact manifold M whose stable and unstable foliations are jointly nonintegrable and g is a Hölder continuous $\wp$-vector function on M. It turns out that if we take the initial point z on an element S of a Markov family (see Section 6) introduced in [8] distributed there at random according to a probability measure equivalent to the volume on S then Assumption 2.1 can be verified. This does not yield though a desirable limit theorem where the initial point is taken at random on the whole manifold M distributed according to the Sinai–Ruelle–Bowen (SRB) measure (or the normalized Riemannian volume). We observe that a suspension representation of Anosov flows employed in [20] to derive limit theorems in the conventional averaging setup does not work in our situation because $F^{q_i(t)}x$, $i = 1, \dots, \ell$ arrive at the ceiling of the suspension at different times for different i's.

In the discrete time case there are several important classes of dynamical systems where our conditions can be verified. First, for transformations where symbolic representations via Markov partitions are available (Axiom A diffeomorphisms (see [3]) and expanding endomorphisms, some one-dimensional maps e.g. the Gauss map (see [15]) etc.) we can rely on standard mixing and approximation assumptions based on two parameter families of σ-algebras as in (2.22) and (2.23). On the other hand, for many transformations Markov partitions are not available but still it is possible to construct one parameter increasing or decreasing filtration of σ-algebras so that our conditions can be verified. For some classes of noninvertible transformations F it is possible to choose an appropriate initial σ-algebra $\mathcal{F}_0$ such that $F^{-1}\mathcal{F}_0 \subset \mathcal{F}_0$ and then to define a decreasing filtration $\mathcal{F}_i = F^{-i}\mathcal{F}_0$ (see [26] and [13]). Passing to the natural extension as in Remark 3.12 of [13] we can turn to an increasing filtration and to verify our conditions. On the other hand, our results can be derived under appropriate conditions with respect to decreasing families of σ-algebras.

Namely, let $\mathcal{F}\supset\mathcal{F}_0\supset\mathcal{F}_1\supset\mathcal{F}_2\supset\cdots$ and define mixing and approximation coefficients by

$$\eta_{p,\kappa,s}(n)=\sup_{t\ge s}\big\{\big\|E\big(g\big(\Xi(t),\Xi(t-s)\big)|\mathcal{F}_{[t]+n}\big)-E_Pg\big(\Xi(t),\Xi(t-s)\big)\big\|_p:\ g=g(\xi,\tilde\xi),|g|_\kappa\le 1\big\},\qquad \eta_{p,\kappa}(n)=\eta_{p,\kappa,0}(n)\tag{2.26}$$

and

$$\zeta_q(n)=\sup_{t\ge n}\big\|E\big(\Xi(t)|\mathcal{F}_{[t]-n}\big)-\Xi(t)\big\|_q.\tag{2.27}$$

Then under Assumption 2.1 we can rely on estimates of Section 3 below and in place of martingales there arrive at reverse martingales and employ a limit theorem for the latter.

Remark 2.3. *If $\bar B\equiv 0$ then according to Theorem 2.2 the process $X^\varepsilon(t)$ is very close to its initial point on the time interval of order $1/\varepsilon$. Thus, in order to see fluctuations of order 1 it makes sense to consider longer time and to deal with $V^\varepsilon(t)=X^\varepsilon(t/\varepsilon^2)$. Under the stronger condition $\int B(x,\xi_1,\dots,\xi_\ell)\,\mathrm{d}\mu(\xi_\ell)\equiv 0$ it is not difficult to mimic the proofs in [19] and [4] relying on the technique of Sections 3 and 4 below in order to obtain that $V^\varepsilon(t),t\in[0,T]$ converges weakly as $\varepsilon\to 0$ to a diffusion process with parameters obtained in the same way as in [19] and [4]. It is not clear whether, in general, this result still holds true assuming only that $\bar B\equiv 0$. Though most of the required estimates still go through in the latter case a convergence of V^ε to a Markov process seems to be problematic in a general nonconventional averaging setup.*

3. Estimates and martingale approximation

The proof of Theorem 2.2 will employ a modification of the machinery developed in [25]. First, we have to study the asymptotical behavior as $\varepsilon\to 0$ of

$$G_i^\varepsilon(t)=\sqrt{\varepsilon}\int_0^{\tau_i(t)/\varepsilon}B_i\big(\bar Z(\varepsilon s),\Xi\big(q_1(s)\big),\dots,\Xi\big(q_i(s)\big)\big)\,\mathrm{d}s\tag{3.1}$$

which is obtained from the definition (2.21) by the change of variables $s\to s/\varepsilon$ and where $\tau_i(t)=t/\alpha_i$ for $i=1,\dots,k$ and $\tau_i(t)=t$ for $i=k+1,\dots,\ell$. Observe that if $\frac{1}{N+1}\le\varepsilon\le\frac{1}{N}$ and $N\ge 1$ then by (2.7),

$$\big|G_i^\varepsilon(t)-G_i^{1/N}(t)\big|\le\frac{2Ktd}{\sqrt N}\tag{3.2}$$

and so it suffices to study the asymptotical behavior of $G_i^{1/N}$ as $N\to\infty$. Set

$$I_{i,N}(n)=\int_n^{n+1}B_i\big(\bar Z(s/N),\Xi\big(q_1(s)\big),\dots,\Xi\big(q_i(s)\big)\big)\,\mathrm{d}s.\tag{3.3}$$

In view of (2.7) the asymptotical behavior of $G_i^{1/N}$ as $N\to\infty$ is the same as of $N^{-1/2}S_{i,N}(t)$ where

$$S_{i,N}(t)=\sum_{n=0}^{[N\tau_i(t)]}I_{i,N}(n).\tag{3.4}$$

There are two obstructions for applying directly the results of [25] to the sum (3.4). First, unlike [25] the integrand in (3.3) depends on the “slow time” s/N. Secondly, our mixing and approximation coefficients look differently from the corresponding coefficients in [25]. Still, it turns out that these obstructions can be dealt with and after minor modifications the method of [25] start working in our situation, as well. Namely, the dependence on the “slow time” being deterministic will not prevent us from making estimates similar to [25] while dependence of I_i^N on N will just require us to deal with martingale arrays which creates no problems as long as we obtain appropriate limits of variances and covariances. Concerning the second obstruction we observe that one half of the approximation estimate from [25] is contained in the coefficient ζ_p while another half is hidden in the coefficient $\eta_{p,\kappa}$ which also suffices for required mixing estimates.

We explain next more precisely why estimates similar to [25] hold true in our circumstances, as well. Let $f(\psi,\xi,\tilde{\xi})$ be a function on $\mathbb{R}^\varrho\times\mathbb{R}^\wp\times\mathbb{R}^\wp$ such that for any $\psi,\psi'\in\mathbb{R}^\varrho$ and $\xi,\tilde{\xi},\xi',\tilde{\xi}'\in\mathbb{R}^\wp$,

$$\bigl|f(\psi,\xi,\tilde{\xi})-f\bigl(\psi',\xi',\tilde{\xi}'\bigr)\bigr|\le C\bigl(\bigl|\psi-\psi'\bigr|^\kappa+\bigl|\xi-\xi'\bigr|^\kappa+\bigl|y-y'\bigr|^\kappa\bigr)\quad\text{and}\quad\bigl|f(\psi,\xi,\tilde{\xi})\bigr|\le C.\tag{3.5}$$

Then setting $g(\psi)=E_P f(\psi,\Xi(0),\Xi(s))$ we obtain from (2.1) and (2.3) that for all $u,v\ge 0$ and $n\in\mathbb{N}$,

$$\bigl\|E\bigl(f\bigl(\psi,\Xi(n+u),\Xi(n+u+v)\bigr)|\mathcal{F}_{[u]}\bigr)-g(\psi)\bigr\|_p\le C\eta_{p,\kappa,v}(n).\tag{3.6}$$

Let $h(\psi,\omega)=E(f(\psi,\Xi(n+u),\Xi(n+u+v))|\mathcal{F}_{[u]})-g(\psi)$. Then by (3.5) we can choose a version of $h(\psi,\omega)$ such that with probability one simultaneously for all $\psi,\psi'\in\mathbb{R}^\varrho$,

$$\bigl|h(\psi,\omega)-h\bigl(\psi',\omega\bigr)\bigr|\le 2C\bigl|\psi-\psi'\bigr|^\kappa.\tag{3.7}$$

Since, in addition, $\|h(\psi,\omega)\|_p\le C\eta_{p,\kappa}(n)$ by (3.6) for all $\psi\in\mathbb{R}^\varrho$, we obtain by Theorem 3.4 from [25] that for any random ϱ-vector $\Psi=\Psi(\omega)$,

$$\bigl\|h\bigl(\Psi(\omega),\omega\bigr)\bigr\|_a\le cC\bigl(\eta_{p,\kappa,v}(n)\bigr)^{1-\varrho/(p\theta)}\bigl(1+\|\Psi\|_m\bigr),\tag{3.8}$$

where $\frac{\varrho}{p}<\theta<\kappa$, $\frac{1}{a}\ge\frac{1}{p}+\frac{1}{m}$ and $c=c(\varrho,p,\kappa,\theta)>0$ depends only on parameters in brackets. Since

$$h\bigl(\tilde{\Psi}(\omega),\omega\bigr)=E\bigl(f\bigl(\tilde{\Psi},\Xi(n+u),\Xi(n+u+v)\bigr)|\mathcal{F}_{[u]}\bigr)(\omega)\quad\text{a.s.}\tag{3.9}$$

provided $\tilde{\Psi}$ is $\mathcal{F}_{[u]}$-measurable we obtain from (3.6)–(3.9) together with the Hölder inequality (cf. Corollary 3.6(ii) in [25]) that,

$$\begin{aligned}&\bigl\|E\bigl(f\bigl(\Psi,\Xi(n+u),\Xi(n+u+v)\bigr)|\mathcal{F}_{[u]}\bigr)-g(\Psi)\bigr\|_a\\&\quad\le C\bigl(\eta_{p,\kappa,v}(n)\bigr)^{1-\varrho/(p\theta)}\bigl(1+\|\Psi\|_m\bigr)+2C\bigl\|\Psi-E(\Psi|\mathcal{F}_{[u]})\bigr\|_q^\delta\end{aligned}\tag{3.10}$$

provided $\frac{1}{a}\ge\frac{1}{p}+\frac{2}{m}+\frac{\delta}{q}$.

We apply the above estimates in two cases. First, when $f(\psi,\xi,\tilde{\xi})=f(\psi,\xi)=B_i(x,\xi_1,\dots,\xi_i)$ with $\psi=(\xi_1,\dots,\xi_{i-1})\in\mathbb{R}^{(i-1)\wp}$, $\xi=\xi_i\in\mathbb{R}^\wp$, $n=[(q_i(t)-q_{i-1}(t))/2]$, $u=q_i(t)-n$ and $\Psi=(\Xi(q_1(t)),\Xi(q_2(t)),\dots,\Xi(q_{i-1}(t))$. In the second case $f(\psi,\xi,\tilde{\xi})=B_i(x,\xi_1,\dots,\xi_i)B_j(y,\xi_1',\dots,\xi_j')$ with $\psi=(\xi_1,\dots,\xi_{i-1},\xi_1',\dots,\xi_{j-1}')\in\mathbb{R}^{(i+j-2)\wp}$, $\xi=\xi_i$, $\tilde{\xi}=\xi_j'\in\mathbb{R}^\wp$, $n=[(\min(q_i(t),q_j(s))-\max(q_{i-1}(t),q_{j-1}(s)))/2]$ when $n>0$, $u=\min(q_i(t),q_j(s))-n$ and $\Psi=(\Xi(q_1(t)),\dots,\Xi(q_{i-1}(t)),\Xi(q_1(s)),\dots,\Xi(q_{j-1}(s)))$. The estimates for the first case are used for martingale approximations while the second case emerges when computing covariances.

Since $\int B_i(x,\xi_1,\dots,\xi_{i-1},\xi_i)\,\mathrm{d}\mu(\xi_i)=0$ we obtain by (3.10) the estimate

$$\bigl\|E\bigl(B_i\bigl(\Xi\bigl(q_1(t)\bigr),\dots,\Xi\bigl(q_i(t)\bigr)\bigr)|\mathcal{F}_{[q_i(t)]-n}\bigr)\bigr\|_a\le C\bigl(\bigl(\eta_{p,\kappa}(n)\bigr)^{1-\varrho(p\theta)}+\bigl(\zeta_q(n)\bigr)^\delta\bigr)\tag{3.11}$$

for some $C>0$ independent of t where $n=n_i(t)=[(q_i(t)-q_{i-1}(t))/2]$. Next, for any $x\in\mathbb{R}^\varrho$, $\xi_1,\dots,\xi_{i-1}\in\mathbb{R}^\varrho$ and $r=1,2,\dots$ set

$$B_{i,r}\bigl(x,\xi_1,\dots,\xi_{i-1},\Xi(t)\bigr)=E\bigl(B_i\bigl(x,\xi_1,\dots,\xi_{i-1},\Xi(t)\bigr)|\mathcal{F}_{[t]+r}\bigr)\quad\text{and}\quad\Xi_r(t)=E\bigl(\Xi(t)|\mathcal{F}_{[t]+r}\bigr).$$

Then by (2.4) and (2.7) together with the Hölder inequality,

$$\begin{aligned}&\bigl\|B_i\bigl(x,\xi_1,\dots,\xi_{i-1},\Xi(t)\bigr)-B_{i,r}\bigl(x,\xi_1,\dots,\xi_{i-1},\Xi(t)\bigr)\bigr\|_q\\&\quad\le 2\bigl\|B_i\bigl(x,\xi_1,\dots,\xi_{i-1},\Xi(t)\bigr)-B_i\bigl(x,\xi_1,\dots,\xi_{i-1},E\bigl(\Xi(t)|\mathcal{F}_{[t]+r}\bigr)\bigr)\bigr\|_q\\&\quad\le 2Kd\bigl\|\bigl|\Xi(t)-E\bigl(\Xi(t)\bigr)|\mathcal{F}_{[t]+r}\bigr)\bigr|^\kappa\bigr\|_q\le 2Kd\zeta_q^\delta(r).\end{aligned}\tag{3.12}$$

This together with the last part of Theorem 3.4 in [25] yields that

$$\left\| B_i\big(x, \Xi\big(q_1(t)\big), \dots, \Xi\big(q_i(t)\big)\big) - B_{i,r}\big(x, \Xi_r\big(q_1(t)\big), \dots, \Xi_r\big(q_{i-1}(t)\big)\big)\right\|_a \le c\zeta_q^{\delta}(r) \tag{3.13}$$

provided $\frac{1}{a} \ge \frac{1}{p} + \frac{2}{m} + \frac{\delta}{q}$ and $\delta < \min(\kappa, 1 - \frac{d}{p\kappa})$ where $c = c(\delta, a, p, q) > 0$ depends only on the parameters in brackets. Set

$$b_{ij}^{l,m}(x, y; s, t) = E\big(B_i^{(l)}\big(x, \Xi\big(q_1(s)\big), \dots, \Xi\big(q_i(s)\big)\big)B_j^{(m)}\big(y, \Xi\big(q_1(t)\big), \dots, \Xi\big(q_j(t)\big)\big)\big),$$

where, recall, $B_i^{(l)}$ is the l-th component of the d-vector B_i. Now, by (2.7), (3.11) and (3.13),

$$\left|b_{ij}^{l,m}(x, y; s, t)\right| \le C\big(\big(\eta_{p,\kappa}(n)\big)^{1-\varrho/(p\theta)} + \big(\zeta_q(n)\big)^{\delta}\big), \tag{3.14}$$

where $C > 0$ does not depend on $s, t \ge 0$ and $n = n_{ij}(s, t) = \max(\hat{n}_{ij}(s, t), \hat{n}_{ji}(t, s))$ with $\hat{n}_{ij}(s, t) = [\frac{1}{2}\min(q_i(s) - q_j(t), q_i(s) - q_{i-1}(s))]$.

Now, set

$$\begin{aligned}
I_{i,N,r}(n) &= \int_{n-1}^{n} B_{i,r}\big(\bar{Z}(s/N), \Xi_r\big(q_1(s)\big), \dots, \Xi_r\big(q_{i-1}(s)\big)\big)\,\mathrm{d}s,\\
S_{i,N,r}(t) &= \sum_{n=1}^{[N\tau_i(t)]} I_{i,N,r}(n),\\
R_{i,r}(m) &= \sum_{l=m+1}^{\infty} E\big(I_{i,N,r}(l)|\mathcal{F}_{m+r}\big),\\
D_{i,N,r}(m) &= I_{i,N,r}(m) + R_{i,r}(m) - R_{i,r}(m-1) \quad \text{and} \quad M_{i,N,r}(t) = \sum_{n=1}^{[N\tau_i(t)]} D_{i,N,r}(n).
\end{aligned} \tag{3.15}$$

In view of (2.6) and (3.11) applied with $a = 2$ we see that the series for $R_{i,r}(m)$ converges in L^2, $D_{i,N,r}(m)$ is $\mathcal{F}_{m+r}$-measurable and since $E(D_{i,N,r}(m)|\mathcal{F}_{m-1+r}) = 0$ we obtain that $\{D_{i,N,r}(m), \mathcal{F}_{m+r}\}_{0\le m\le[N\tau_i(T)]}$ is a martingale differences array. Next, we rely on (3.11) and the inequality

$$\left|S_{i,N,r}(t) - M_{i,N,r}(t)\right| \le \left|R_{i,r}\big(\big[N\tau_i(t)\big]\big)\right| + \left|R_{i,r}(0)\right|$$

to observe that as $N \to \infty$ the limiting behavior of $N^{-1/2}S_{i,N,r}$ as $N \to \infty$ is the same as of $N^{-1/2}M_{i,N,r}$. Once we derive that appropriate covariances converge as $N \to \infty$, which will be done in the next section, we can invoke for the latter expression a version of the functional central limit theorem for martingale arrays (see, for instance, Section 2 in Ch. VIII of [17]) yielding that $N^{-1/2}M_{i,N,r}$, and so also $N^{-1/2}S_{i,N,r}$ converge as $N \to \infty$ to a d-dimensional Gaussian process with independent increments. Next, we write

$$S_{i,N}(t) = S_{i,N,1}(t) + \sum_{r=1}^{\infty}\big(S_{i,N,2^r}(t) - S_{i,N,2^{r-1}}(t)\big) \tag{3.16}$$

and relying on uniform moment estimates of the terms in this series similar to Proposition 5.9 of [25] we obtain that also $N^{-1/2}S_{i,N}$ converges in distribution as $N \to \infty$ to a d-dimensional Gaussian process with independent increments. Next, by a version of the Cramér–Wold argument (see Corollary 5.7 in [25]) we obtain that $(S_{i,N}, i = 1, \dots, \ell)$ converges in distribution as $N \to \infty$ to a ℓd-dimensional Gaussian process with independent increments while covariances computations in the next section show that last $\ell - k$ random d-vectors of the limiting process are independent of each other and of other random d-vectors there. Thus, we will end up with limiting d-dimensional

Gaussian processes $G_i^0, i=1,\dots,\ell$ with independent increments such that $(G_i^0, i=1,\dots,\ell)$ is an ℓd-dimensional Gaussian process and $G_{k+1}^0,\dots,G_\ell^0$ are independent of each other and of G_i^0 with $i\le k$. Now we write

$$G^0(t)=\sum_{i=1}^k G_i^0(it)+\sum_{i=k+1}^{\ell} G_i^0(t)=\sum_{j=1}^k\sum_{i=1}^k \lambda_{ij}\big(G_j^0(it)-G_j^0\big((i-1)t\big)\big),$$

where $\lambda_{ij}=1$ if $i\le j$ and $\lambda_{ij}=0$, otherwise, obtaining from the above that G^0 is a Gaussian d-dimensional process (for more details see [25]).

Observe that under a bit stronger assumptions we could employ in our setup another martingale approximation construction from [23] which is based on increasing blocks and negligible gaps between them and which does not require r-approximations as above. In order to complete this programm it remains only to compute limiting covariances as in Section 4 of [25] taking care also of the slow time s/N entering (3.3) and (3.15).

4. Limiting covariances

In this section we show the existence and compute the limit as $\varepsilon\to 0$ of the expression

$$E\big(G_{i,l}^\varepsilon(s)G_{j,m}^\varepsilon(t)\big)=\varepsilon\int_0^{\tau_i(s/\varepsilon)}\int_0^{\tau_j(t/\varepsilon)} b_{ij}^{l,m}\big(\bar Z(\varepsilon u),\bar Z(\varepsilon v);u,v\big)\,\mathrm{d}u\,\mathrm{d}v. \tag{4.1}$$

We start with showing that there exists a constant $C>0$ such that for all $t\ge s>0, l=1,\dots,d,\ N\ge 1$ and $i=1,\dots,\ell$,

$$\sup_{N\ge 1} E\big|G_{i,l}^{1/N}(t)-G_{i,l}^{1/N}(s)\big|^2\le C(t-s). \tag{4.2}$$

In order to obtain (4.2) we note that by (2.9) and (2.10) for $t\ge s$,

$$q_i(t)-q_i(s)\ge \alpha_i(t-s)\quad\text{and}\quad q_i(t)-q_{i-1}(t)\ge\alpha_{i-1}t\quad\text{when } i=2,\dots,k \tag{4.3}$$

and for any $\gamma>0$ there exists t_γ such that for all $t\ge t_\gamma$ and $i=k+1,\dots,\ell$,

$$q_i(t)-q_i(s)\ge (t-s)+\gamma^{-1}\quad\text{and}\quad q_i(t)-q_{i-1}(t)\ge t+\gamma^{-1}. \tag{4.4}$$

Now (4.2) follows from (2.6), (3.14), (4.1), (4.3) and (4.4). Observe, that by (3.2) and (4.1) if $\frac{1}{N+1}\le\varepsilon\le\frac{1}{N}$ then

$$\big|EG_{i,l}^\varepsilon(s)G_{j,m}^\varepsilon(t)-EG_{i,l}^{1/N}(s)G_{j,m}^{1/N}(t)\big|\le\frac{4KdC\sqrt{\mathcal{T}}}{\sqrt{N}},$$

and so it suffices to study (4.1) as $\varepsilon=\frac{1}{N}$ and $N\to\infty$.

Next, we claim that if $i>j$ and $i>k$ then the limit in (4.1) as $\frac{1}{\varepsilon}=N\to\infty$ exists and equals zero. Indeed, in this case for any small $\gamma>0$ with $\gamma T\le s$,

$$\big|EG_{i,l}^{1/N}(s)G_{j,m}^{1/N}(t)\big|\le I_1+I_2, \tag{4.5}$$

where by (4.2),

$$I_1=\big|EG_{i,l}^{1/N}(\gamma\mathcal{T})G_{j,m}^{1/N}(t)\big|\le\big(E\big(G_{i,l}^{1/N}(\gamma\mathcal{T})\big)^2\big)^{1/2}\big(E\big(G_{j,m}^{1/N}(t)\big)^2\big)^{1/2}\le C\sqrt{\gamma\mathcal{T}t} \tag{4.6}$$

and by (3.14),

$$\begin{aligned}I_2&=\big|E\big(G_{i,l}^{1/N}(s)-G_{i,l}^{1/N}(\gamma\mathcal{T})\big)G_{j,m}^{1/N}(t)\big|\\&=\frac{1}{N}\int_{\gamma\mathcal{T}N}^{sN}\mathrm{d}u\int_0^{\tau_j(tN)} b_{ij}^{l,m}\big(\bar Z(u/N),\bar Z(v/N);u,v\big)\,\mathrm{d}v\le\frac{C}{N}\int_{\gamma\mathcal{T}N}^{sN}\mathrm{d}u\int_0^{\tau_j(tN)}\rho_{ij}(u,v)\,\mathrm{d}v,\end{aligned} \tag{4.7}$$

where

$$\rho_{ij}(u,v)=\big(\eta_{p,\kappa}\big(n_{ij}(u,v)\big)\big)^{1-\varrho/(p\theta)}+\big(\zeta_q\big(n_{ij}(u,v)\big)\big)^{\delta} \tag{4.8}$$

with $n_{ij}(s,t)$ defined after (3.14). It follows from (2.6), (2.9), (2.10) and (4.8) that for any $\gamma>0$ there exists N_γ such that whenever $N\ge N_\gamma$ and $v\in[0,\mathcal{T}N]$ (cf. Proposition 4.5 in [25]),

$$\int_{\gamma\mathcal{T}N}^{sN}\rho_{ij}(u,v)\,\mathrm{d}u\le\gamma$$

and so $I_2\le C\mathcal{T}\gamma$. Since $\gamma>0$ is arbitrary this together with (4.5) and (4.6) yields that for all $l,m=1,\dots,d$, $i>k$ and $j<i$,

$$\lim_{N\to\infty}EG_{i,l}^{1/N}(s)G_{j,m}^{1/N}(t)=0. \tag{4.9}$$

Next, we claim that when $i>k$ then also for all $l,m=1,\dots,d$,

$$\lim_{N\to\infty}EG_{i,l}^{1/N}(s)G_{i,m}^{1/N}(t)=0. \tag{4.10}$$

Indeed, by (3.14) and (4.8) for $t\ge s$,

$$\big|EG_{i,l}^{1/N}(s)G_{i,m}^{1/N}(t)\big|\le\frac{1}{N}\int_0^{sN}\mathrm{d}u\int_0^{tN}\rho_{ii}(u,v)\,\mathrm{d}v=I_3+I_4, \tag{4.11}$$

where

$$I_3=\frac{2}{N}\int_0^{sN}\mathrm{d}u\int_u^{sN}\rho_{ii}(u,v)\,\mathrm{d}v\quad\text{and}\quad I_4=\frac{1}{N}\int_0^{sN}\mathrm{d}u\int_{sN}^{tN}\rho_{ii}(u,v)\,\mathrm{d}v.$$

Now

$$\begin{aligned}I_3&=\frac{2}{N}\int_0^{sN}\mathrm{d}u\int_u^{u+\gamma}\rho_{ii}(u,v)\,\mathrm{d}v+\frac{2}{N}\int_0^{\gamma N}\mathrm{d}u\int_{u+\gamma}^{sN}\rho_{ii}(u,v)\,\mathrm{d}v\\&\quad+\frac{2}{N}\int_{\gamma N}^{sN}\mathrm{d}u\int_{u+\gamma}^{sN}\rho_{ii}(u,v)\,\mathrm{d}v\le C\big(s\gamma+\gamma+s\beta_\gamma(\gamma N)\big)\end{aligned} \tag{4.12}$$

for some $C>0$ where by (2.6) and (2.10) for any $\gamma>0$,

$$\beta_\gamma(M)=\sup_{u\ge M}\int_{u+\gamma}^{\infty}\rho_{ii}(u,v)\,\mathrm{d}v<\infty\quad\text{and}\quad\lim_{M\to\infty}\beta_\gamma(M)=0. \tag{4.13}$$

Next,

$$\begin{aligned}I_4&=\frac{1}{N}\int_0^{sN}\mathrm{d}u\int_{sN}^{sN+\gamma}\rho_{ii}(u,v)\,\mathrm{d}v\\&\quad+\frac{1}{N}\int_0^{sN}\mathrm{d}u\int_{sN+\gamma}^{tN}\rho_{ii}(u,v)\,\mathrm{d}v\le Cs\gamma+Cs\beta_s(N).\end{aligned} \tag{4.14}$$

Finally, (4.10) follows from (4.11)–(4.14) letting, first, $N\to\infty$ and then $\gamma\to 0$.

In order to compute the limit as $\frac{1}{\varepsilon}=N\to\infty$ of (4.1) for $i,j=1,2,\dots,k$ we recall an argument of Lemma 3.1 from [18] which yields that if uniformly in $\sigma\ge 0$ and x,y from a compact set the limit

$$\lim_{N\to\infty}\frac{1}{N}\int_{\sigma/\alpha_i}^{(\sigma+sN)/\alpha_i}\mathrm{d}u\int_{\sigma/\alpha_j}^{(\sigma+sN)/\alpha_j}b_{ij}^{lm}(x,y;u,v)\,\mathrm{d}u\,\mathrm{d}v=sD_{ij}^{l,m}(x,y) \tag{4.15}$$

exists and has the form of the right hand side with a continuous $D_{i,j}^{l,m}$ then the limit (4.1) exists, as well, and it has the form

$$\lim_{N\to\infty} E\big(G_{i,l}^{1/N}(s)G_{j,m}^{1/N}(t)\big) = \int_0^{\min(s,t)} D_{ij}^{l,m}\big(\bar{Z}(u), \bar{Z}(u)\big)\,\mathrm{d}u. \tag{4.16}$$

Namely, set $M = M(N) = [N^{2/3}]$ and let $s_\iota = \frac{\iota s}{M}$, $\iota = 0, 1, \dots, M-1$. Assume also that $s \le t$. Let

$$A_N = \bigcup_{\iota=0}^{M-1} A_{N,\iota} \quad \text{with } A_{N,\iota} = \left\{(u,v)\colon\ s_\iota N \le u, v < \left(s_\iota + \frac{s}{M}\right)N\right\}$$

and $B_N = \{(u,v)\colon\ 0 \le u \le sN, 0 \le v \le tN\} \setminus A_N$. Then

$$EG_{i,l}^{1/N}(s)G_{j,m}^{1/N}(t) = I_5 + I_6, \tag{4.17}$$

where

$$I_5 = \frac{1}{Nij}\int_{B_N} b_{ij}^{lm}\big(\bar{Z}(u/N), \bar{Z}(v/N), u/i, v/j\big)\,\mathrm{d}u\,\mathrm{d}v$$

and

$$I_6 = \frac{1}{Nij}\int_{A_N} b_{ij}^{lm}\big(\bar{Z}(u/N), \bar{Z}(v/N), u/i, v/j\big)\,\mathrm{d}u\,\mathrm{d}v.$$

Now, by (3.14) and (4.8),

$$\begin{aligned}|I_5| \le \frac{C}{Nij}\Bigg(&\sum_{\iota=0}^{M-1}\int_0^{s_\iota/\varepsilon}\int_{s_\iota/\varepsilon}^{(s_\iota+s/M)/\varepsilon}\big(\rho_{ij}(u/\alpha_i, v/\alpha_j)\\ &+ \rho_{ji}(u/\alpha_j, v/\alpha_i)\big)\,\mathrm{d}u\,\mathrm{d}v + \int_0^{s_\iota/\varepsilon}\int_{s_\iota/\varepsilon}^{t_\iota/\varepsilon}\rho_{ij}(u/\alpha_i, v/\alpha_j)\,\mathrm{d}u\,\mathrm{d}v\Bigg).\end{aligned} \tag{4.18}$$

Observe that by the definition of $n_{ij}(u,v)$ after (3.14) we can write for $i, j = 1, \dots, k$,

$$\rho_{ij}(u/\alpha_i, v/\alpha_j) = \zeta\big(|u-v|\big), \tag{4.19}$$

where $\zeta \ge 0$ satisfies $\int_0^\infty w\zeta(w)\,\mathrm{d}w < \infty$. Integrating by parts we obtain for any $V \ge U \ge 0$,

$$\int_0^U \mathrm{d}u \int_U^V \zeta(v-u)\,\mathrm{d}v \le \int_0^U \mathrm{d}u \int_{U-u}^\infty \zeta(w)\,\mathrm{d}w = \int_0^U r\zeta(r)\,\mathrm{d}r \le \int_0^\infty r\zeta(r)\,\mathrm{d}r. \tag{4.20}$$

This together with (2.6), (4.18) and (4.19) gives by the choice of $M = M(N)$ that

$$|I_5| \le \tilde{C}\frac{M}{N\alpha_i\alpha_j} \to 0 \quad \text{as } N \to \infty \tag{4.21}$$

for some $\tilde{C} > 0$ independent of M and N.

Next,

$$I_6 = \frac{1}{M\alpha_i\alpha_j}\sum_{\iota=0}^{M-1} J_{M,N}(\iota) + I_7, \tag{4.22}$$

where

$$J_{M,N}(\iota)=\frac{M}{N}\int_{s_\iota\le u,v<(s_\iota+\frac{s}{M})N} b_{ij}^{lm}\left(\bar Z(s_\iota),\bar Z(s_\iota);\frac{u}{\alpha_i},\frac{v}{\alpha_j}\right)\mathrm{d}u\,\mathrm{d}v$$

and by (2.7) and the choice of $M=M(N)$,

$$|I_7|\le Cs^3NM^{-2}\to 0\quad\text{as } N\to\infty, \tag{4.23}$$

where $C>0$ does not depend on s, N and M. By (4.15) we obtain that

$$\left|J_{M,N}(\iota)-s\alpha_i\alpha_j D_{ij}^{l,m}\big(\bar Z(s_\iota),\bar Z(s_\iota)\big)\right|\to 0\quad\text{as } N\to\infty \tag{4.24}$$

and so

$$\left|I_6-\int_0^s D_{ij}^{l,m}\big(\bar Z(u),\bar Z(u)\big)\,\mathrm{d}u\right|\to 0\quad\text{as } N\to\infty \tag{4.25}$$

completing the proof of (4.16).

In order to describe $D_{ij}^{l,m}(x,y), i,j\le k$ consider all indices $1\le i'_1<i'_2<\dots<i'_{\iota_{ij}}=i$ and $1\le j'_1<j'_2<\dots<j'_{\iota_{ij}}=j$ such that there exist $0<\rho_1<\dots<\rho_{\iota_{ij}}=1$ satisfying $\alpha_{i'_l}\rho_l,\alpha_{j'_l}\rho_l\in\{\alpha_1,\dots,\alpha_k\}$ for all $l=1,\dots,\iota_{ij}$. Define

$$\begin{aligned}a_{ij}^{l,m}(x,y;s_1,\dots,s_{\iota_{ij}})&=\int B_i^{(l)}(x,\xi_1,\dots,\xi_i)B_j^{(m)}(y,\tilde\xi_1,\dots,\tilde\xi_j)\\&\quad\times\prod_{\beta=1}^{\iota_{ij}}\mathrm{d}\mu_{s_\beta}(\xi_{i'_\beta},\tilde\xi_{j'_\beta})\prod_{i_\gamma\notin\{i'_1,\dots,i'_{\iota_{ij}}\},1\le i_\gamma<i}\mathrm{d}\mu(\xi_{i_\gamma})\prod_{j_\zeta\notin\{j'_1,\dots,j'_{\iota_{ij}}\},1\le j_\zeta<j}\mathrm{d}\mu(\xi_{j_\zeta}).\end{aligned}\tag{4.26}$$

Then in the same way as in the proof of Lemma 4.4 from [25] (see also Section 6 there) we obtain relying on (2.6), (3.10) and (3.14) that

$$\lim_{u_N,v_N\to\infty,\alpha_iu_N-\alpha_jv_N=w} b_{ij}^{l,m}(x,y;u_N,v_N)=a_{ij}^{l,m}(x,y;\rho_1w,\rho_2w,\dots,\rho_{\iota_{ij}}w). \tag{4.27}$$

This is the only place where we need Assumption 2.1 for $\eta_{p,\kappa,s}$ with $s>0$. It follows similarly to Section 6 of [25] that the limit (4.15) exists and it can be written in the form

$$D_{ij}^{l,m}(x,y)=\frac{1}{\alpha_i\alpha_j}\int_{-\infty}^{\infty}a_{ij}^{l,m}(x,y;\rho_1w,\rho_2w,\dots,\rho_{\iota_{ij}}w)\,\mathrm{d}w. \tag{4.28}$$

Roughly speaking, we derive (4.27) in the following way. First, observe that when $u_N,v_N\to\infty$ so that $\alpha_iu_N-\alpha_jv_N=w$ we can split the collection of random variables $\Xi(\alpha_1u_N),\dots,\Xi(\alpha_iu_N);\Xi(\alpha_1v_N),\dots,\Xi(\alpha_jv_N)$ appearing in $b_{ij}^{l,m}(x,y;u_N,v_N)$ into groups consisting of singletons and pairs such that time differences within each group are bounded while time differences between different groups tend to infinity. Indeed, time differences between any two terms in either sequence of times $\alpha_1u_N,\dots,\alpha_iu_N$ and $\alpha_1v_N,\dots,\alpha_jv_N$ tend to infinity, and so such groups may consist of at most one member from each of these sequences. Next, if the distance $|\alpha_iu_N-\alpha_jv_N|$ remains bounded as $u_N,v_N\to\infty$ then only for pairs $\alpha_{i'},\alpha_{j'}$ with the same ratio the distance $|\alpha_{i'}u_N-\alpha_{j'}v_N|$ remains bounded, as well. But then $\alpha_{i'}=\rho\alpha_i$ and $\alpha_{j'}=\rho\alpha_j$ for some positive $\rho<1$. Now observe that an estimate of the form (3.10) enables us, making only negligible errors, to compute expectation in $b_{ij}^{l,m}$ doing this separately for described above groups of random variables Ξ with large time differences between them (as if they were independent) which leads to the limit (4.27) in the form (4.26).

Collecting the results of Sections 3 and 4 together we conclude that each $G_i^\varepsilon, i=1,\dots,k$ converges weakly as $\varepsilon\to 0$ to the corresponding Gaussian process G_i^0 having independent increments while the process $G_i^\varepsilon, i>k$ converge weakly as $\varepsilon\to 0$ to zero (in the continuous time case we are dealing with now). Moreover, the processes G^ε converge weakly as $\varepsilon \to 0$ to a Gaussian process G^0 (with not necessarily independent increments as an example in [25] shows) having the representation

$$G^0(t) = \sum_{i=1}^{k} G_i^{\varepsilon}(it). \tag{4.29}$$

Furthermore, the covariances of different components $G_i^0(s) = (G_i^{0,1}(s), \dots, G_i^{0,d}(s))$ of this processes are described in view of the above by

$$EG_i^{0,l}(s)G_j^{0,m}(t) = \int_0^{\min(s,t)} D_{ij}^{l,m}\big(\bar{Z}(u), \bar{Z}(u)\big)\,\mathrm{d}u, \tag{4.30}$$

and so by (4.29),

$$EG^{0,l}(s)G^{0,m}(t) = \int_0^{\min(s,t)} A^{l,m}(u)\,\mathrm{d}u, \tag{4.31}$$

where

$$A^{l,m}(u) = \sum_{1 \le i,j \le k} D_{ij}^{l,m}\big(\bar{Z}(iu), \bar{Z}(ju)\big).$$

5. Gaussian approximation of the slow motion and discrete time case

In order to complete the proof of Theorem 2.2 we proceed similarly to [18]. First, we consider the process $H^{\varepsilon}(t)$ which solves the linear equation

$$H^{\varepsilon}(t) = G^{\varepsilon}(t) + \int_0^t \nabla\bar{B}\big(\bar{Z}(s)\big)H^{\varepsilon}(s)\,\mathrm{d}s. \tag{5.1}$$

By (2.7), for some $C > 0$ independent of t and ε,

$$\big|H^{\varepsilon}(t)\big| \le \big|G^{\varepsilon}(t)\big| + C\int_0^t \big|H^{\varepsilon}(s)\big|\,\mathrm{d}s.$$

Then

$$\big||H^{\varepsilon}(t)| - |G^{\varepsilon}(t)|\big| \le C\int_0^t \big|G^{\varepsilon}(s)\big|\,\mathrm{d}s + C\int_0^t \big||H^{\varepsilon}(s)| - |G^{\varepsilon}(s)|\big|\,\mathrm{d}s$$

and by Gronwall's inequality

$$\big|H^{\varepsilon}(t)\big| \le \big|G^{\varepsilon}(t)\big| + C\mathrm{e}^{Ct}\int_0^t \big|G^{\varepsilon}(s)\big|\,\mathrm{d}s. \tag{5.2}$$

It follows from Section 3 that the family of processes $\{G^{\varepsilon}(t), t \in [0, \mathcal{T}]\}$ is tight which together with (5.2) implies that the family of processes $\{H^{\varepsilon}(t), t \in [0, \mathcal{T}]\}$, as well, as the family of pairs $V^{\varepsilon} = \{G^{\varepsilon}, H^{\varepsilon}\}$ are tight.

It follows that any weak limit $V^0 = \{G^0, H^0\}$ of V^{ε} as $\varepsilon \to 0$ must satisfy the equation

$$H^0(t) = G^0(t) + \int_0^t \nabla\bar{B}\big(\bar{Z}(s)\big)H^0(s)\,\mathrm{d}s \tag{5.3}$$

which has a unique solution. Moreover, its solution H^0 is a Gaussian process. Indeed, the equation (5.3) can be solved by successive approximations starting from G^0 so that on each step we will get a Gaussian process (in view of linearity) and the limiting process will be Gaussian, as well. Moreover, H^0 depends linearly on G^0 having an integral representation of the form

$$H^0(t) = G^0(t) + \int_0^t K(t,s)G^0(s)\,\mathrm{d}s \tag{5.4}$$

with a differentiable kernel K (Green's function). The latter follows considering an operator A given by

$$Af(t) = \int_0^t \nabla \bar{B}\big(\bar{Z}(s)\big) f(s)\,\mathrm{d}s$$

which has the supremum norm less than 1 if $t \in [0, \Delta]$ for Δ small enough, and so we can write

$$H^0 = (I - A)^{-1} G^0 = G^0 + \sum_{n=1}^{\infty} A^n G^0.$$

In view of the form of the integral operator A above this representation yields (5.4) on the interval $[0, \Delta]$ and then employing the same argument successively to time itervals $[\Delta, 2\Delta]$, $[2\Delta, 3\Delta], \ldots$ we extend the representation (5.4) for any t.

Observe that

$$\begin{aligned} Q^\varepsilon(t) &= \varepsilon^{-1/2} \int_0^t \Big(B\big(Z_x^\varepsilon(s), \Xi\big(q_1(s/\varepsilon)\big), \ldots, \Xi\big(q_\ell(s/\varepsilon)\big)\big) - \bar{B}\big(\bar{Z}_x(s)\big)\Big)\,\mathrm{d}s \\ &= G^\varepsilon(t) + \int_0^t \nabla_x B\big(Z_x^\varepsilon(s), \Xi\big(q_1(s/\varepsilon)\big), \ldots, \Xi\big(q_\ell(s/\varepsilon)\big)\big) Q^\varepsilon(s)\,\mathrm{d}s + \int_0^t J_1^\varepsilon(s)\,\mathrm{d}s, \end{aligned} \tag{5.5}$$

where

$$\begin{aligned} J_1^\varepsilon(s) = \varepsilon^{-1/2}\Big(& B\big(\bar{Z}_x(s) + \sqrt{\varepsilon} Q^\varepsilon(s), \Xi\big(q_1(s/\varepsilon)\big), \ldots, \Xi\big(q_\ell(s/\varepsilon)\big)\big) \\ & - B\big(\bar{Z}_x(s), \Xi\big(q_1(s/\varepsilon)\big), \ldots, \Xi\big(q_\ell(s/\varepsilon)\big)\big) - \nabla_x B\big(\bar{Z}_x(s), \Xi\big(q_1(s/\varepsilon)\big), \ldots, \Xi\big(q_\ell(s/\varepsilon)\big)\big)\sqrt{\varepsilon} Q^\varepsilon(s)\Big). \end{aligned}$$

If H^ε solves (5.1) then $U^\varepsilon(t) = Q^\varepsilon(t) - H^\varepsilon(t)$ satisfies by (5.4) the equation

$$U^\varepsilon(t) - \int_0^t \nabla_x B\big(\bar{Z}_x(s), \Xi\big(q_1(s/\varepsilon)\big), \ldots, \Xi\big(q_\ell(s/\varepsilon)\big)\big) U^\varepsilon(s)\,\mathrm{d}s = \int_0^t \big(J_1^\varepsilon(s) + J_2^\varepsilon(s)\big)\,\mathrm{d}s, \tag{5.6}$$

where

$$J_2^\varepsilon(s) = \Big(\nabla_x B\big(\bar{Z}_x(s), \Xi\big(q_1(s/\varepsilon)\big), \ldots, \Xi\big(q_\ell(s/\varepsilon)\big)\big) - \nabla_x \bar{B}\big(\bar{Z}_x(s)\big)\Big) H^\varepsilon(s).$$

By Gronwall's inequality we obtain that

$$\big|U^\varepsilon(t)\big| \le C t \mathrm{e}^{ct} \int_0^t \big|J_1^\varepsilon(s) + J_2^\varepsilon(s)\big|\,\mathrm{d}s \tag{5.7}$$

for some $C > 0$ independent of ε and $t \in [0, \mathcal{T}]$.

Thus, in order to prove that Q^ε converges weakly as $\varepsilon \to 0$ to a Gaussian process Q^0 solving (2.14) it suffices to show that $\int_0^t J_1^\varepsilon(s)\,\mathrm{d}s$ and $\int_0^t J_2^\varepsilon(s)\,\mathrm{d}s$ converge to zero in probability as $\varepsilon \to 0$. By (2.7),

$$\big|Z_x^\varepsilon(t) - Y_x^\varepsilon(t)\big| \le C \int_0^t \big|Z_x^\varepsilon(s) - \bar{Z}_x(s)\big|\,\mathrm{d}s = C\sqrt{\varepsilon} \int_0^t \big|Q_x^\varepsilon(s)\big|\,\mathrm{d}s$$

with $C = Kd$, and so

$$\big|Q_x^\varepsilon(t)\big| \le \big|G^\varepsilon(t)\big| + C \int_0^t \big|Q_x^\varepsilon(s)\big|\,\mathrm{d}s.$$

Hence, in the same way as in (5.2),

$$\left|Q_x^\varepsilon(t)\right| \le \left|G^\varepsilon(t)\right| + C\mathrm{e}^{Ct}\int_0^t \left|G^\varepsilon(s)\right| \mathrm{d}s. \tag{5.8}$$

By (2.7) and the Taylor formula with a reminder we conclude that

$$\left|J_1^\varepsilon(s)\right| \le C\sqrt{\varepsilon}\left|Q^\varepsilon(s)\right|^2 \tag{5.9}$$

which together with (4.2) yields that $E|J_1^\varepsilon(s)| \to 0$ as $\varepsilon \to 0$.

The proof of convergence to zero in probability of $\int_0^t J_2^\varepsilon(s)\,\mathrm{d}s$ as $\varepsilon \to 0$ is based on the integral representation (5.4). Set

$$\Phi(x, \xi_1, \dots, \xi_\ell) = B(x, \xi_1, \dots, \xi_\ell) - \bar{B}(x)$$

and

$$\Psi(x, \xi_1, \dots, \xi_\ell) = \nabla_x B(x, \xi_1, \dots, \xi_\ell) - \nabla_x \bar{B}(x).$$

Relying on the representation (5.4) we obtain that

$$\left|E\int_0^t J_2^\varepsilon(s)\,\mathrm{d}s\right| \le \left|J_3^\varepsilon(t)\right| + \left|J_4^\varepsilon(t)\right|, \tag{5.10}$$

where

$$\begin{aligned} J_3^\varepsilon(t) = \varepsilon^{3/2}\int_0^{t/\varepsilon} \mathrm{d}s \int_0^s \mathrm{d}u\, E\big(\Psi\big(\bar{Z}_x(\varepsilon s), \Xi\big(q_1(s)\big), \dots, \Xi\big(q_\ell(s)\big)\big) \\ \times \Phi\big(\bar{Z}_x(\varepsilon u), \Xi\big(q_1(u)\big), \dots, \Xi\big(q_\ell(u)\big)\big)\big) \end{aligned} \tag{5.11}$$

and

$$\begin{aligned} J_4^\varepsilon(t) = \varepsilon^{3/2}\int_0^{t/\varepsilon} \mathrm{d}s \int_0^{\varepsilon s} \mathrm{d}u \int_0^{u/\varepsilon} \mathrm{d}v\, K(\varepsilon s, \varepsilon v) \\ \times E\big(\Psi\big(\bar{Z}_x(\varepsilon s), \Xi\big(q_1(s)\big), \dots, \Xi\big(q_\ell(s)\big)\big)\Phi\big(\bar{Z}_x(\varepsilon u), \Xi\big(q_1(u)\big), \dots, \Xi\big(q_\ell(u)\big)\big)\big). \end{aligned} \tag{5.12}$$

Estimating the expectations in (5.11) and (5.12) via (3.10) similarly to (3.14) we obtain that both $J_3^\varepsilon(t)$ and $J_4^\varepsilon(t)$ are of order $\sqrt{\varepsilon}$, and so the left hand side of (5.10) is of order $\sqrt{\varepsilon}$, as well. For more details of a similar argument we refer the reader to [18]. This completes the proof of Theorem 2.2 concerning the continuous time case.

In the discrete time case the proofs are similar but slightly simpler. Namely, set

$$R_i^{1/N}(t) = N^{-1/2}\sum_{n=0}^{[Nt/i]} B_i\big(\bar{Z}(nt/N), \Xi\big(q_1(n)\big), \dots, \Xi\big(q_i(n)\big)\big), \tag{5.13}$$

where B_i's are the same as in (2.16)–(2.18). Then for all $N \ge 1$,

$$\left|G_i^{1/N}(t) - R_i^{1/N}(t)\right| \le CN^{-1/2} \tag{5.14}$$

for some $C > 0$ independent of N. The asymptotical behavior of $R^{1/N}$ as $N \to \infty$ can be studied in the same way as in [25] taking into account that we have here slightly different mixing conditions, and so the corresponding estimates should be done as above via (3.10)–(3.14). The main difference of the discrete vis-á-vis continuous time case is that now each $G_i^{1/N}(t), i = k+1, \dots, \ell$ converges weakly as $N \to \infty$ to a nondegenerate Gaussian process $G_i^0(t)$ having the covariances

$$E\big(G_i^0(t)G_i^0(s)\big) = \int_0^{\min(s,t)} \mathrm{d}u \int \big(B_i\big(\bar{Z}(u), \xi_1, \dots, \xi_i\big)\big)^2 \,\mathrm{d}\mu(\xi_1)\dots\mathrm{d}\mu(\xi_i) \tag{5.15}$$

which is proved combining arguments of Proposition 4.5 in [25] and of Section 4 above. The computation of other limiting covariances proceeds in the same way as in the continuous time case. It follows that in the discrete time case the processes G^ε converge weakly as $\varepsilon \to 0$ to a Gaussian process G^0 having the representation

$$G^0(t) = \sum_{i=1}^{k} G_i^0(it) + \sum_{i=k+1}^{\ell} G_i^0(t),$$

where each process $G_i^0, i > k$ is independent of each G_j^0 with $j \neq i$ while the processes $G_i^0, i \le k$ are correlated with covariances described at the end of Section 4 taken with $\alpha_i = i, i = 1, \dots, k$. The argument concerning the convergence of processes Q^ε to Q^0 solving (2.14) remains the same as in the continuous time case.

6. Some dynamical systems applications

We start with recalling the setup from [8] and [9]. A C^2-diffeomorphism F of a compact Riemannian manifold Ω is called partially hyperbolic if there is a F-invariant splitting $E^u \oplus E^c \oplus E^s$ of the tangent bundle of Ω with $E^u \neq 0$ and constants $\lambda_1 \le \lambda_2 < \lambda_3 \le \lambda_4 < \lambda_5 \le \lambda_6$, $\lambda_2 < 1, \lambda_5 > 1$ such that $\|\mathrm{d}F(v)\|/\|v\|$ is between λ_1 and λ_2 on E^s, between λ_3 and λ_4 on E^c and between λ_5 and λ_6 on E^u. Denote by W^u the foliation tangent to E^u and call S a u-set if S belongs to a single leaf of W^u. F-invariant probability measures which are absolutely continuous with respect to the volume on leafs W^u are called u-Gibbs measures. It is assumed that F has a unique u-Gibbs measure μ^{SRB} which is called the Sinai–Ruelle–Bowen (SRB) measure.

An important role in the construction is played by Markov families which are collections $\mathcal{S}$ of u-sets which cover Ω and have certain regularity properties (see [8] and [9]) but we formulate here only their "Markov property" saying that for any $S \in \mathcal{S}$ there are $S_i \in \mathcal{S}$ such that $FS = \bigcup_i S_i$. Now let $\mathcal{S}$ be a Markov family. Following [8] and [9] we construct on each $S \in \mathcal{S}$ an increasing sequence of σ-algebras $\mathcal{F}_n$ in the following recursive way. Let $\mathcal{F}_0^S = \{\varnothing, S\}$. Suppose that $\mathcal{F}_n^S$ is generated by $\{S_{j,n}\}$ with $F^n S_{j,n} \in \mathcal{S}$. By the "Markov property" we can decompose $F^{n+1}S_{j,n} = \bigcup_l S_{jl,n}$ and now let $\mathcal{F}_{n+1}^S$ be generated by $F^{-n-1}S_{jl,n}$.

Next, for each x_1 and x_2 in a u-set S put

$$\rho(x_1, x_2) = \prod_{j=0}^{\infty} \frac{\det(\mathrm{d}F^{-1}|E^u)(F^{-1}x_1)}{\det(\mathrm{d}F^{-1}|E^u)(F^{-1}x_2)}.$$

Fix $x_0 \in S$ and let $\rho_S(x) = \rho(x, x_0)(\int_S \rho(x, x_0)\,\mathrm{d}x)^{-1}$. For a Markov family $\mathcal{S}$ and nonnegative constants R, α denote by $E_1(\mathcal{S}, R, \alpha)$ the set of probability measures σ defined for each continuous function $g \in C(\Omega)$ by

$$\sigma(g) = \int_S g(x)\mathrm{e}^{G(x)}\rho_S(x)\,\mathrm{d}x, \tag{6.1}$$

where $S \in \mathcal{S}$ and G is Hölder continuous with the exponent α and the constant R. Denote also by $E = E(\mathcal{S}, R, \alpha)$ the closure of the convex hull of $E_1(\mathcal{S}, R, \alpha)$. The decay of correlations is measured in [8] and [9] via a sequence $a(n) \to 0$ as $n \to \infty$ such that for any $\sigma \in E$ and each Hölder continuous g on Ω,

$$\big|\sigma\big(g \circ F^n\big) - \mu^{\mathrm{SRB}}(g)\big| \le a(n)\|g\| \tag{6.2}$$

where $\|\cdot\|$ is a Hölder norm. An argument from Section 5 of [11] compares the coefficient $a(n)$ above with the more familiar rate of decay of correlations $|\mu^{\mathrm{SRB}}(f\cdot(g\circ F^n)) - \mu^{\mathrm{SRB}}(f)\mu^{\mathrm{SRB}}(g)|$ and it follows from there that the latter decays superpolynomially if and only if $a(n)$ decays superpolynomially. According to [7] such decay of correlations holds true for C^2 Anosov flows with jointly nonintegrable stable and unstable foliations and for their time-one maps. By [14] this remains true for an open dense set of C^2 Axion A flows as well, as for their time-one maps. For other partially hyperbolic dynamical systems with fast decay of correlations see [8], [9], [14] and references there.

In order to estimate $\eta_{p,\kappa,s}(n)$ from (2.3) we write in the same way as in Lemma 4 from [8] that on each element S in $\mathcal{F}_{[t]}$,

$$A_{n,s,t} = E\big(g\big(f \circ F^{n+t}, f \circ F^{n+t+s}\big)|\mathcal{F}_{[t]}\big) = \int_S \rho_S(y) g_{s,t}\big(F^n y\big)\,\mathrm{d}y, \tag{6.3}$$

where the expectation is with respect to σ on S and $g_{s,t}(z) = g(f(F^{t-[t]}z), f(F^{t-[t]+s}z))$. If f and g are Hölder continuous then $g_{s,t}$ is Hölder continuous for fixed s and t and it is uniformly in t Hölder continuous when $s = 0$. Thus, by (6.2) we have that $|A_{n,s,t} - EA_{n,s,t}|$ decays in n with the speed of at least $a(n)$ and this decay is uniform in t if $s = 0$. Hence, if $a(n)$ decays superpolynomially then (2.6) holds true. This yields Theorem 2.2 for $\Xi(t) = \Xi(t, z) = g(F^t z)$ on a probability space (S, σ) for $\sigma \in E$ and an element S of a Markov family while g is a Hölder continuous function. We observe that the measure σ here plays the role of the probability Pr in the setup of Section 2 while μ^{SRB} plays the role of P there.

7. Concluding remarks: Fully coupled averaging

In the nonconventional framework as discussed in this paper even the setup of fully coupled averaging, i.e. when the fast motion depends on the slow one, is not quite clear. On the first sight we may want to deal with the equations

$$\begin{aligned} &X^\varepsilon(n+1) = X^\varepsilon(n) + \varepsilon B\big(X^\varepsilon(n), \Xi(n), \Xi(2n), \dots, \Xi(\ell n)\big),\\ &\Xi(n+1) = F_{X^\varepsilon(n)}\big(\Xi(n)\big) \end{aligned} \tag{7.1}$$

in the discrete time case and

$$\frac{\mathrm{d}X^\varepsilon(t)}{\mathrm{d}t} = \varepsilon B\big(X^\varepsilon(t), \Xi(t), \Xi(2t), \dots, \Xi(\ell t)\big), \qquad \frac{\mathrm{d}\Xi(t)}{\mathrm{d}t} = b\big(X^\varepsilon(t), \Xi(t)\big) \tag{7.2}$$

in the continuous time case. The problem is that $\Xi(kn)$ or $\Xi(kt)$ are not yet defined for $k > 1$ at time n or t so we cannot insert them into the first equation in (7.1) or (7.2) respectively, and so these equations do not define properly X^ε and Ξ.

A reasonable modification of this setup is to consider

$$\begin{aligned} &X^\varepsilon(n+1) = X^\varepsilon(n) + \varepsilon B\big(X^\varepsilon(n), \eta_1(n), \eta_2(n), \dots, \eta_\ell(n)\big),\\ &\eta_i^\varepsilon(n+1) = F^i_{X^\varepsilon(n)}\big(\eta_i^\varepsilon(n)\big), \quad i = 1, \dots, \ell \end{aligned} \tag{7.3}$$

in the discrete time case and

$$\begin{aligned} &\frac{\mathrm{d}X^\varepsilon(t)}{\mathrm{d}t} = \varepsilon B\big(X^\varepsilon(t), \eta_1(t), \eta_2(t), \dots, \eta_\ell(t)\big),\\ &\frac{\mathrm{d}\eta_i^\varepsilon(t)}{\mathrm{d}t} = ib\big(X^\varepsilon(t), \eta_i^\varepsilon(t)\big), \quad i = 1, 2, \dots, \ell \end{aligned} \tag{7.4}$$

in the continuous time case. We consider (7.3) and (7.4) as sets of $\ell + 1$ equations but require that $\eta_1^\varepsilon(0) = \eta_2^\varepsilon(0) = \cdots = \eta_\ell^\varepsilon(0)$. This approach seems to be reasonable if we consider (7.3) and (7.4) as perturbations of equations with constants of motion

$$\eta^{(x)}(n+1) = F_x\big(\eta^{(x)}(n)\big) \quad \text{and} \quad \frac{\mathrm{d}\eta^{(x)}(t)}{\mathrm{d}t} = B\big(x, \eta^{(x)}(t)\big), \tag{7.5}$$

i.e. when x variable remains fixed in unperturbed equations but start moving slowly in perturbed ones. Then $\eta^{(x)}(i(n+1)) = F_x^i(\eta^{(x)}(in))$ and $\mathrm{d}\eta^{(x)}(it)/\mathrm{d}t = iB(x, \eta^{(x)}(it))$.

As it is well known in the fully coupled setup the averaging principle not always holds true and when it takes place then usually only in the sense of convergence in average or in measure. In the nonconventional situation the problem is even more complicated. Consider, for instance,

$$\begin{aligned}
&\frac{\mathrm{d}\alpha^{\varepsilon}_{\alpha,\varphi}(t)}{\mathrm{d}t}=\varepsilon B\big(\alpha^{\varepsilon}_{\alpha,\varphi}(t),\varphi^{\varepsilon}_{1,\alpha}(t),\dots,\varphi^{\varepsilon}_{\ell,\alpha}(t)\big),\\
&\frac{\mathrm{d}\varphi^{\varepsilon}_{i,\alpha,\varphi}(t)}{\mathrm{d}t}=i\alpha^{\varepsilon}_{i,\alpha,\varphi}(t),\quad \alpha^{\varepsilon}_{\alpha,\varphi}(0)=\alpha,\varphi^{\varepsilon}_{1,\alpha,\varphi}(0)=\cdots=\varphi^{\varepsilon}_{\ell,\alpha,\varphi}(0)=\varphi,
\end{aligned}\tag{7.6}$$

where φ denotes a point on an n-dimensional torus $\mathbb{T}^n$ and α denotes a constant n-vector (constant vector field on $\mathbb{T}^n$). Then $\varphi^{\varepsilon}_{i,\alpha,\varphi}=i\varphi^{\varepsilon}_{1,\alpha,\varphi}-(i-1)\varphi$. Set $\tilde{B}(\psi,\varphi)=B(\alpha,\psi,\psi-\varphi,\dots,\psi-(\ell-1)\varphi)$. Then the right hand side of (7.6) can be replaced by $\varepsilon\tilde{B}(\alpha^{\varepsilon}_{\alpha,\varphi}(t),\varphi^{\varepsilon}_{1,\alpha}(t),\varphi)$. If $\bar{B}(\alpha)=\int\tilde{B}(\alpha,\varphi_1,\varphi)\,\mathrm{d}\varphi_1\,\mathrm{d}\varphi$ and $\frac{\mathrm{d}\bar{\alpha}_\alpha(t)}{\mathrm{d}t}=\bar{B}(\bar{\alpha}_\alpha(t)),\bar{\alpha}_\alpha(0)=\alpha$ then employing the technique from the proof of Theorem 2.1 in [21] it is not difficult to see that for any compact K,

$$\int_K\sup_{0\le t\le\mathcal{T}/\varepsilon}|\alpha^{\varepsilon}_{\alpha,\varphi}(t)-\bar{\alpha}_\alpha(\varepsilon t)|\,\mathrm{d}\alpha\,\mathrm{d}\varphi\to 0\quad\text{as }\varepsilon\to 0.\tag{7.7}$$

Acknowledgment

A part of this paper was done during my visit to the Fields Institute in Toronto in June 2011 whose support and excellent working conditions I gratefully acknowledge.

References

[1] I. Assani. Multiple recurrence and almost sure convergence for weakly mixing dynamical systems. *Israel J. Math.* **103** (1998) 111–124. MR1613556
[2] V. Bergelson. Weakly mixing PET. *Ergodic Theory Dynam. Systems* **7** (1987) 337–349. MR0912373
[3] R. Bowen. *Equilibrium States and the Ergodic Theory of Anosov Diffeomorphisms. Lecture Notes in Math.* **470**. Springer, Berlin, 1975. MR0442989
[4] A. N. Borodin. A limit theorem for solutions of differential equations with random right-hand side. *Theory Probab. Appl.* **22** (1977) 482–497. MR0517995
[5] R. C. Bradley. *Introduction to Strong Mixing Conditions.* Kendrick Press, Heber City, 2007.
[6] V. Bergelson, A. Leibman and C. G. Moreira. From discrete-to continuous time ergodic theorems. *Ergodic Theory Dynam. Systems*. **32** (2012) 383–426.
[7] D. Dolgopyat. On decay of correlations in Anosov flows. *Ann. of Math. (2)* **147** (1998) 357–390. MR1626749
[8] D. Dolgopyat. Limit theorems for partially hyperbolic systems. *Trans. Amer. Math. Soc.* **356** (2003) 1637–1689. MR2034323
[9] D. Dolgopyat. Averaging and invariant measures. *Mosc. Math. J.* **5** (2005) 537–576. MR2241812
[10] J. Doob. *Stochastic Processes*. Wiley, New York, 1953. MR0058896
[11] D. Dolgopyat and C. Liverani. Energy transfer in a fast–slow Hamiltonian system. *Comm. Math. Phys.* **308** (2011) 201–225. MR2842975
[12] H. Furstenberg. Nonconventional ergodic averages. *Proc. Sympos Pure Math.* **50** (1990) 43–56. MR1067751
[13] M. Field, I. Melbourne and A. Torok. Decay of correlations, central limit theorems and approximation by Brownian motion for compact Lie group extensions. *Ergodic Theory Dynam. Systems* **23** (2003) 87–110. MR1971198
[14] M. Field, I. Melbourne and A. Torok. Stability of mixing and rapid mixing for hyperbolic flows. *Ann. of Math. (2)* **166** (2007) 269–291. MR2342697
[15] L. Heinrich. Mixing properties and central limit theorem for a class of non-identical piecewise monotonic C^2-transformations. *Math. Nachr.* **181** (1996) 185–214. MR1409076
[16] I. A. Ibragimov and Y. V. Linnik. *Independent and Stationary Sequences of Random Variables*. Wolters–Noordhoff, Groningen, 1971. MR0322926
[17] J. Jacod and A. N. Shiryaev. *Limit Theorems for Stochastic Processes*, 2nd edition. Springer, Berlin, 2003. MR1943877
[18] R. Z. Khasminskii. On stochastic processes defined by differential equations with a small parameter. *Theory Probab. Appl.* **11** (1966) 211–228. MR0203788
[19] R. Z. Khasminskii. A limit theorem for solutions of differential equations with random right-hand side. *Theory Probab. Appl.* **11** (1966) 390–406.
[20] Yu. Kifer. Limit theorems in averaging for dynamical systems. *Ergodic Theory Dynam. Systems* **15** (1995) 1143–1172. MR1366312
[21] Yu. Kifer. Averaging principle for fully coupled dynamical systems and large deviations. *Ergodic Theory Dynam. Systems* **24** (2004) 847–871. MR2062922

[22] Y. Kifer. Nonconventional law of large numbers and fractal dimensions of some multiple recurrence sets. *Stoch. Dyn.* **12** (2012) 1150023.
[23] Y. Kifer. A strong invariance principle for nonconventional sums. *Probab. Theory Related Fields* **155**(1–2) (2013) 463–486.
[24] A. Katok and B. Hasselblatt. *Introduction to the Modern Theory of Dynamical Systems*. Cambridge Univ. Press, Cambridge, 1995. MR1326374
[25] Yu. Kifer and S. R. S. Varadhan. Nonconventional limit theorems in discrete and continuous time via martingales. *Ann. Probab.* To appear.
[26] C. Liverani. Central limit theorems for deterministic systems. In *International Conference on Dynamical Systems (Montevideo, 1995)* 56–75. *Pitman Research Notes in Math.* **363**. Longman, Harlow, 1996. MR1460797
[27] D. L. McLeish. Invariance principles for dependent variables. *Z. Wahrsch. Verw. Gebiete* **32** (1975) 165–178. MR0388483
[28] D. L. McLeish. On the invariance principle for nonstationary mixingales. *Ann. Probab.* **5** (1977) 616–621. MR0445583
[29] J. A. Sanders, F. Verhurst and J. Murdock. *Averaging Methods in Nonlinear Dynamical Systems*, 2nd edition. Springer, New York, 2007. MR2316999