Convergence in distribution of stochastic dynamics

Mathias Rousset[(1)]

(1) INRIA Roquencourt, France

MICMAC project-team.

Motivation

- Consider a stochastic dynamical model in the form

$$t \mapsto (X_t^\varepsilon, E_t^\varepsilon),$$

where X denotes an *effective variable*, and E is an *environment variable*.

- General problem: we want to prove (rigorously) the convergence when $\varepsilon \to 0$ of the dynamics of the effective variable towards a dynamics in closed form.

Ex1: Overdamped Langevin dynamics

- Model: a classical Hamiltonian system $H : \mathbb{R}^{6N} \to \mathbb{R}$:

$$H(p,q) = \frac{1}{2}|p|^2 + V(q)$$

- $M = \mathrm{Id}$ rescaled mass coordinates.
- Introduction of a strong coupling with a stochastic thermostat of temperature, $\beta^{-1} = k_b T$.

Ex1: Overdamped Langevin dynamics

- The "simplest " case is given by the following equations of motion:

$$\begin{cases} dQ_t^\varepsilon = P_t^\varepsilon \, dt, \\ dP_t^\varepsilon = -\nabla V(Q_t^\varepsilon) \, dt - \underbrace{\frac{1}{\varepsilon} P_t^\varepsilon \, dt}_{Dissipation} + \underbrace{\sqrt{\frac{2}{\beta\varepsilon}} \, dW_t}_{Fluctuation} \end{cases}$$

- Physically: $\varepsilon =$ ratio between the timescale of vibrations in the Hamiltonian (slow), and the timescale of dissipation (fast).
- The invariant probability distribution is Gibbs $\propto \mathrm{e}^{-\beta H(q,p)} dq \, dp$ and independant of ε.

Ex1: Overdamped equations

- On large times of order $1/\varepsilon$, it is well known that the position variable is solution to the overdamped equation:

$$dQ_t = -\nabla V(Q_t)\, dt + \sqrt{2\beta^{-1}} dW_t.$$

- Thus in this case momenta p are the environment variables, and positions q are the effective variables .

- Model: a classical Hamiltonian system $H : \mathbb{R}^6 \to \mathbb{R}$ with one particle :

$$H(p,q) = \frac{1}{2}p^T p + V(q).$$

- V is a mixing and stationary random potential on $\mathbb{R}^3$.
- V is smooth and has vanishing average ($\mathbb{E}(\partial^k V(0)) = 0 \quad \forall k \geq 0$).
- The particle travels at high kinetic energy compare to V (weak coupling).

Ex2: Stochastic acceleration

- Efective dynamics occurs at diffusive scaling for momenta ("central limit theorem scaling").
- We look at a space scale of order $1/\varepsilon^2$, a particle kinetic energy of order 1, a potential energy of order ε.
- If V is made of "obstacles" the particle on time 1 hits $1/\varepsilon^2$ obstacles of null average and of size ε ("central limit scaling").
- Hamiltonian + Equation of motion:

$$\begin{cases} H_\varepsilon(p,q) = \frac{1}{2} p^T p + \varepsilon V(q/\varepsilon^2) \\ p_{t=0} = O(1). \end{cases}$$

$$\begin{cases} \frac{d}{dt} Q_t^\varepsilon = P_t^\varepsilon, \\ \frac{d}{dt} P_t^\varepsilon = -\frac{1}{\varepsilon} \nabla V(Q_t^\varepsilon/\varepsilon^2) \end{cases}$$

- When $\varepsilon \to 0$, the particle exhibits a Landau diffusion (diffusion of velocity on the unit sphere).
- Define

$$\begin{cases} R(q) = \mathbb{E}\left(V(0)V(q)\right) & \text{[Two point correl.]}, \\ A(p) = -\int_0^{+\infty} \mathrm{Hess} R(p\,t)\, dt & \underbrace{(\geq 0)}_{\text{sym. matrix sense}} \end{cases} .$$

- Equations of motion (SDE):

$$\begin{cases} dQ_t = P_t\, dt \\ dP_t = \mathrm{div} A(P_t)\, dt + A^{1/2}(P_t)\, dW_t \end{cases}$$

Ex2: Asymptotic stochastic acceleration

- In this case position and momenta (Q, P) are the effective variables, and the field $V(q)$ is the environment variable.

Many references

- Overdamped Langevin (stochastic averaging): Khas'minskii ('66), Papanicolaou Stroock Varhadan ('77), Kushner ('79), Stuart Pavliotis ('08).
- Stochastic acceleration: Kesten Papanicoalou ('85), Dürr Goldstein Lebowitz ('87), Ryzhik ('06), Kirkpatrick ('07).
- Problem: either extremely technical and ad hoc, or restricted to stochastic averaging with an environment variable in compact space.
- Goal: Give a more user-friendly general setting, robust to different models.

- The steps of the proof are standard
  - (i) Put a topology on path spaces (say uniform convergence).
  - (ii) Consider for each small parameter $\varepsilon > 0$ the probability distribution μ_ε on path space (say the space of continuous trajectories) of the effective variable.
  - (iii) Prove tightness = relative compacity for convergence in probability distribution of μ_ε when $\varepsilon \to 0$.
  - (iv) Extract a limit, denoted μ_0.
  - (v) Prove that under μ_0 and for a sufficiently large set of tests functions φ, then

$$t \mapsto \varphi(X_t^0) - \varphi(X_0^0) - \int_0^t \mathcal{L}_0 \varphi(X_s^0) \, ds$$

is a $\sigma(X^0)$-martingale, where $\mathcal{L}_0$ is a Markov generator.

- Define a metric or norm on path space for instance the uniform norm on $C_{\mathbb{R}^d}[0,T]$ (continuous paths) such that the topological space is Polish (separable = countable base of open sets, complete).
- The σ-field on $C_{\mathbb{R}^d}[0,T]$ is the Borel σ-field = all the sets obtained by a countable set operation of open sets. Topology $\Rightarrow$ measurable sets. You can now consider probability measures on it .
- Brownian motion is the only probability on $C_{\mathbb{R}^d}[0,T]$ such that a random variable realization $(W_t)_{t\geq 0}$ verifies for any $0 \leq s \leq t \leq T$

$$\begin{cases} \mathrm{Law}(W_t - W_s) = \mathcal{N}(\text{mean} = 0, \text{co-variance} = (t-s)\times \mathrm{Id}) \\ W_t - W_s \text{ independant of } W_{0\leq r\leq s}. \end{cases}$$

- The set of probability distribution on $C_{\mathbb{R}^d}[0,T]$ is topologized (again Polish:= separable, metric, complete) with weak convergence on continuous bounded test functions = convergence in distribution.
- Prohorov theorem: whatever the state space E (Polish:= separable, metric, complete), say here $E = C(\mathbb{R}^d, [0,T])$. Then tightness of $(X^\varepsilon)_{\varepsilon \geq 0}$ = "the main mass stays in a compact set" = for any $\varepsilon > 0$ there is a compact $K_{C(\mathbb{R}^d,[0,T]),\varepsilon} \subset E$ such that $\mathbb{P}\left(X^\varepsilon \in K_{C(\mathbb{R}^d,[0,T]),\varepsilon}\right) \geq 1 - \varepsilon$ is equivalent to relative compactness of convergence in distribution .
- Ascoli theorem characterize compact sets in path space $C(\mathbb{R}^d, [0,T])$ through uniform modulus of continuity.

- Filtration = $(\mathcal{F}_t)_{t\geq}$ = information of interest until time t = σ-field generated by the random processes of interest until time t.
- Adaptation of process X = the past of X until t is contained in $(\mathcal{F}_t)_{t\geq}$.
- Markov process X with respect to $(\mathcal{F}_t)_{t\geq 0}$ = " the future law depends on the past only through the present state " = for any $t_0 \geq 0$ the law of $X_{t_0 \leq t \leq T}$ conditionally on $\mathcal{F}_{t_0}$ and the present position of the process $\sigma(X_{t_0})$ are the same.
- Martingale $M \in \mathbb{R}^d$ with respect to $(\mathcal{F}_t)_{t\geq 0}$ = " whatever the past, the future average is zero " = for any $t_0 \geq 0$ the $\mathbb{E}(M_{t_0+h}|\mathcal{F}_{t_0}) = 0$.
- Stopping times with respect to $(\mathcal{F}_t)_{t\geq 0}$ = $\inf \{t \geq 0 | S_t = 0\}$ with $S_t \in \{0, 1\}$ adapted.

- Consider $t \mapsto X_t$ a Markov process, and $\mathcal{L}$ its generator that is to say (formally):

$$\mathcal{L}(\varphi)(x) := \frac{d}{dt}\Big|_{t=0} \mathbb{E}(\varphi(X_t)|X_0 = x).$$

- Ex: For Brownian motion, $\mathcal{L} = \frac{\Delta}{2}$, for ODE, $\mathcal{L} = F\nabla$ where F is a vector field, Kernel operators for processes with jumps, etc... General classification in $\mathbb{R}^d$ through the Levy-Kintchine formula .
- The Markov property implies the martingale property: if $\varphi \in D(\mathcal{L})$, then:

$$M_t := \varphi(X_t) - \varphi(X_0) - \int_0^t \mathcal{L}\varphi(X_s)\, ds$$

is a martingale (same reference filtration).

- Well-posed martingale problems gives the uniqueness: if $\forall\,\varphi \in D(\mathcal{L})$,

$$M_t := \varphi(X_t) - \varphi(X_0) - \int_0^t \mathcal{L}\varphi(X_s)\,ds$$

is a $\sigma(X_s, 0 \leq s \leq t)$-martingale, then the probability distribution of $t \mapsto X_t$ is unique and is Markov with respect to $\sigma(X_s, 0 \leq s \leq t)$ and of generator $\mathcal{L}$.

- Enables identification of limits obtained by compacity.
- Can be generalized to non-Markov.
- NB: Typically Lipschitz generators in $\mathbb{R}^d$ yields well-posed martingale problems through well-posed strong solutions of stochastic differential equations and a coupling argument ($\sim$ coupling + Cauchy-Lipschitz).

Standard references for limit theorems

- Ethier, Kurtz: Markov processes: characterization convergence, 87 (Markov generator oriented).
- Jacod, Shiryaev: Limit Theorems for Stochastic Processes, 87 (cad-lag semi-martingales oriented).
- Rq: Very technical to rather tedious.

Plugging in perturbation analysis

- We now want to ”plug in” some singular perturbation analysis in the martingale approach (Papanicolaou, Stroock, Varhadan ’77).
- Typically, the Markov generator of the full process $t \mapsto (X_t^\varepsilon, E_t^\varepsilon)$ is of the form:

$$\mathcal{L}_\varepsilon := \frac{1}{\varepsilon^2}\mathcal{L}_e + \frac{1}{\varepsilon}\mathcal{L}_x,$$

- $\mathcal{L}_e$ can be interpreted as the ” $\frac{1}{\varepsilon^2}$ fast” dynamics of the environment e. $\mathcal{L}_x$ is the ”$\frac{1}{\varepsilon}$ fast” dynamics of the effective variable x, null ”on average” of the effective variable.

- We assume the existence of an averaging operator $\langle \quad \rangle$ of the environment variables, such that:
  - (i) $\langle \quad \rangle$ is an invariant probability of $\mathcal{L}_e$ in the sense that we have (in a perhaps "very formal" way) the following representation: if $t \mapsto E_t$ is Markov with generator $\mathcal{L}_e$:

$$\mathcal{L}_e^{-1}\varphi(e,x) = -\int_0^{+\infty} \mathbb{E}\left(\varphi(E_t,x)|E_0 = e\right) dt.$$

    if $\langle \varphi \rangle = 0$ for any x.
  - (ii) Dynamics of the effective variable null on average: $\langle \mathcal{L}_x \varphi_0 \rangle = 0$, where $\varphi_0 \equiv \varphi_0(x)$ depends on x only.

- We now seek for a perturbed test function φ_ε of $\varphi_0(x)$ such that:

$$\begin{cases} \varphi_\varepsilon(x,e) = \varphi_0(x) + o_\varepsilon(1) \\ \mathcal{L}_\varepsilon \varphi_\varepsilon(x,e) = \mathcal{L}_0 \varphi_0 + o_\varepsilon(1). \end{cases}$$

- Formally, the perturbation analysis yields at order N:

$$\begin{cases} \varphi_\varepsilon(x,e) = \sum_{n=0}^{N} \varepsilon^n \varphi_n(x,e) \\ \varphi_{n+1} = \mathcal{L}_e^{-1} \left(\langle \mathcal{L}_x \varphi_n \rangle - \mathcal{L}_x \varphi_n \right). \end{cases}$$

Plugging perturbation analysis

- This yields the effective generator $\mathcal{L}_0$ through:

$$\begin{aligned}\mathcal{L}_\varepsilon \varphi_\varepsilon &= \langle \mathcal{L}_x \varphi_1 \rangle + o_\varepsilon(1) \\ &= \underbrace{-\left\langle \mathcal{L}_x \mathcal{L}_e^{-1} \mathcal{L}_x \right\rangle}_{\mathcal{L}_0} \varphi_0 + o_\varepsilon(1)\end{aligned}$$

- Typically, $\mathcal{L}_x$ is a first order differential operator, and $\mathcal{L}_0$ a second-order (diffusion) operator.

- Let $\Gamma_{\varepsilon,\delta}(\varphi_0)$ be the sup forward averaged variation of the path $t \mapsto \varphi_0(X_t^\varepsilon)$:

$$\Gamma_{\varepsilon,\delta}(\varphi_0) := \sup_{t,|h|\leq\delta} \left|\mathbb{E}\left(\varphi_0(X_{t+h}^\varepsilon) - \varphi_0(X_t^\varepsilon)|X_s^\varepsilon, 0 \leq s \leq t\right)\right|.$$

Then tightness follows from:

(i) Compact containment: For any $\varepsilon > 0$ there is a compact K_ε such that $\mathbb{P}\left(X_t^\varepsilon \in K_{\mathbb{R}^d,\varepsilon}, \forall t \in [0,T]\right) \geq 1 - \varepsilon$.

(ii) Uniform continuity of mean forward variation

$$\lim_{\delta\to 0}\sup_{\varepsilon} \mathbb{E}\Gamma_{\varepsilon,\delta}(\varphi_0) = 0, \quad \forall \varphi_0 \in D.$$

(iii) D is a dense algebra in $C_b(\mathbb{R}^d)$ for uniform convergence on compacts.

- This is good since perturbation analysis yields formally:

$$\begin{aligned}
&\varphi_0(X^\varepsilon_{t+h}) - \varphi_0(X^\varepsilon_t) \\
&\quad = \varphi_\varepsilon(X^\varepsilon_{t+h}, E^\varepsilon_{t+h}) - \varphi_\varepsilon(X^\varepsilon_t, E^\varepsilon_t) + o_\varepsilon(1) \\
&\quad = \int_t^{t+h} \mathcal{L}_\varepsilon \varphi_\varepsilon(X^\varepsilon_s, E^\varepsilon_s)\, ds + \text{martingale} + o_\varepsilon(1) \\
&\quad = \int_t^{t+h} \mathcal{L}_0 \varphi_0(X^\varepsilon_s)\, ds + \text{martingale} + o_\varepsilon(1).
\end{aligned}$$

And formally the forward averaged variation satisfies:

$$\mathbb{E}\left(\varphi_0(X^\varepsilon_{t+h}) - \varphi_0(X^\varepsilon_t) | X^\varepsilon_s, 0 \le s \le t\right) = o_\varepsilon(1) + O(h).$$

- To prove that the limit of $t \mapsto X_t^\varepsilon$ is solution of a martingale problem , we use the facts that:
  - By definition of $\mathcal{L}_\varepsilon$:

$$M_t^\varepsilon := \varphi_\varepsilon(X_t^\varepsilon, E_t^\varepsilon) - \varphi_\varepsilon(X_0^\varepsilon, E_0^\varepsilon) - \int_0^t \mathcal{L}_\varepsilon \varphi_\varepsilon(X_s^\varepsilon, E_s^\varepsilon)\, ds,$$

    is a $\sigma(X^\varepsilon)$-martingale
  - Being a martingale is an information on finite dimensional path functionals for any $k \geq 0$, $t_1 > \cdots > t_{-k}$ and $\varphi_1 \cdots \varphi_k$

$$\mathbb{E}\left((M_{t_1} - M_{t_0})\varphi_1(M_{t_{-1}}) \cdots \varphi_k(M_{t_{-k}})\right) = 0.$$

  - Pass to the limit $\varepsilon \to 0$ using Lebesgue dominated convergence and the perturbation analysis.

- Then several final sufficient criteriae for convergence to an effective dynamics may be obtained, e.g.:

  (i) $t \mapsto X_t^\varepsilon$ stays in some compact uniformly in ε with high probability.

  (ii) Define the rest terms:

$$\begin{cases} A_t^\varepsilon(\varphi_0) := \mathbb{E}\left((\varphi_\varepsilon - \varphi_0)(X_t^\varepsilon, E_t^\varepsilon)|X_t^\varepsilon, 0 \leq s \leq t\right), \\ B_t^\varepsilon(\varphi_0) := \mathbb{E}\left((\mathcal{L}_\varepsilon\varphi_\varepsilon - \mathcal{L}_0\varphi_0)\left(X_t^\varepsilon, E_t^\varepsilon\right)|X_s^\varepsilon, 0 \leq s \leq t\right). \end{cases}$$

  We ask $\forall \varphi_0 \in D$:

$$\begin{cases} \lim\limits_{\varepsilon\to 0} \mathbb{E} \sup\limits_t |A_t^\varepsilon(\varphi_0)| = 0, \\ \lim\limits_{\varepsilon\to 0} \mathbb{E} \int_0^T |B_t^\varepsilon(\varphi_0)| \, dt = 0. \end{cases}$$

- The martingale problem associated to $\mathcal{L}_0$ is well-posed.

- In practice, one introduces e.g. a stopping time cut-off τ_η with cut-off parameter $\eta > 0$ of the form:

$$\tau_\eta := \inf\left(t \geq 0 | s(X^\varepsilon)_t = 0\right),$$

where s is a continuous adapted functional $C(\mathbb{R}^d, [0,T]) \to C(\mathbb{R}^+, [0,T])$.

- We then ask that all the machinery holds for the processes stopped at τ_η.
- We then ask that under the law of the solution of the limit martingale problem associated to $\mathcal{L}_0$:

$$\lim_{\eta \to 0} \mathbb{P}\left(\tau_\eta = +\infty\right) = 1.$$

- Example: τ_η = time of exiting a compact of size $1/\eta$.

- Equations of motion:

$$\begin{cases} dQ_t = P_t\, dt, \\ dP_t = -\nabla V(Q_t)\, dt - \underbrace{\frac{1}{\varepsilon} P_t\, dt}_{Dissipation} + \underbrace{\sqrt{\frac{2}{\beta\varepsilon}}\, dW_t}_{Fluctuation} \end{cases}$$

- Generator:

$$\begin{cases} \mathcal{L}_\varepsilon = \frac{1}{\varepsilon^2}\mathcal{L}_p + \frac{1}{\varepsilon}\mathcal{L}_q, \\ \mathcal{L}_p = \frac{1}{\beta} \mathrm{e}^{\beta\frac{|p|^2}{2}} \mathrm{div}_p \left(\mathrm{e}^{-\beta\frac{|p|^2}{2}} \nabla_p\, . \right), \quad \text{[Orstein-Uhlenbeck process]} \\ \mathcal{L}_q = p\nabla_q - \nabla V(q)\nabla_p \quad \text{[Hamilton/Liouville operator]} \end{cases}$$

- Then the environment variable (e) is p here and the averaging operator is:

$$\langle \quad \rangle = \int \quad \mathrm{e}^{-\beta \frac{|p|^2}{2}} \frac{dp}{(2\pi)^{d/2}},$$

- And the effective dynamics is the drifted diffusion on q only :

$$\mathcal{L}_0 = -\left\langle \mathcal{L}_q \mathcal{L}_p^{-1} \mathcal{L}_q \right\rangle = -\nabla V(q) \nabla_q + \frac{1}{\beta} \Delta_q.$$

- Theorem: Let V be Lipschitz. The probability distribution of the path $t \mapsto Q_t^\varepsilon$ converges when $\varepsilon \to 0$ towards the Markov dynamics with generator $\mathcal{L}_0$.

- Equations fo motion (NB: V mixing, stationary, null avearge)

$$\begin{cases} \frac{d}{dt} Q_t^\varepsilon = P_t^\varepsilon, \\ \frac{d}{dt} P_t^\varepsilon = -\frac{1}{\varepsilon} \nabla V(Q_t^\varepsilon/\varepsilon^2) \end{cases}$$

- Effective variables: momenta p (the position q is unuseful), Environment variables: the "microscopic position" $(y = q/\varepsilon^2)$ and the random field V.
- The generator writes down:

$$\begin{cases} \mathcal{L}_\varepsilon = \frac{1}{\varepsilon^2} \mathcal{L}_y + \frac{1}{\varepsilon} \mathcal{L}_p, \\ \mathcal{L}_y = p \nabla_y \quad \text{[transport]} \\ \mathcal{L}_p = -\nabla V(y) \nabla_p. \end{cases}$$

- Then the averaging operator is the expectation with respect to the field randomness

$$\langle \quad \rangle = \mathbb{E}\,(\quad) = \int_{\text{fields}} \cdot \quad \mu(dv_{\text{pot}}),$$

- Technical trick in the perturbed test funtion, transport operators cannot be inverted and $\mathcal{L}_y^{-1}$ is replaced by:

$$\mathcal{L}_y^{\theta,-1}\psi(p,y,v_{\text{pot}}) = -\int_0^\theta \psi(y+pt,p,v_{\text{pot}})\,dt$$

where θ is a cut-off parameter and $\psi(p,y,v_{\text{pot}})$ is a test function with null average with respect to $\mu(dv_{\text{pot}})$ field integration , and depends on v_{pot} through future directed points y_+ only ($(y_+ - y).p \geq 0$).

- We can the exploit the mixing properties of the field V using estimates similar to :

$$\int_0^{+\infty} \left| \mathbb{E}\left(\psi(y+pt,p,V) \middle| V(y_-), y_- \text{past directed}\right)\right| \, dt < +\infty,$$

where "past directed" means $(y_- - y).p \leq 0$.

- In this sense, the transport operator along velocities is invertible.

- The effective generator is then:

$$\mathcal{L}_0 := -\left\langle \mathcal{L}_p \mathcal{L}_y^{-1} \mathcal{L}_p \right\rangle = \mathrm{div}_p \int_0^{+\infty} \mathbb{E}\left(\nabla V(0) \otimes \nabla V(pt) \right) \, dt \, \nabla_p$$

- Theorem: Let $d \geq 3$. Let V and its derivatives be (sufficiently) polynomially mixing, with p_0-moments for $p_0 \geq 0$ large enough. Then the probability distribution of the path $t \mapsto P_t^\varepsilon$ converges when $\varepsilon \to 0$ towards the (Landau) diffusion with generator $\mathcal{L}_0$.
- NB: technical cut-off in the proof necessary to prevent self-intersection of paths. (This explains $d \geq 3$).

- The results improves previous results by including fields with ”rare but peaky obstacles”, in which case the field $V \equiv V_\varepsilon$ depends on ε so that:

$$\begin{cases} \mathbb{E}\left(V_\varepsilon(0) \otimes V_\varepsilon(x)\right) = O(1). \\ \|V_\varepsilon\|_\infty = O(1/\varepsilon^\alpha), \quad \alpha \in [0,1]. \end{cases}$$

- Ref: MR, *Effective dynamics, perturbed test functions, and the stochastic acceleration problem.*, in preparation.