0 Warm-up

0.1 Processes with Fast Markov Modulations

0.1.1 Model Formulation

We consider here a two-scale system with the "slow" dynamics given by a one-dimensional conditionally Gaussian process X^ε with the drift modulated by a "fast" finite-state Markov process θ^ε. When θ^ε is in the state i the process X^ε behaves like the Wiener process with drift λ^i. If θ^ε is stationary, it is natural to expect that the process X^ε approximates in distribution the Wiener process with drift obtained by averaging of λ^i with weights proportional to the time spent by θ^ε in corresponding states.

Combining techniques based on the bounds for the total variation distance in terms of the Hellinger processes with methods of singular perturbations we prove a strong limit theorem for the slow variable even in the case of nonhomogeneous Markov modulations and establish a bound for the rate of convergence in the total variation norm. Notice that the model specification does not involve singular perturbed stochastic equations but they appear immediately when we look for an intrinsic description of the slow variable dynamics.

Let $(\Omega, \mathcal{G}, \mathbf{G} = (\mathcal{G}_t), P)$ be a stochastic basis with a one-dimensional Wiener process w and a nonhomogeneous Markov process $\theta^\varepsilon = (\theta^\varepsilon_t)_{t \le T}$ taking values in the finite set $\{1, 2, \dots, K\}$. The small parameter ε takes values in $]0, 1]$.

We shall consider the process X^ε given by

$$dX^\varepsilon_t = \lambda^\star J^\varepsilon_t dt + dw_t, \quad X^\varepsilon_0 = 0, \tag{0.1.1}$$

where $\lambda := (\lambda^1, \dots, \lambda^K)^\star$ is a fixed (column) vector and $J^\varepsilon = (J^{1,\varepsilon}, \dots, J^{K,\varepsilon})^\star$ is a vector with components $J^{i,\varepsilon} := I_{\{\theta^\varepsilon = i\}}$. In other words, (0.1.1) is just a convenient abbreviation for

$$X^\varepsilon_t = \int_0^t \sum_{i=1}^K \lambda^i I_{\{\theta^\varepsilon_s = i\}} \, ds + w_t. \tag{0.1.2}$$

Let $p^\varepsilon := (p^{1,\varepsilon}, \dots, p^{K,\varepsilon})^\star := E J^\varepsilon_0$ be the initial distribution of θ^ε. Notice that in the theory of Markov processes it is convenient to represent distributions as row vectors; to make notations of our model consistent with the further development we deviate here from this tradition.

We assume that the transition intensity matrices of θ^ε have the form $Q_t^\varepsilon := \varepsilon^{-1}Q_t$ where $Q = (Q_t)$ is a continuous matrix function with the following properties:

(1) for any $t \in [0,T]$ there is a unique probability distribution

$$\pi_t = (\pi_t^1, \dots, \pi_t^K)^\star$$

satisfying the equation

$$Q_t^\star \pi_t = 0, \tag{0.1.3}$$

i.e. zero is a simple eigenvalue and $\pi_t^\star$ is the corresponding left eigenvector of the matrix Q_t;

(2) $\pi = (\pi_t)$ is a continuous function;

(3) there exists $\kappa > 0$ such that for any $t \in [0,T]$

$$\operatorname{Re}\lambda(Q_t) < -2\kappa \tag{0.1.4}$$

where $\lambda(Q_t)$ runs the set of nonzero eigenvalues of Q_t.

The above hypotheses need some comments. We recall that a transition intensity matrix is a matrix with nonnegative elements except those in the diagonal and the sum of the elements in each row is equal to zero (hence, zero is always an eigenvalue). It is a well-known fact (see, e.g., [17]) that all other eigenvalues of such a matrix have strictly negative real parts and there are left eigenvectors which are probability distributions spanning the eigenspace corresponding to the zero eigenvalue. Thus, the assumption (1) is, actually, the requirement that zero is of multiplicity one while the properties (2) and (3) follow from (1) and continuity of Q_t. In a probabilistic language the property (1) means that for any fixed t the matrix Q_t can be viewed as the transition intensity matrix of an irreducible homogeneous Markov process and π_t is its invariant distribution. In particular, if Q does not depend on t, the process θ^ε is ergodic.

0.1.2 Asymptotic Behavior of Distributions

Let P_T^ε be the distribution of X^ε in the space $C[0,T]$ and R_T be the distribution of the process $X = (X_t)_{t\le T}$ given by

$$dX_t = \lambda^\star \pi_t dt + dw_t, \quad X_0 = 0, \tag{0.1.5}$$

i.e. of the Wiener process with drift $\lambda^\star \pi_t$.

Theorem 0.1.1 *(a)* $\lim_{\varepsilon\to 0} \operatorname{Var}(P_T^\varepsilon - R_T) = 0.$

(b) If $Q = (Q_t)$ is a continuously differentiable function then

$$\mathrm{Var}\,(P^\varepsilon_T - R_T) \le C(1+\delta_\lambda)\delta_\lambda \varepsilon^{1/2} \tag{0.1.6}$$

where $\delta_\lambda := \max \lambda^i - \min \lambda^i$ *and* C *is a constant depending only on* Q *and* T*.*

(c) If Q *does not depend on* t *and* $\pi = p^\varepsilon$ *there is a simpler bound*

$$\mathrm{Var}\,(P^\varepsilon_T - R_T) \le C\delta^2_\lambda \varepsilon^{1/2}. \tag{0.1.7}$$

Proof. Let $\mathbf{F}^\varepsilon$ be the filtration generated by X^ε and null sets and let $\widehat{J}^\varepsilon$ be the $\mathbf{F}^\varepsilon$-optional projection of J^ε, i.e. $\mathbf{F}^\varepsilon$-optional process such that

$$\widehat{J}^\varepsilon_\tau = E(J^\varepsilon_\tau | \mathcal{F}^\varepsilon_\tau)$$

for any $\mathbf{F}^\varepsilon$-stopping time τ.

Put

$$\widetilde{w}_t := X^\varepsilon_t - \int_0^t \lambda^\star \widehat{J}^\varepsilon_s ds = w_t + \int_0^t \lambda^\star J^\varepsilon_s ds - \int_0^t \lambda^\star \widehat{J}^\varepsilon_s ds. \tag{0.1.8}$$

Then $\widetilde{w}$ is an $\mathbf{F}^\varepsilon$-adapted Wiener process (this simple observation is known as the innovation theorem) and X^ε can be represented as a diffusion-type process with

$$dX^\varepsilon_t = \lambda^\star \widehat{J}^\varepsilon_t dt + d\widetilde{w}_t, \quad X^\varepsilon_0 = 0, \tag{0.1.9}$$

(see, e.g., [66], Th. 7.12). According to [66], Th. 9.1, $\widehat{J}^\varepsilon$ satisfies the filtering equation

$$d\widehat{J}^\varepsilon_t = \varepsilon^{-1} Q^\star_t \widehat{J}^\varepsilon_t dt + \phi(\widehat{J}^\varepsilon_t) d\widetilde{w}_t, \quad \widehat{J}_0 = p^\varepsilon, \tag{0.1.10}$$

where

$$\phi(\widehat{J}^\varepsilon_t) := \mathrm{diag}\,\lambda\, \widehat{J}^\varepsilon_t - \widehat{J}^\varepsilon_t(\widehat{J}^\varepsilon_t \lambda) \tag{0.1.11}$$

and $\mathrm{diag}\,\lambda$ is the diagonal matrix with $\lambda_{ii} := \lambda_i$.

Let $|.|_1$ be the absolute norm of a matrix (or a vector), that is, the sum of the absolute values of its components. It is easily seen that

$$|\phi(\widehat{J}^\varepsilon_t)|_1 = \sum_{i=1}^K J^{i,\varepsilon}_t |\lambda^i - J^\varepsilon_t \lambda| \le \sum_{i=1}^K J^{i,\varepsilon}_t \delta_\lambda = \delta_\lambda$$

and hence

$$|\phi(\widehat{J}^\varepsilon_t)|^2 \le |\phi(\widehat{J}^\varepsilon_t)|_1^2 \le \delta^2_\lambda. \tag{0.1.12}$$

Applying for the pair of measures P^ε and R the upper bound in (A.3.3) we get that

$$\mathrm{Var}\ (P^\varepsilon_T - R_T) \le 4\sqrt{E h^\varepsilon_T} \tag{0.1.13}$$

where the Hellinger process h^ε is given by

$$h^\varepsilon_t := \frac{1}{8}\int_0^t (\lambda^\star(\widehat{J}^\varepsilon_s - \pi_s))^2 ds. \tag{0.1.14}$$

For any $a=(a^1,\dots,a^K)$ with $\sum a^i=0$ we have

$$a\lambda=\sum a^i I_{\{a^i\ge 0\}}\lambda^i+\sum a^i I_{\{a^i<0\}}\lambda^i\le |a|/2(\max\lambda^i-\min\lambda^i)=(1/2)|a|\delta_\lambda.$$

Thus,

$$|\lambda^\star(\widehat{J}^\varepsilon_s-\pi_s)|\le(1/2)\delta_\lambda|\widehat{J}^\varepsilon_s-\pi_s|_1\le(1/2)\delta_\lambda|\widehat{J}^\varepsilon_s-\pi_s|$$

and we get by virtue of (0.1.14) that

$$Eh^\varepsilon_t\le\frac{1}{32}\delta^2_\lambda E\int_0^t|\widehat{J}^\varepsilon_s-\pi_s|^2\,ds. \tag{0.1.15}$$

Put $z^\varepsilon:=\widehat{J}^\varepsilon-\pi$. It follows from (0.1.3) and (0.1.10) that

$$z^\varepsilon_t=z^\varepsilon_0+\varepsilon^{-1}\int_0^t Q^\star_s z^\varepsilon_s\,ds+\int_0^t\phi(\widehat{J}^\varepsilon_s)\,d\widetilde{w}_s-(\pi_t-\pi_0). \tag{0.1.16}$$

Let us consider the subspace $\mathcal{L}:=\{x\in\mathbf{R}^K:\ x^\star\mathbf{1}=0\}$ where

$$\mathbf{1}:=(1,\dots,1)^\star.$$

Clearly, $\mathcal{L}$ is an invariant subspace for every operator Q_t and the restriction A_t of Q_t to $\mathcal{L}$ has the same eigenvalues as Q_t except zero. Thus, we can view (0.1.16) as the operator equation in $\mathbf{R}^{K-1}$. If the function Q_t is continuous differentiable, (0.1.16) can be written as

$$dz^\varepsilon_t=\varepsilon^{-1}A_t z^\varepsilon_t dt+\phi(\widehat{J}^\varepsilon_t)d\widetilde{w}_t+\dot{\pi}_t dt,\quad z^\varepsilon_0=p^\varepsilon-\pi_0, \tag{0.1.17}$$

and this we consider as a matrix equation in $\mathbf{R}^{K-1}$ (by choosing an orthonormal basis in $\mathcal{L}$). By the Cauchy formula we have

$$z^\varepsilon_t=\Phi^\varepsilon(t,0)z^\varepsilon_0+\int_0^t\Phi^\varepsilon(t,s)\phi(\widehat{J}^\varepsilon_s)\,d\widetilde{w}_s+\int_0^t\Phi^\varepsilon(t,s)\dot{\pi}_s\,ds \tag{0.1.18}$$

where $\Phi^\varepsilon(t,s)$ is the fundamental (or transition) matrix corresponding to $\varepsilon^{-1}A$, i.e. the solution of the equation

$$\frac{\partial\Phi^\varepsilon(t,s)}{\partial t}=\varepsilon^{-1}A_t\Phi^\varepsilon(t,s)dt,\quad \Phi^\varepsilon(s,s)=I.$$

Using the exponential inequality

$$|\Phi^\varepsilon(t,s)|\le ce^{-\kappa(t-s)/\varepsilon} \tag{0.1.19}$$

(see Proposition A.2.3) and taking into account (0.1.12) we easily obtain the following bounds:

$$|\Phi^\varepsilon(t,0)z_0^\varepsilon|^2 \le c^2|p^\varepsilon-\pi_0|e^{-2\kappa t/\varepsilon}, \tag{0.1.20}$$

$$E\left|\int_0^t \Phi^\varepsilon(t,s)\phi(\widehat{J}_s^\varepsilon)\,d\widetilde{w}_s\right|^2 = E\int_0^t |\Phi^\varepsilon(t,s)|^2|\phi(\widehat{J}_s^\varepsilon)|^2\,ds \le \delta_\lambda^2\frac{c^2\varepsilon}{2\kappa}, \tag{0.1.21}$$

$$\left|\int_0^t \dot\pi_s\Phi^\varepsilon(t,s)\,ds\right|^2 \le \|\dot\pi\|_T^2\left(\int_0^t |\Phi^\varepsilon(t,s)|\,ds\right)^2 \le \|\dot\pi\|_T^2\frac{c^2\varepsilon^2}{\kappa^2} \tag{0.1.22}$$

where $\|\dot\pi\|_T := \sup_{t\le T}|\dot\pi_t|$.

From (0.1.18) and (0.1.20)–(0.1.22) we get that for some constant C

$$\int_0^T E|z_s^\varepsilon|^2\,ds \le C^2(1+\delta_\lambda^2)\varepsilon \tag{0.1.23}$$

and, in the homogeneous case with $p^\varepsilon=\pi$,

$$\int_0^T E|z_s^\varepsilon|^2\,ds \le C^2\delta_\lambda^2\varepsilon. \tag{0.1.24}$$

Now the assertion (b) is evident in view of (0.1.13) and (0.1.15).

In the case of the assertion (a) where the function Q (and hence π) is supposed to be only continuous the equation (0.1.16) cannot be written as (0.1.17) and the usual Cauchy formula is not applicable. Nevertheless, we can represent z as follows:

$$z_t^\varepsilon = \Phi^\varepsilon(t,0)z_0^\varepsilon + \int_0^t \Phi^\varepsilon(t,s)\phi(\widehat{J}_s^\varepsilon)\,d\widetilde{w}_s + r_t^\varepsilon \tag{0.1.25}$$

where

$$r_t^\varepsilon := \Phi^\varepsilon(t,0)(\pi_t-\pi_0) + \int_0^t \frac{\partial\Phi^\varepsilon(t,s)}{\partial s}(\pi_t-\pi_s)\,ds. \tag{0.1.26}$$

Arguing as above we infer from (0.1.24) the bound

$$\int_0^T E|z_s^\varepsilon|^2\,ds \le C(\varepsilon+\|r^\varepsilon\|_T)$$

and it remains to show that $\|r^\varepsilon\|_T\to 0$ as $\varepsilon\to 0$. But taking into account the relation

$$\frac{\partial\Phi^\varepsilon(t,s)}{\partial s} = -\varepsilon^{-1}A_s\Phi^\varepsilon(t,s), \quad \Phi^\varepsilon(t,t)=I,$$

(which follows from the semigroup property) and using the exponential inequality we obtain that

$$|r_t^\varepsilon| \le ce^{-\kappa t/\varepsilon}|\pi_t-\pi_0| + c\varepsilon^{-1}\|A\|_T\int_0^t e^{-\kappa(t-s)/\varepsilon}|\pi_t-\pi_s|\,ds.$$

The uniform convergence of r^ε to zero follows now from Lemma A.2.4. □

Conclusion. The bounds of Theorem 0.1.1 hold by virtue of properties of the equation (0.1.16) which is a **singularly perturbed stochastic equation** because it can be written as

$$\varepsilon dz_t^\varepsilon = A_t z_t^\varepsilon dt + \sigma(\varepsilon) G(z^\varepsilon) d\tilde{w}_t + \sigma(\varepsilon) b_t dt \tag{0.1.27}$$

with $\sigma(\varepsilon) = o(1)$ as $\varepsilon \to 0$. It is important to note that the matrix function A_t admits the exponential bound (0.1.19). Another essential feature is that the parameter $\sigma(\varepsilon) := \varepsilon$ at the diffusion coefficient tends to zero faster than $\sqrt{\varepsilon}$ providing the convergence of trajectories to zero. Singularly perturbed stochastic equations of this type and, especially, more general systems involving also “slow” variables are the objects of principal interest of this book. Some techniques developed in the sequel will be used in Section 6.2 where we present a more deep analysis of our model and show that the bounds (0.1.6) and (0.1.7) give the correct order of convergence in ε by calculating the limit of $\varepsilon^{-1/2}\mathrm{Var}\,(P_T^\varepsilon - R_T)$.

0.2 The Liénard Oscillator Under Random Force

In this section we discuss briefly an important example of a two-scale stochastic system, namely, the Liénard oscillator driven by random force. This classic model arises in a mathematical description of the motion of a small particle in a viscous media. On an intuitive level, it can be described by the second-order equation

$$\varepsilon \ddot{x} + \dot{x} - h(x) = \dot{w} \tag{0.2.1}$$

where $\dot{w}$ is a white noise and ε is a small positive parameter. The standard reduction transforms it to the system of two equations of the first order

$$\dot{x} = v, \tag{0.2.2}$$

$$\varepsilon \dot{v} = -v + h(x) + \dot{w}. \tag{0.2.3}$$

The rigorous formulation can be given by the following system of stochastic equations in the usual Ito sense where we exhibit explicitly the dependence on ε:

$$dx_t^\varepsilon = v_t^\varepsilon dt, \quad x_0^\varepsilon = x^o, \tag{0.2.4}$$

$$\varepsilon dv_t^\varepsilon = -v_t^\varepsilon dt + h(t, x_t^\varepsilon) dt + dw_t, \quad v_0^\varepsilon = v^o. \tag{0.2.5}$$

Here w is a Wiener process and the initial condition can be random.

It is worth noting that this system is quite specific: the equation for the position x^ε does not contain a diffusion term, while the equation for the velocity v^ε does not involve a small parameter at diffusion. For $h = 0$ the process v^ε is simply the Ornstein–Uhlenbeck process. Therefore, in general, we cannot expect the convergence of v^ε. Nevertheless, it is easy to prove that under mild assumptions on h the position process x^ε converges uniformly on any compact interval in probability to the process x (in the literature referred to as the Smoluchowski–Kramers approximation) satisfying the stochastic equation

$$dx_t = h(t, x_t)dt + dw_t, \quad x_0 = x^o. \tag{0.2.6}$$

We give here a bit more precise result of this kind assuming that the processes x^ε, v^ε, and w are n-dimensional.

Proposition 0.2.1 *Let $T \in \mathbf{R}_+$, $p \in [1, \infty[$. Assume that $x^o, v^o \in L^p(\Omega)$ and h satisfies on $[0,T] \times \mathbf{R}^n$ the global Lipschitz condition and the linear growth condition. Then there exists a constant C (depending on p and T) such that*

$$\lim_{\varepsilon \to 0} (E||x^\varepsilon - x||_T^p)^{1/p} \le C\sqrt{\varepsilon |\ln \varepsilon|} \tag{0.2.7}$$

where $||.||_T$ is the norm in $C[0,T]$.

Proof. Put $\Delta_t^\varepsilon := x_t^\varepsilon - x_t$. Using the Cauchy formula we "resolve" (0.2.5) and obtain the representation

$$v_s = e^{-s/\varepsilon} v^o + \frac{1}{\varepsilon}\int_0^s e^{-(s-u)/\varepsilon} h(u, x_u^\varepsilon)du + \frac{1}{\varepsilon}\int_0^s e^{-(s-u)/\varepsilon} dw_u. \tag{0.2.8}$$

Substituting it into (0.2.4), we get from (0.2.4) and (0.2.6) that

$$\Delta_t^\varepsilon = \varepsilon(1 - e^{-t/\varepsilon})v^o + \zeta_t^\varepsilon + \xi_t^\varepsilon + \eta_t^\varepsilon \tag{0.2.9}$$

where

$$\begin{aligned}
\zeta_t^\varepsilon &:= \int_0^t \left(\frac{1}{\varepsilon}\int_0^s e^{-(s-u)/\varepsilon}(h(u, x_u^\varepsilon) - h(u, x_u))du \right) ds, \\
\xi_t^\varepsilon &:= \int_0^t \left(\frac{1}{\varepsilon}\int_0^s e^{-(s-u)/\varepsilon} h(u, x_u)du \right) ds - \int_0^t h(s, x_s)ds, \\
\eta_t^\varepsilon &:= \frac{1}{\varepsilon}\int_0^t \left(\int_0^s e^{-(s-u)/\varepsilon} dw_u \right) ds - w_t.
\end{aligned}$$

By virtue of the Fubini theorems

$$\zeta_t^\varepsilon = \int_0^t (1 - e^{-(t-u)/\varepsilon})(h(u, x_u^\varepsilon) - h(u, x_u))du, \tag{0.2.10}$$

$$\xi_t^\varepsilon = \int_0^t e^{-(t-u)/\varepsilon} h(u, x_u)du, \tag{0.2.11}$$

$$\eta_t^\varepsilon = -\int_0^t e^{-(t-u)/\varepsilon} dw_u. \tag{0.2.12}$$

By assumption, there is a constant L such that

$$|h(t,y_1)-h(t,y_2)|\le L|y_1-y_2| \quad \forall t\in[0,T],\ y_1,y_2\in\mathbf{R}^n,$$

and

$$|h(t,y)|\le L(1+|y|) \quad \forall t\in[0,T],\ y\in\mathbf{R}^n.$$

Using this we deduce from (0.2.9)–(0.2.11) that for every $t\le T$

$$||\Delta^\varepsilon||_t\le\varepsilon|v^o|+L\int_0^t||\Delta^\varepsilon||_u du+L\varepsilon(1+||x||_T)+||\eta^\varepsilon||_T$$

and, hence, by the Gronwall–Bellman lemma

$$||\Delta^\varepsilon||_T\le\big(\varepsilon|v^o|+\varepsilon L(1+||x||_T)+||\eta^\varepsilon||_T\big)e^{LT}. \tag{0.2.13}$$

This bound implies the result. Indeed, we assume that the initial conditions are random variables belonging to $L^p(\Omega)$. Due to the linear growth and global Lipschitz conditions $||x||_T\in L^p(\Omega)$ for all finite p. It remains to notice that for some constant C_p we have

$$(E||\eta^\varepsilon||_T^p)^{1/p}\sim C_p\sqrt{\varepsilon|\ln\varepsilon|},\quad \varepsilon\to 0, \tag{0.2.14}$$

see Chapter 1. □

Remark. Notice that the equation (0.2.5) for the fast variable does not contain the small parameter at diffusion and, thus, the model looks different from the basic one considered in this book. However, if we choose for analysis of (0.2.1) the so-called Liénard coordinates by putting $u=\varepsilon\dot{x}+x$, the resulting system will be

$$du_t=h(x_t)dt+dw_t, \tag{0.2.15}$$
$$\varepsilon dx_t=(-x_t+u)dt, \tag{0.2.16}$$

with the diffusion coefficient of the fast variable equal to zero. The general theory of Chapter 4 includes this case if h is Lipschitz.

Comment. It would be more consistent with the modern methodology to start with the models (0.2.4), (0.2.5) each defined on its own probability space and indexed (together with its Wiener process) by the parameter ε. The physically meaningful question is the convergence of the distribution of x^ε in the space $C[0,T]$ to that of x. To do this, there is a powerful method to construct a realization of processes on a single probability space and prove the convergence of x^ε to x (as random variables with values in $C[0,T]$) in probability. We start, actually, from the point where this transfer has been done. Does $||x^\varepsilon-x||_T$ converge to zero almost surely? The positive answer, in view of the bound (0.2.13), seems obvious. But take care: (0.2.13) holds only a.s. and the exceptional set may depend on ε. In fact, for the process η^ε we cannot ensure even the convergence $||\eta^\varepsilon||_T$ to zero a.s. because stochastic integrals are defined up to P-null sets. To make this question mathematically correct (and admitting the positive answer) we should construct a good realization of the whole family to ensure the continuity of paths $||\Delta^\varepsilon||_T$ and $||\eta^\varepsilon||_T$ for $\varepsilon > 0$. For this simple model it is not difficult. For instance, integrating by parts we get that for every $\varepsilon > 0$

$$\eta_t^\varepsilon = w_t - \varepsilon^{-1} \int_0^t e^{-(t-s)/\varepsilon} w_s ds$$

almost surely for all t. The right-hand side of this formula can be used to define the appropriate version of η^e.

0.3 Filtering of Nearly Observed Processes

The problem of nonlinear filtering consists in estimating a stochastic process (a signal) that is not directly observed. A lot of studies are devoted to the practically important case where the process is *nearly* observed. This is an asymptotic setting in which computable asymptotic filters can be easily studied. The aim of this section is to provide a simple illustrative example where the singular perturbed stochastic equations appear in a natural way.

Let us consider the model described by two processes x (unobservable signal) and y^ε (observations), both, for simplicity, n-dimensional, given by

$$dx_t = f_t dt + \sigma_t dw^x, \quad x_0 = x^0, \tag{0.3.1}$$

$$dy_t^\varepsilon = x_t dt + \varepsilon dw^y, \quad y_0^\varepsilon = y^0, \tag{0.3.2}$$

where w^x and w^y are independent Wiener processes in $\mathbf{R}^n$, and f and σ are continuous processes of corresponding dimensions adapted to the filtration generated by w^x. The parameter $\varepsilon \in]0,1]$ is small; it formalizes the fact that noises in the signal and observations are of different scales and the signal-to-noise ratio is large.

A filter is any process adapted with respect to the filtration of y^ε. Engineers are looking often for filters which approximate x in some sense. Such filters may not perform so well as but are easier to implement.

Let us consider the filter $\widehat{x}^\varepsilon$ admitting the following representation:

$$d\widehat{x}_t^\varepsilon = \widehat{f}_t^\varepsilon dt - \varepsilon^{-1} A_t (dy_t^\varepsilon - \widehat{x}_t^\varepsilon dt), \tag{0.3.3}$$

where the continuous vector-valued process $\widehat{f}^\varepsilon$ (assumed to be a function of y^ε), the continuous function A with values in the set of $n \times n$ matrices, and the initial condition can be viewed as filter parameters. We assume that there is a constant $\kappa > 0$ such that for $\operatorname{Re} \lambda(A_t) < -2\kappa$ for all t. For the error process $\Delta^\varepsilon := \widehat{x}^\varepsilon - x$ we get from (0.3.1)–(0.3.3) the equation

$$d\Delta_t^\varepsilon = \varepsilon^{-1} A_t \Delta_t^\varepsilon dt + (\widehat{f}_t^\varepsilon - f_t) dt + G_t d\tilde{w}_t, \quad \Delta_0^\varepsilon = \widehat{x}_0^\varepsilon - x^0, \tag{0.3.4}$$

where $G_t := (A_t A_t^\star + \sigma_t \sigma_t^\star)^{1/2}$ and

$$\tilde{w}_t := -\int_0^t G_s^{-1} A_s dw_s^y - \int_0^t G_s^{-1} \sigma_s dw_s^x$$

is a Wiener process in $\mathbf{R}^n$.

Let $\Phi^\varepsilon(t,s)$ be the fundamental matrix defined by the linear equation

$$\varepsilon \frac{\partial \Phi^\varepsilon(t,s)}{\partial t} = A_t \Phi^\varepsilon(t,s), \quad \Phi^\varepsilon(s,s) = I. \tag{0.3.5}$$

Using the Cauchy formula we can write the solution of (0.3.4) as

$$\Delta_t^\varepsilon = \Phi^\varepsilon(t,0)(\widehat{x}_0^\varepsilon - x^0) + \int_0^t \Phi^\varepsilon(t,s)(\widehat{f}_s^\varepsilon - f_s)ds + \xi_t^\varepsilon \tag{0.3.6}$$

where

$$\xi_t^\varepsilon := \int_0^t \Phi^\varepsilon(t,s) G_s d\tilde{w}_s. \tag{0.3.7}$$

The process ξ_t^ε is the solution of

$$\varepsilon d\xi_t^\varepsilon := A_t \xi_t^\varepsilon dt + \varepsilon G_t d\tilde{w}_t, \quad \xi_0^\varepsilon = 0. \tag{0.3.8}$$

For us it is important to note that the asymptotic behavior of the approximate filter is determined by properties of solutions of a singularly perturbed stochastic equation (with a small parameter at the diffusion term of order ε).

Using the exponential bound for $|\Phi^\varepsilon(t,s)|$ (see Lemma A.2.2) we obtain that

$$|\Delta_t^\varepsilon| \le Ce^{-\kappa t/\varepsilon}|\widehat{x}_0^\varepsilon - x^0| + C\int_0^t e^{-\kappa(t-s)/\varepsilon}|\widehat{f}_s^\varepsilon - f_s|ds + |\xi_t^\varepsilon|. \tag{0.3.9}$$

This implies the following less precise but simpler inequality which gives a clear idea of the filter behavior:

$$|\Delta_t^\varepsilon| \le Ce^{-\kappa t/\varepsilon}|\widehat{x}_0^\varepsilon - x^0| + C\varepsilon\kappa^{-1}||\widehat{f}^\varepsilon - f||_t + |\xi_t^\varepsilon| \tag{0.3.10}$$

where $||.||_t$ denotes the uniform norm on $[0,t]$.

In the particular case of constant A and σ we have $\Phi^\varepsilon(t,s) = e^{(t-s)A/\varepsilon}$, the process ξ^ε is Gaussian, and $\varepsilon^{-1/2}\xi_t^\varepsilon$ converges in distribution as $\varepsilon \to 0$ to the centered Gaussian vector with covariance matrix

$$S_A = \int_0^\infty e^{rA}(AA^\star + \sigma\sigma^\star)e^{rA^\star}dr. \tag{0.3.11}$$

Assuming, e.g., that $|\widehat{x}_0^\varepsilon|$ is bounded and $\varepsilon^{1/2}||\widehat{f}^\varepsilon - f||_t$ converges to zero in probability as $\varepsilon \to 0$ we infer from (0.3.6) that $\varepsilon^{-1/2}(\widehat{x}_t^\varepsilon - x_t)$ is asymptotically Gaussian with zero mean and covariance S_A.

In a more specific situation of scalar processes ($n = 1$) with the filter parameter $A = -\gamma > 0$ we have $S_A := (\gamma^2 + \sigma^2)/(2\gamma)$. If $\sigma > 0$ is known one can attain the smallest value of asymptotic variance $S_A = \sigma$ by choosing $\gamma = \sigma$.

In the vector model with a known nondegenerated matrix σ it is reasonable to choose $A = -(\sigma\sigma^\star)^{1/2}$ and get the limit covariance $S_A = (\sigma\sigma^\star)^{1/2}$. To justify such a choice we notice that for any symmetric negative definite matrix B commuting with $\sigma\sigma^\star$ the difference $S_B - (\sigma\sigma^\star)^{1/2}$ is positive definite because, in this case,

$$S_B = -\frac{1}{2}B^{-1}(B^2 + (\sigma\sigma^\star)) = (\sigma\sigma^\star)^{1/2} - \frac{1}{2}B^{-1}(B + (\sigma\sigma^\star)^{1/2})^2 \qquad (0.3.12)$$

and the last term is negative definite.

One can expect that the filter will exhibit a better performance also in a non-asymptotic sense (i.e. for realistic values of the signal-to-noise ratio) if $\widehat{f}_t^\varepsilon$ tracks f_t. We return to this model in Sections 6.3 and 6.4.

0.4 Stochastic Approximation

The stochastic approximation theory, initially developed for discrete-time models but now treated more and more often in a very general semimartingale setting, deals with the problems of estimating a root of an unknown function F on the basis of observations of a controlled random process $\theta = \theta^\gamma$. We consider here a rather particular continuous-time white-noise model which, nevertheless, covers several approximation procedures studied in the literature. Our aim is to show that, being rescaled, it comes into the framework of the theory of singularly perturbed stochastic differential equations which allows us to analyze stochastic approximation procedures in a systematic and transparent way and get asymptotic expansions of estimators.

Let $\theta = \theta^\gamma$ be given on $[t_0, \infty[$ by the SDE

$$d\theta_t = \gamma_t F(\theta_t)dt + \gamma_t dw_t, \quad \theta_{t_0} = \theta_0, \qquad (0.4.1)$$

where w is a Wiener process in $\mathbf{R}^n$, the function $F : \mathbf{R}^n \to \mathbf{R}^n$ is continuously differentiable, the "control" $\gamma = (\gamma_t)_{t>0}$ is a nonnegative continuous deterministic function, and the initial condition is posed at some point $t_0 > 0$.

We assume that F satisfies the following hypotheses:

H.0.4.1 There is a unique root θ_* of the equation

$$F(\theta) = 0$$

and

$$(\theta - \theta_*)^\star F(\theta) < 0 \quad \forall\, \theta \in \mathbf{R}^n \setminus \{\theta_*\}. \qquad (0.4.2)$$

H.0.4.2 The real parts of all eigenvalues of the matrix

$$A := F'(\theta_*)$$

are strictly negative: $\operatorname{Re}\lambda(A) < -2\kappa < 0$.

Resembling the standard problems of optimal control, the model has specific features: it is not completely specified since F is unknown (but some extra information on F may be available) and the class of controls is quite restrictive. For these reasons, the traditional paradigm of stochastic approximation does not formulate the optimal control problem by stipulating in a precise way an objective function but uses instead the ideology and concepts of mathematical statistics. There is a vast literature devoted to analysis, for particular stochastic procedures γ, of the asymptotic behavior as $T \to \infty$ of θ_T or, more recently, of the average

$$\hat{\theta}_T := \frac{1}{T - t_1} \int_{t_1}^{T} \theta_s ds, \tag{0.4.3}$$

as statistical estimators of θ_*. For instance, the continuous-time version of the classic Robbins–Monro procedure claims that θ_T is a strongly consistent estimator of θ_*. Its precise formulation is as follows.

Proposition 0.4.1 *Assume that* **H.0.4.1** *holds and*

$$\int_{t_0}^{\infty} \gamma_u \, du = \infty, \quad \int_{t_0}^{\infty} \gamma_u^2 \, du < \infty. \tag{0.4.4}$$

Then

$$\lim_{t\to\infty} \theta_t = \theta_* \quad \text{a.s. and in } L^4. \tag{0.4.5}$$

Proof. Put $U_t := \theta_t - \theta_*$. Then

$$dU_t = F(\theta_t)\gamma_t dt + \gamma_t dw_t, \quad \theta_{t_0} = \theta_0,$$

and, by the Ito formula,

$$|U_t|^2 = |U_{t_0}|^2 + 2\int_{t_0}^{t} U_s^\star F(\theta_s)\gamma_s ds + 2M_t + \int_{t_0}^{t} \gamma_s^2 ds \tag{0.4.6}$$

where

$$M_t := \int_{t_0}^{t} U_s^\star \gamma_s dw_s.$$

Notice that $U_s^\star F(\theta_s) \le 0$ by (0.4.2) and hence the first integral in the right-hand side of (0.4.6) defines a decreasing process. Localizing the stochastic integral M and taking the expectation, we get, with help of the Fatou lemma, that

$$E|U_t|^2 \leq E|U_{t_0}|^2 + \int_{t_0}^{\infty} \gamma_s^2 ds.$$

It follows that

$$E\langle M\rangle_\infty = \int_{t_0}^{\infty} E|U_s|^2 \gamma_s^2 ds \leq \left(E|U_{t_0}|^2 + \int_{t_0}^{\infty} \gamma_s^2 ds\right) \int_{t_0}^{\infty} \gamma_s^2 ds < \infty$$

by the second relation in (0.4.4). The square integrable martingale M bounded in L^2 converges a.s. to a finite limit. Thus, the processes on the right-hand side of (0.4.6) converge at infinity to finite limits (a.s.). The continuity of F and the relation **H.0.4.2** imply that for every $r \in]0,1[$ there is a constant $c_r > 0$ such that

$$(\theta - \theta^*)F(\theta) \leq -c_r$$

when $r \leq |\theta - \theta^*| \leq 1/r$. The divergence of the integral of γ implies that on the set $\{\lim |U_t| > 0\}$ the first integral in (0.4.6) diverges to $-\infty$. Hence, U_t converges to zero a.s. At last, $||M||_{t_0,\infty} \in L^2$ and the process $|U|^2$, being bounded by a square integrable random variable, converges to zero in L^2. □

We consider here two stochastic approximation procedures and study asymptotic expansions of the estimator (0.4.3). The first procedure, depending on a parameter $\rho \in]1/2, 1[$, corresponds to the choice

$$\gamma_t := t^{-\rho} \tag{0.4.7}$$

and $t_1 = t_1(T) = Tr_1(T)$ with

$$r_1(T) := \frac{1}{\ln(\gamma_T T)} = \frac{1}{(1-\rho)\ln T}. \tag{0.4.8}$$

The second one, with the characteristics marked by the superscript ^{o}, is given by

$$\gamma_t^o := e\frac{\ln t}{t \ln_3 t} \tag{0.4.9}$$

with

$$r_1^o(T) := \frac{1}{\ln_2 T} \tag{0.4.10}$$

where $\ln_n$ denotes the n-times-iterated logarithm.

Theorem 0.4.2 *Suppose that $F \in C^3$ and* **H.0.4.1**, **H.0.4.2** *are fulfilled. Then for the procedure given by (0.4.7), (0.4.8) we have*

$$\widehat{\theta}_T = \theta_* + \xi_T \frac{1}{T^{1/2}} + h\frac{1}{1-\rho}\frac{1}{T^\rho} + R_T \frac{1}{T^\rho} \tag{0.4.11}$$

where $h \in \mathbf{R}^n$, ξ_T is a centered Gaussian random vector with covariance matrix converging to $(A^\star A)^{-1}$, and $R_T \to 0$ in probability as $T \to \infty$.

Theorem 0.4.3 *Suppose that $F \in C^3$ and* **H.0.4.1**, **H.0.4.2** *are fulfilled. Then for the procedure given by (0.4.9), (0.4.10) we have*

$$\widehat{\theta}_T^o = \theta_* + \xi_T^o \frac{1}{T^{1/2}} + he\frac{\ln T}{T} + R_T^o \frac{\ln T}{T} \tag{0.4.12}$$

where $h \in \mathbf{R}^n$, ξ_T^o is a centered Gaussian random vector with covariance matrix converging to $(A^\star A)^{-1}$, and $R_T^o \to 0$ in probability as $T \to \infty$.

Remark 1. The vector h in the above theorems depends only on F. In the scalar case we have $h = (1/4)A^{-2}F''(\theta_*)$. The explicit expression in the general case can be found in Section 2.4.

Remark 2. An inspection of (0.4.11) makes plausible the idea that the third term on its right-hand side is responsible for the bias of the estimator. Obviously, for sufficiently large T

$$\max_{\rho \in]1/2,1[} (1-\rho)T^\rho = \frac{1}{e}\frac{T}{\ln T}.$$

Thus, the minimum over $\rho \in]1/2, 1[$ of the third term on the right-hand side of (0.4.11) coincides with the corresponding terms in (0.4.12). This observation explains our interest in the second procedure. Indeed, under a certain auxiliary condition $E|R_T|$ and $E|R_T^o|$ converge to zero, see Theorem 0.4.6 below.

We prove the above results in Chapter 2 providing here only the reduction to the framework of singular perturbations.

First, let us consider the procedure with the function γ_t defined by (0.4.7). A simple rescaling leads to a problem on the interval with a fixed right extremity. Indeed, put $\tilde{\theta}_r := \theta_{rT}$. Then, by virtue of (0.4.1), on the interval $[t_0/T, 1]$

$$d\tilde{\theta}_r = \gamma_T T F(\tilde{\theta}_r)\gamma_r dr + \gamma_T T^{1/2}\gamma_r d\tilde{w}_r, \quad \tilde{\theta}_{t_0/T} = \theta_0, \tag{0.4.13}$$

where $\tilde{w}_r := T^{-1/2} w_{rT}$ is a Wiener process. Obviously,

$$\widehat{\theta}_T = \frac{1}{1-r_1}\int_{r_1}^1 \tilde{\theta}_r dr.$$

Now we reparameterize the problem by introducing instead of the large parameter T the small parameter

$$\varepsilon := \frac{1}{\gamma_T T} = \frac{1}{T^{1-\rho}}. \tag{0.4.14}$$

Then

$$T = T(\varepsilon) = \frac{1}{\varepsilon^{1/(1-\rho)}}. \tag{0.4.15}$$

Setting $y_r^\varepsilon := \tilde{\theta}_{rT(\varepsilon)}$, we rewrite (0.4.13) as the singularly perturbed stochastic equation

$$\varepsilon dy_r^\varepsilon = F(y_r^\varepsilon)\gamma_r dr + \beta\varepsilon^{1/2}\gamma_r d\tilde{w}_r, \quad y_{t_0/T(\varepsilon)} = \theta_0, \tag{0.4.16}$$

where

$$\beta := \sqrt{\gamma_{T(\varepsilon)}} = \varepsilon^{(1/2)\rho/(1-\rho)}. \tag{0.4.17}$$

With this new parameterization $\widehat{\theta}_T$ becomes equal to

$$\widehat{y}_1^\varepsilon := \frac{1}{1-r_1^\varepsilon}\int_{r_1^\varepsilon}^1 y_r^\varepsilon dr \tag{0.4.18}$$

where

$$r_1^\varepsilon = -\frac{1}{\ln\varepsilon}. \tag{0.4.19}$$

Theorem 0.4.3 has the following equivalent form:

Theorem 0.4.4 *Suppose that $F \in C^3$ and* **H.0.4.1**, **H.0.4.2** *are fulfilled. Then for the model (0.4.14)–(0.4.19)*

$$\widehat{y}_1^\varepsilon = \theta_* + \xi^\varepsilon\varepsilon^{1/2}\beta + h\frac{1}{1-\rho}\beta^2 + R^\varepsilon\beta^2 \tag{0.4.20}$$

where $h \in \mathbf{R}^n$, ξ^ε is a centered Gaussian random vector with covariance matrix converging to $(A^\star A)^{-1}$, and $R^\varepsilon \to 0$ in probability as $\varepsilon \to 0$.

One can notice that the small parameters are involved in (0.4.16) in a very simple, multiplicative, way. The only particular feature is that the starting time depends on ε and the function γ has a singularity at zero which is integrable. The coefficient $1/(1-\rho)$ in (0.4.20) is equal to the integral of γ over $[r_1^\varepsilon, 1]$ up to $o(1)$.

Similarly, the rescaling of the model with γ^o defined in (0.4.9) results in the stochastic equation on $[t_0/T, 1]$

$$d\tilde{\theta}_r = \gamma_{rT}^o TF(\tilde{\theta}_r)dr + \gamma_{rT}^o T^{1/2}d\tilde{w}_r, \quad \tilde{\theta}_{t_0/T} = \theta_0. \tag{0.4.21}$$

For sufficiently large T we define the function $\varepsilon = \varepsilon(T)$ by putting

$$\varepsilon := \frac{1}{\gamma_T^o T} = \frac{1}{e}\frac{\ln_3 T}{\ln T}. \tag{0.4.22}$$

Let $T(\varepsilon)$ be the inverse of the above function. We rewrite (0.4.21) as

$$\varepsilon dy_r^\varepsilon = F(y_r^\varepsilon)\gamma_r^\varepsilon dr + \beta\varepsilon^{1/2}\gamma_r^\varepsilon d\tilde{w}_r, \quad y_{t_0/T(\varepsilon)} = \theta_0, \tag{0.4.23}$$

where

$$\gamma_r^\varepsilon := \frac{1}{r}\frac{\ln(rT(\varepsilon))}{\ln T(\varepsilon)}\frac{\ln_3 T(\varepsilon)}{\ln_3(rT(\varepsilon))} \tag{0.4.24}$$

and

$$\beta := \sqrt{\gamma^o_{T(\varepsilon)}} = \frac{1}{\sqrt{\varepsilon T(\varepsilon)}}. \tag{0.4.25}$$

Again $\widehat{\theta}_T$ is equal to

$$\widehat{y}_1^\varepsilon := \frac{1}{1-r_1^\varepsilon}\int_{r_1^\varepsilon}^1 y_r^\varepsilon dr \tag{0.4.26}$$

but now

$$r_1^\varepsilon = \frac{1}{\ln_2 T(\varepsilon)}. \tag{0.4.27}$$

The corresponding equivalent version of Theorem 0.4.3 is

Theorem 0.4.5 *Suppose that $F \in C^3$ and* **H.0.4.1**, **H.0.4.2** *are fulfilled. Then for the model (0.4.22)–(0.4.27)*

$$\widehat{y}_1^\varepsilon = \theta_* + \xi^{o,\varepsilon}\varepsilon^{1/2}\beta + h\beta^2 \ln_3 T(\varepsilon) + R^{o,\varepsilon}\beta^2 \ln_3 T(\varepsilon) \tag{0.4.28}$$

where $h \in \mathbf{R}^n$, $\xi^{o,\varepsilon}$ is a centered Gaussian random vector with covariance matrix converging to $(A^\star A)^{-1}$, and $R^{o,\varepsilon} \to 0$ in probability as $\varepsilon \to 0$.

Of course, more systematic notations require the superscript ^{o} at $T(\varepsilon)$, β, etc, but we skip it for obvious reasons.

Remark. Clearly, the equations (0.4.16) and (0.4.23) are of the same structure. However, in the latter case the function γ_r^ε has a singularity at zero like $1/r$ which is not integrable and which yields in the term $\ln_3 T(\varepsilon)$ after integrating over the interval $[r_1^\varepsilon, 1]$.

To get a convergence of residual terms we add to our assumption the following hypothesis on a “global” behavior of F:

H.0.4.3 There exists a bounded matrix-valued function $A(y_1, y_2)$ such that for all y_1, y_2

$$F(y_1) - F(y_2) = A(y_1, y_2)(y_1 - y_2) \tag{0.4.29}$$

and

$$z^\star A(y_1, y_2)z \le -\kappa|z|^2 \qquad \forall z \in \mathbf{R}^n \tag{0.4.30}$$

for some constant $\kappa > 0$.

Clearly, **H.0.4.3** implies the Lipschitz and linear growth condition. In the one-dimensional case this hypothesis holds if $F \in C^1$ and $F' \le -\kappa < 0$.

Theorem 0.4.6 *Suppose that $F \in C^3$, the second derivative F'' is bounded and satisfies the Lipschitz condition, and the conditions* **H.0.4.1**–**H.0.4.3** *are fulfilled. Then $E|R^\varepsilon| = o(1)$ and $E|R^{o,\varepsilon}| = o(1)$ as $\varepsilon \to 0$.*

As a corollary we obtain, under the assumptions of Theorem 0.4.6, that

$$E\widehat{\theta}_T = \theta_* + h\frac{1}{1-\rho}\frac{1}{T^\rho} + \frac{1}{T^\rho}o(1), \tag{0.4.31}$$

$$E\widehat{\theta}_T^o = \theta_* + he\frac{\ln T}{T} + \frac{\ln T}{T}o(1) \tag{0.4.32}$$

as $T \to \infty$.

http://www.springer.com/978-3-540-65332-5

Two-Scale Stochastic Systems
Asymptotic Analysis and Control
Kabanov, Y.; Pergamenshchikov, S.
2003, XIV, 266 p., Hardcover
ISBN: 978-3-540-65332-5