Chapter 15

Introduction to Stochastic Approximation Algorithms

[1]Stochastic approximation algorithms are recursive update rules that can be used, among other things, to solve optimization problems and fixed point equations (including standard linear systems) when the collected data is subject to noise. In engineering, optimization problems are often of this type, when you do not have a mathematical model of the system (which can be too complex) but still would like to optimize its behavior by adjusting certain parameters. For this purpose, you can do experiments or run simulations to evaluate the performance of the system at given values of the parameters. Stochastic approximation algorithms have also been used in the social sciences to describe collective dynamics: fictitious play in learning theory and consensus algorithms can be studied using their theory. In short, it is hard to overemphasized their usefulness. In addition, the theory of stochastic approximation algorithms, at least when approached using the ODE method as done here, is a beautiful mix of dynamical systems theory and probability theory. We only have time to give you a flavor of this theory but hopefully this will motivate you to explore further on your own. For our purpose, essentially all approximate DP algorithms encountered in the following chapters are stochastic approximation algorithms. We will not have time to give formal convergence proofs for all of them, but this chapter should give you a starting point to understand the basic mechanisms involved. Most of the material discussed here is taken from [Bor08].

15.1 Example: The Robbins-Monro Algorithm

Suppose we wish to find the root $\bar{\theta}$ of the function $f : \mathbb{R} \to \mathbb{R}$. We can use Newton's procedure, which generates the sequence of iterates

$$\theta_{n+1} = \theta_n - \frac{f(\theta_n)}{f'(\theta_n)}.$$

[1]This version: October 31 2009

Suppose we also know a neighborhood of $\bar{\theta}$, where $f(\theta) < 0$ for $\theta < \bar{\theta}$, $f(\theta) > 0$ for $\theta > \bar{\theta}$, and f in nondecreasing in this neighborhood. Then if we start at θ_0 close enough of $\bar{\theta}$, the following simpler (but less efficient) scheme also converges to $\bar{\theta}$, and does not require the derivative of f:

$$\theta_{n+1} = \theta_n - \alpha f(\theta_n), \tag{15.1}$$

for some fixed and sufficiently small $\alpha > 0$. Note that if f is itself the derivative of a function F, these schemes correspond to Newton's method and a fixed-step gradient descent procedure for minimizing F, respectively (more precisely, finding a critical point of F or root of the gradient of F).

Very often in applications, we do not have access to the mathematical model f, but we can do experiments or simulations to sample the function at particular values of θ. These samples are typically noisy however, so that we can assume that we have a black-box at our disposal (the simulator, the lab where we do the experiments, etc.), which on input $x\theta$ returns the value $y = f(\theta) + d$, where d is a noise, which will soon be assumed to be random. The point is that we only have access to the value y, and we have no way of removing the noise from it, i.e., of isolating the exact value of $f(\theta)$. Now suppose that we still want to find a root of f as in the problem above, with access only to this noisy black box.

Assume for now that we know that the noise is i.i.d. and zero-mean. A first approach to the problem could be, for a given value of θ, to sample sufficient many time at the same point θ and get values $y_1, \ldots, y_N$, and then form an estimate of $f(\theta)$ using the empirical average

$$f(\theta) \approx \frac{1}{N} \sum_{i=1}^{N} y_i. \tag{15.2}$$

With sufficiently many samples at every iterate θ_n of (15.1), we can reasonably hope to find approximately the root of f. The problem is that we might spend a lot of time taking samples at points θ that are far from $\bar{\theta}$ and are not really relevant, except for telling us in which direction to move next. This can be a real issue if obtaining each sample is time-consuming or costly.

An alternative procedure, studied by Robbins and Monro [RM51][2], is to simply use directly the noisy version of f in a slightly modified version of algorithm (15.1):

$$\theta_{n+1} = \theta_n - \gamma_n y_n, \tag{15.3}$$

where γ_n is a sequence of positive numbers converging to 0 and such that $\sum_n \gamma_n = \infty$ (for example, $\gamma_n = 1/(n+1)$), and $y_n = f(\theta_n) + d_n$ is the noisy version of $f(\theta_n)$. Note that the iterates θ_n are now random variables.

The intuition behing the decreasing step size γ_n is that it provides a sort of averaging of the observations. For an analogy in a simpler setting, suppose

[2]In fact, recursive stochastic algorithms have been used in signal processing (e.g., for smoothing radar returns) even before the work of Robbins and Monro. However, there was apparently no general asymptotic theory.

we have i.i.d. observations $\xi_1, \ldots, \xi_N$ of a random variable and wish to form their empirical average as in (15.2). A recursive alternative to (15.2), extremely useful in settings where the samples become available progressively with time (recall for example the Kalman filter), is to form

$$\theta_1 = \xi_1, \;\; \theta_{n+1} = \theta_n - \gamma_n[\theta_n - \xi_{n+1}],$$

with $\gamma_n = 1/(n+1)$. One can immediately verify that $\theta_n = (\sum_{i=1}^n \xi_i)/n$, for all n.

This chapter is concerned with recurrences generalizing (15.3) of the form:

$$\theta_{n+1} = \theta_n + \gamma_n[f(\theta_n) + b_n + D_{n+1}] \tag{15.4}$$

where $\theta_0 \in \mathbb{R}^d$ is possibly random, f is a function $\mathbb{R}^d \to \mathbb{R}^d$, b_n is a small systematic perturbation term, such as a bias in our estimator of $f(\theta_n)$, and D_{n+1} is a random noise with zero mean (conditioned on the past). The assumptions and exact definitions of these terms will be made precise in section 15.3. In applications, we are typically first interested in the asymptotic behavior of the sequence $\{\theta_n\}$.

15.2 The ODE Approach and More Application Examples

The ODE (Ordinary Differential Equation) method says roughly that if the step sizes γ_n are appropriately chosen, the bias terms b_n decrease appropriately, and the noise D_n is zero-mean, then the iterates (15.4) asymptotically track the trajectories of the dynamical system[3]

$$\dot{\theta} = f(\theta).$$

We will give a more formal proof of this fact in the basic case in section 15.3. Typically for the simplest proofs γ_n must be decreasing to 0 and satisfy

$$\sum_n \gamma_n = \infty, \;\; \sum_n \gamma_n^2 < \infty.$$

However other choices are possible, including constant small step sizes in some cases, and in practice the choice of step sizes requires experimentation because it controls the convergence rate. Some theoretical results regarding convergence rates are also available but will not be covered here. The ODE is extremely useful in any case, even if another technique is chosen for formal convergence proofs, in order to get a quick idea of the behavior of an algorithm. Moreover, another big advantage of this method is that it can be used to easily create new stochastic approximation algorithms from convergent ODEs. We now describe a few more classes of problems where these algorithms arise.

[3]By definition, $\dot{x} := \frac{d}{dt}x(t)$

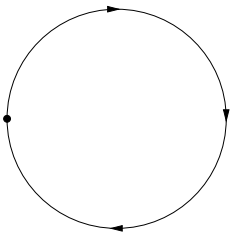


Figure 15.1: Consider a flow on a circle that moves clockwise everywhere except at a single rest point. This rest point is the unique ω-limit point of the flow. Now suppose the flow represents the expected motion of some underlying stochastic process. If the stochastic process reaches the rest point, its *expected* motion is zero. Nevertheless, actual motion may occur with positive probability and in particular the process can jump past the rest point and begin another circuit. Therefore in the long run all regions of the circle are visited infinitely often. The long run behavior is captured by the notion of chain recurrence, as all points on the circle are chain recurrent under the flow.

Brief Review of Some Concepts from Dynamical Systems

Consider an (autonomous) ordinary differential equation (ODE)

$$\dot{x}(t) = f(x(t)),\ x(0) = x_0,\ x(t) \in \mathbb{R}^d, t \in \mathbb{R}. \tag{15.5}$$

We assume that the ODE is well-posed, i.e., for each initial condition $x_0 \in \mathbb{R}^d$ it has a unique solution $x(\cdot)$ defined for all $t \geq 0$ and the map associating an initial condition x_0 to its corresponding solution $x(\cdot) \in C([0,\infty), \mathbb{R}^d)$ is continuous (for the topology of uniform convergence on compacts). One sufficient condition for this is to assume that f is *Lipschitz*, i.e., there exists $L > 0$ such that

$$\|f(x) - f(y)\| \leq L\|x - y\|,\ \forall x, y \in \mathbb{R}^d.$$

A closed set $A \subset \mathbb{R}^d$ is an *invariant set* for this ODE if any trajectory $x(t), -\infty < t < +\infty$ with $x(0) \in A$ satisfies $x(t) \in A$ for all $t \in \mathbb{R}$. In the basic convergence theorem in section 15.3, the concept of *chain transitivity* appears. A close set $A \subset \mathbb{R}^d$ set is said to be *internally chain transitive* for the ODE if for any $x, y \in A$ and any $\epsilon > 0, T > 0$, there exists points $x_0 = x, x_1, \ldots, x_{n-1}, x_n = y$ in A, for some $n \geq 1$, such that the trajectory of (15.5) starting at x_i, for $0 \leq i < n$ meets with the ϵ-neighborhood of x_{i+1} after a time greater or equal to T (take $x = y$ in this definition to obtain the notion of *chain recurrence*). The small jumps at the points of the chain is a natural assumption for stochastic approximations, where the noise pushes the iterates away from the trajectories of the ODE, see Fig. 15.1.

Given a trajectory $x(\cdot)$ of (15.5), the set $\Omega = \cap_{t>0}\overline{\{x(s) : s > t\}}$, i.e., the set of its limit points as $t \to \infty$, is called its ω-limit set. Note that Ω depends on the actual trajectory. It is easy to verify that Ω is an invariant set for the ODE.

Def of Lyapunov function for a CT system.
Lasalle's invariance principle.

Stochastic Gradient Algorithms

The simplest set-up where stochastic approximation algorithms arise is in the context of noisy versions of optimization algorithms. Consider the Robbins-Monro scheme, but not the function for which we wish to find a root is itself the gradient of another function f. That is, we consider a gradient descent iteration of the type

$$x_{n+1} = x_n + \gamma_n[-\nabla f(x_n) + D_{n+1}],$$

where f is a continuously differentiable function we want to minimize. We do not have access to the gradient of f directly however, only to a noisy version of it. The limiting ODE is then

$$\dot{x}(t) = -\nabla f(x(t)), \tag{15.6}$$

i.e., describes a gradient flow, and such dynamical system are among the simplest ones to study. Indeed, f itself serves as a Lyapunov function to study convergence:

$$\frac{d}{dt} f(x(t)) = -\|\nabla f(x(t))\|^2 \leq 0,$$

where the inequality is strict when $\nabla f(x(t)) \neq 0$. The set of equilibria of (15.6) is $H = \{x : \nabla f(x) = 0\}$. By Lasalle's invariance principle, the only limit sets that can occur as ω-limit sets for (15.6) are subsets of H, and the ODE method tells us that the iterates converge almost surely (a.s.) to such an invariant set. Moreover, they avoid convergence to critical points $\nabla f(x) = 0$ that are either maxima or saddle-points, as these represent unstable equilibria of the ODE. In particular if f has only isolated local minima, we can expect that the iterates $\{x_n\}$ converge to one of them. In another variation, f is not smooth and the noisy gradients must be replaced by noisy subgradients. The theory uses a limiting differential inclusion instead of a limiting ODE to prove a.s. convergence.

Often we do not even have access to the gradient of f, but must compute it approximately, say using finite differences. We obtain then an algorithm of the type

$$x_{n+1} = x_n + \gamma_n[-\nabla f(x_n) + b_n + D_{n+1}],$$

where $\{b_n\}$ is the additional error in the gradient estimation. If we have $\sup_n \|b_n\| < \epsilon_0$ for some small ϵ_0, then the iterates converge a.s. to a small neighborhood of some point in H, in fact of a local minimum. The first such scheme goes back to Kiefer and Wolfowitz [KW52], who used a central difference approximation. Denoting v^i the i^{th} coordinate of a vector $v \in \mathbb{R}^d$, and e_i the i^{th} unit vector in $\mathbb{R}^d$, we have

$$x_{n+1}^i = x_n^i + \gamma_n \left[-\left(\frac{f(x_n + \delta e_i) - f(x_n - \delta e_i)}{2\delta} \right) + D_{n+1}^i \right],$$

where $\delta > 0$ is a small positive scalar. An issue with this algorithm is that it requires $2d$ function evaluations, and using one-sided differences still requires $d+1$ function evaluations, which might still be too costly. A nice development in this context is the *simultaneous perturbation stochastic approximation* (SPSA) due to Spall. A basic version of this method considers random variables $\Delta_n \in \mathbb{R}^d$ i.i.d., with Δ_n independent of $D_1, \ldots, D_{n+1}$ and $x_0, \ldots, x_n$ and $P(\Delta_m^i = 1) = P(\Delta_m^i = -1) = \frac{1}{2}$. Then replace the algorithm above by

$$x_{n+1}^i = x_n^i + \gamma_n \left[- \left(\frac{f(x_n + \delta\Delta_n) - f(x_n)}{\delta \Delta_n^i} \right) + D_{n+1}^i \right],$$

which requires only two function evaluations. By Taylor's theorem, for each i,

$$\frac{f(x_n + \delta\Delta_n) - f(x_n)}{\delta \Delta_n^i} \approx \frac{\partial f}{\partial x^i}(x_n) + \sum_{j \neq i} \frac{\partial f}{\partial x^i}(x_n) \frac{\Delta_n^j}{\Delta_n^i}.$$

Now the expected value of the second term above is zero, and so it acts just like another noise term that can be included in D_{n+1} for the purpose of analysis.

A type of applications quite close to our subject considers the optimization of an expected performance measure

$$J(\theta) = E_\theta[f(X)],$$

where X is a random variable with a distribution F_θ that depends on a parameter θ to be adjusted in order to minimize $J(\theta)$ (in our context, θ is a policy). Now it is typically difficult to compute $J(\theta)$, but if we fix $\theta = \theta_n$, we can generate samples $f(X)$ with X distributed according to F_{θ_n}. Suppose that the laws μ_θ corresponding to F_θ (i.e., $\mu_\theta([-\infty, x)) = F_\theta(x)$ for real values random variables) are all uniformly continuous with respect to a probability measure μ, i.e., $d\mu_\theta(x) = \Lambda_\theta(x) d\mu(x)$, where the likelihood ratio $\Lambda_\theta(x)$ (or Radon-Nykodym derivative) is continuously differentiable in θ. Then

$$J(\theta) = \int f(x) d\mu_\theta(x) = \int f(x) \Lambda_\theta(x) d\mu(x).$$

If the interchange of expectation and differentiation can be justified, then

$$\frac{d}{d\theta} J(\theta) = \int f \frac{d}{d\theta} \Lambda_\theta d\mu,$$

and the stochastic approximation

$$\theta_{n+1} = \theta_n + \gamma_n [f(X_{n+1}) \frac{d}{d\theta} \Lambda_\theta(X_{n+1})|_{\theta=\theta_n}]$$

will track the ODE

$$\dot{\theta}(t) = \frac{d}{d\theta} J(\theta),$$

which is again a gradient flow converging asymptotically to a local minimum of J. This method is called the *likelihood ratio method* and is used in stochastic control to do gradient descent in the space of policies, see section 17.6. Another close idea is *infinitesimal perturbation analysis* (IPA).

Stochastic Fixed Point Iterations

A stochastic approximation of the form

$$x_{n+1} = x_n + \gamma_n[F(x_n) - x_n + D_{n+1}] \tag{15.7}$$

can be used to converge to a solution x^* of the equation $F(x^*) = x^*$, i.e., to a fixed point of F. The limiting ODE of (15.7) is

$$\dot{x}(t) = F(x(t)) - x(t). \tag{15.8}$$

We consider the case where F is an α-contraction ($0 \leq \alpha < 1$) with respect to a weighted norm on $\mathbb{R}^d$

$$\|x\|_{p,w} := \left(\sum_{i=1}^{d} w_i |x_i|\right)^{1/p},$$
$$\text{or} \quad \|x\|_{\infty,w} := \max w_i |x_i|,$$

where $w = [w_1, \ldots, w_d]^T$ with $w_i \geq 0$ for all i. Recall the Banach fixed point theorem 6.4.1 which says that a contraction has a unique fixed point. To analyze the behavior of the ODE (15.8), where F is an α-contraction with fixed point x^*, we consider the Lyapunov function $V(x) = \|x - x^*\|_{p,w}$ for $x \in \mathbb{R}^d$ (the notation includes the case $p = \infty$). Note that the only equilibrium of (15.8) is x^* and the only constant trajectory is $x(\cdot) \equiv x^*$.

Theorem 15.2.1. *The function $t \to V(x(t))$ is a strictly decreasing function of t for any non-constant trajectory of (15.8).*

Corollary 15.2.2. *x^* is the unique globally asymptotically stable equilibrium of (15.8).*

Proof of the theorem. We start with the case $1 < p < \infty$. Define $\text{sgn}(x) = +1, -1$, or 0 depending on whether $x > 0$, $x < 0$, or $x = 0$. For $x(t) \neq x^*$, we

have

$$
\begin{aligned}
&\frac{d}{dt}V(x(t)) \\
=&\frac{1}{p}\left(\sum_{i=1}^{d} w_i|x_i(t)-x_i^*|^p\right)^{(1-p)/p} \times \\
&p\left(\sum_{i=1}^{d} w_i \mathrm{sgn}(x_i(t)-x_i^*)|x_i(t)-x_i^*|^{p-1}\ \dot{x}_i(t)\right) \\
=&\|x(t)-x^*\|_{p,w}^{1-p}\left(\sum_{i=1}^{d} w_i \mathrm{sgn}(x_i(t)-x_i^*)|x_i(t)-x_i^*|^{p-1}(F_i(x(t))-x_i(t))\right) \\
=&\|x(t)-x^*\|_{p,w}^{1-p}\left(\sum_{i=1}^{d} w_i \mathrm{sgn}(x_i(t)-x_i^*)|x_i(t)-x_i^*|^{p-1}(F_i(x(t))-F_i(x^*)\right) \\
&-\|x(t)-x^*\|_{p,w}^{1-p}\left(\sum_{i=1}^{d} w_i |x_i(t)-x_i^*|^{p-1}\mathrm{sgn}(x_i(t)-x_i^*)(x_i(t)-x_i^*)\right) \\
=&\|x(t)-x^*\|_{p,w}^{1-p}\left(\sum_{i=1}^{d} w_i \mathrm{sgn}(x_i(t)-x_i^*)|x_i(t)-x_i^*|^{p-1}(F_i(x(t))-F_i(x^*)\right) \\
&-\|x(t)-x^*\|_{p,w} \\
\leq&\|x(t)-x^*\|_{p,w}^{1-p}\ \|x(t)-x^*\|_{p,w}^{p-1}\ \|F(x(t))-F(x^*)\|_{p,w}-\|x(t)-x^*\|_{p,w} \\
\leq&-(1-\alpha)\|x(t)-x^*\|_{p,w},
\end{aligned}
$$

where the first inequality is obtained using Hölder's inequality, valid for $1 < p < \infty$. Hence the time derivative is strictly negative for $x(t) \neq x^*$, which proves the claim for $1 < p < \infty$. The inequality can be written, for $t > s \geq 0$, as

$$\|x(t)-x^*\|_{p,w} \leq \|x(s)-x^*\|_{p,w} - (1-\alpha)\int_s^t \|x(\tau)-x^*\|_{p,w}d\tau.$$

The claim then follows for $p = 1$ and $p = \infty$ by continuity of $p \to \|x\|_{p,w}$ on $[1, \infty]$. □

Explanation of Collective Behaviors

Learning in Games One well studied learning mechanism for games is the "fictitious play" model introduced by Brown [Bro51]. In the simplest setting, let us consider two agents that play repeatedly a game in which two strategy choices are available for each of them at each time, say $\{s_1, t_1\}$ for agent 1 and $\{s_2, t_2\}$ for agent 2. If the (noncooperative) agents choose a strategy pair (ξ_n^1, ξ_n^2) at time n, agent i receives a payoff $h_i(\xi_n^1, \xi_n^2)$, for $i = 1, 2$. Define the empirical frequency for each player

$$\nu_i(n) := \frac{\sum_{t=1}^{n} 1\{\xi_t^i = s_i\}}{n}, i = 1, 2; n \geq 0,$$

i.e., $\nu_i(n)$ is the frequency with which player i played strategy 1 up to time n. In the fictitious play model, an agent records the empirical frequency of its opponent and plays at each stage the best response assuming the the opponent chooses its strategy randomly according to its empirical frequency[4]. This best response for player i is a map $f_i(p_{-i}) : [0,1] \to [0,1]$ which, based on the one stage game, prescribes the probability with which player i should choose its strategy s_i if the probability that its opponent chooses s_{-i} is p_{-i}. In the model, the empirical frequencies then evolve according to

$$\nu_i(n+1) = \nu_i(n) + \frac{1}{n+1}(1\{\xi^i_{n+1} = s_i\} - \nu_i(n)),\ i = 1,2$$

and the corresponding ODE is

$$\begin{aligned}\dot{\nu}_1(t) &= f_1(\nu_2(t)) - \nu_1(t)\\ \dot{\nu}_2(t) &= f_2(\nu_1(t)) - \nu_2(t).\end{aligned}$$

An equilibrium of this ODE is by definition a *Nash equilibrium*, and so the goal is to understand under which circumstances the fictitious play model converges to the players playing a Nash equilibrium. The 2 player 2 strategy case is fairly well understood, but in general the ODEs obtained from game theoretical models can have quite complex dynamics and further assumptions of the right hand side must typically be made.

Averaging (Consensus) Under Stochastic Perturbations Another well-studied algorithm is the averaging algorithm in a multiagent system. We have n agents starting with an initial value $x_i(0), i = 1, \ldots, n$. Often the problem is motivated by saying that the agents should asymptotically one a common value, but from an engineering perspective this is not well defined. First we need to rule out the trivial solution that has all agents agree on (say) 0. In the distributed algorithm literature, this is usually done by requiring that the final value be one of the initial value. Then in the synchronous setting considered here, there is again a trivial algorithm that chooses the maximum of the initial values. Most of the recent related literature instead studies variants of the following successive averaging scheme

$$x_i(k+1) = x_i(k) + \epsilon \sum_{j \in \mathcal{N}_i} (x_j(k) - x_i(k)),\ i = 1, \ldots, n, \tag{15.9}$$

where $\mathcal{N}_i$ represents the neighbors of i as specified by a graph for example, and ϵ is a small positive constant used to obtain convergence. This variant is often justified by saying that terminal value is required to be the average of the initial values, but perhaps a more convincing argument is it see this simple update rule as again an explanation of opinion formation in social systems, much like fictitious play, instead of a practical engineering tool.

[4]This is clearly not an optimal strategy. The point is that the economics literature attempts to argue that it is a reasonable model of observed behavior.

Now we can consider many variations of the basic averaging rule (15.9). For example, suppose that at period k the communication link from j to i fails with probability $1 - p_{ij}$. This probability can be made dependent on the past and time dependent without change, but for simplicity, let us assume here that the failures are i.i.d. Moreover, let's assume that the difference $x_j(k) - x_i(k)$ in (15.9) is also perturbed by a zero mean noise $\nu_{ij}(k)$ (due say to quantization or communication errors), independent of the random link failures. The perturbed averaging rule becomes then

$$x_i(k+1) = x_i(k) + \epsilon_k \sum_{j \in \mathcal{N}_i} [\delta_{ij}(k+1)(x_j(k) - x_i(k) + \nu_{ij}(k+1))], \ i = 1, \ldots, n.$$

where $\{\delta_{ij}(k)\}_k$ is i.i.d. Bernoulli, with $P(\delta_{ij}(k) = 1) = p_{ij}$, and we allow for a time-varying (typically diminishing) step size ϵ_k. Under broad conditions, this stochastic approximation tracks asymptotically the corresponding ODE

$$\dot{x}_i(t) = \sum_{j \in \mathcal{N}_i} p_{ij}(x_j(t) - x_i(t)).$$

Note that the set of equilibria of this equation is the one-dimensional subspace $x_1 = \ldots = x_n$ under reasonable conditions on the underlying connectivity graph and failure probabilities, hence consensus is obtained asymptotically. However, the choice of step sizes, as often in such simple stochastic approximation algorithms, has a strong influence on the practical (transient) behavior of the trajectories, see Fig. 15.2. One can also study asynchronous versions of the averaging algorithm using the ODE method, which is perhaps more useful from an engineering point of view.

15.3 Basic Convergence Analysis via the ODE Method

We will discuss a basic convergence analysis result, first for a special case of the stochastic recurrence (15.4) with no bias term b_n

$$x_{n+1} = x_n + \gamma_n[f(x_n) + D_{n+1}], \ n \geq 0, \ x_0 \text{ prescribed } (x_0 \text{ can be random}). \tag{15.10}$$

The following assumptions are made for the analysis

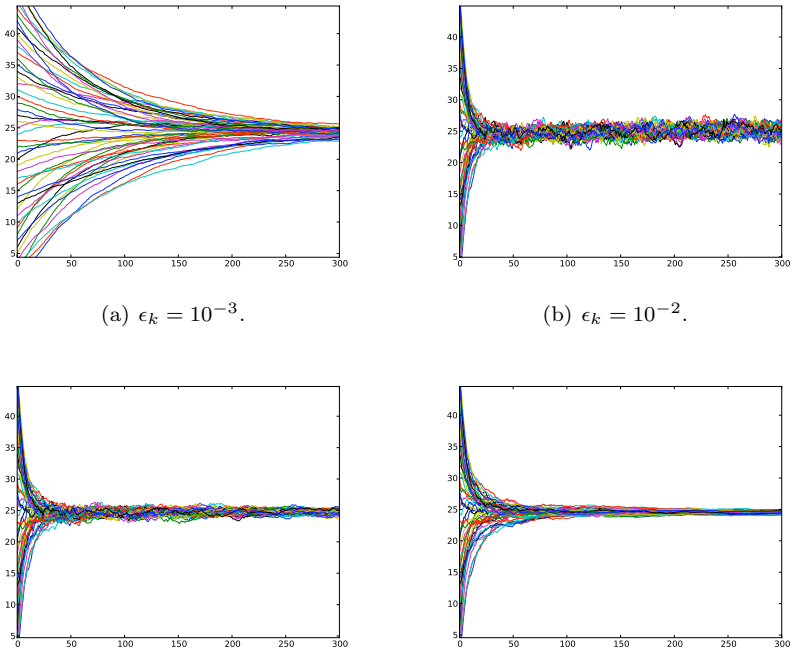
1. The map $f : \mathbb{R}^d \to \mathbb{R}^d$ is Lipschitz: $\|h(x) - h(y)\| \leq L\|x - y\|$ for some $0 < L < \infty$.

2. The stepsizes are positive scalars satisfying

$$\sum_n \gamma_n = \infty, \ \sum_n \gamma_n^2 < \infty.$$

3. $\{D_n\}$ is a *martingale difference sequence* with respect to the increasing family of σ-fields (filtration, or history generated by the sequence of random variables)

$$\mathcal{F}_n = \sigma(x_m, D_m, m \leq n) = \sigma(x_0, D_0, \ldots, D_n), n \geq 0.$$



(a) $\epsilon_k = 10^{-3}$.

(b) $\epsilon_k = 10^{-2}$.

(c) $\epsilon_k = 10^{-2}/(1 + 0.01k)$.

(d) $\epsilon_k = 10^{-2}/(1 + 0.05k)$.

Figure 15.2: Transient behavior of the local averaging algorithm for different choices of step sizes. If we choose a constant step size, increasing it improves the convergence speed but the communication noise is not well filtered. Decreasing step sizes for larger values of k improves the asymptotic filtering property of the algorithm, but can also reduce the convergence speed if decreasing too fast. In fact for constant step sizes in this problem, we only obtain asymptotic convergence in a neighborhood of the limit set of the ODE.

This means that

$$E[D_{n+1}|\mathcal{F}_n] = 0 \text{ a.s.}, n \geq 0. \tag{15.11}$$

Furthermore $\{D_n\}$ are square-integrable with

$$E[\|D_{n+1}\|^2|\mathcal{F}_n] \leq K(1 + \|x_n\|^2) \text{ a.s.}, n \geq 0,$$

for some constant K.

4. The iterates of (15.10) remain bounded a.s., i.e.,

$$\sup_n \|x_n\| < \infty, a.s.$$

The "assumption" 4 is not easy to establish in general, and specific techniques must be developed to verify it for different problems. Sometimes the

analysis is done by artificially forcing the iterates to remain bounded in (15.10) (say by truncation), which can actually be a useful device in applications. This requires to consider a limiting ODE with reflection terms on the domain boundary [KY03]. But for general unbounded state spaces this is a stability assumption that must be proved separately perhaps via other means than the ODE method, e.g. a method based on stochastic Lyapunov functions [KY03]. Under the stability assumption, the iterates (15.10) are expected to track asymptotically the ODE

$$\dot{x}(t) = f(x(t)),\ t \geq 0. \tag{15.12}$$

Assumption 1 ensures that this ODE has a unique solution for each $x(0)$, which depends continuously on $x(0)$. The martingale difference assumption (15.11) is a more precise definition of our earlier assumption of zero-mean noise D_n. We allow conditioning on the past iterates, so this is a quite general set-up. Any deterministic trend in the noise should be captured in f or the bias terms b_n in (15.4).

To make the idea that the stochastic approximation asymptotically tracks the trajectories of the ODE more formal, first define the sequence of times

$$t_0 = 0,\ t_n = \sum_{m=0}^{n-1} \gamma_m.$$

We construct a continuous time trajectory $\bar{x}(t)$ interpolating the iterates $\{x_n\}$ at times $\{t_n\}$ and show that this trajectory almost surely approaches the solution set of the ODE (15.12). That is,

$$\bar{x}(t_n) = x_n, n \geq 0,$$

and $\bar{x}$ is piecewise linear, which defines it on the intervals $[t_n, t_{n+1}]$. We see now that the assumption $\sum_{m=0}^{\infty} \gamma_m = \infty$ is necessary in order to cover the whole time axis and be able to track the ODE asymptotically. Next define for $s \geq 0$ the unique solution x^s of the ODE (15.12), defined for $t \geq s$, with initial condition $x^s(s) = \bar{x}(s)$.

We now give a relatively general set of results, mostly without proofs.

Lemma 15.3.1. *For any $T > 0$,*

$$\lim_{s \to \infty} \sup_{t \in [s, s+T]} \|\bar{x}(t) - x^s(t)\| = 0,\ \textit{a.s.}$$

Thus as $s \to \infty$, the interpolated trajectory $\bar{x}$ starting from s remains arbitrarily close to a trajectory of the ODE, a formalization of the idea that the noise becomes asymptotically too weak to push the iterates away from the trajectories of the ODE. A general convergence theorem for stochastic approximations is given below.

Theorem 15.3.2. *Assume that the assumptions 1-4 are satisfied. Then almost surely, the sequence $\{x_n\}$ generated by (15.10) converges to a (possibly sample path dependent) compact connected internally chain transitive invariant set of the ODE (15.12).*

Note that the chain transitive invariant set of the theorem can be much larger than the ω-limit set of the ODE, because it must essentially be "stable under small perturbations", recall Fig. 15.1. In practice, Lyapunov functions are useful for further narrowing down the potential candidates for the limit set. Suppose that we have a Lyapunov function $V : \mathbb{R}^d \to [0, \infty)$, continuously differentiable, such that $\lim_{\|x\|\to\infty} V(x) = \infty$, $H = \{x \in \mathbb{R}^d : V(x) = 0\} \neq \emptyset$, and $\frac{d}{dt}V(x(t)) = \langle \nabla V(x), f(x)\rangle \leq 0$ with equality if and only if $x \in H$. Then we have the following corollary, under the same assumptions as for the theorem.

Corollary 15.3.3. *Almost surely the sequence $\{x_n\}$ converges to an internally chain transitive invariant set contained in H.*

Proof. Consider a sample sequence $x_0, x_1, \ldots$ (on the probability 1 set where the assumptions are satisfied). Let $C' = \sup_n \|x_n\|$ and $C = \sup_{\|x\|\leq C'} V(x)$. Define the level sets of V by $H^a = \{x \in \mathbb{R}^d : V(x) < a\}$, and note that $\bar{x}(t) \in \bar{H}^C$ for all $t \geq 0$, where $\bar{H}^a$ is the closure of H^a. Fix $0 < \epsilon < C/2$. Then let

$$\Delta := \min_{x\in\bar{H}^C\setminus H^\epsilon} |\langle \nabla V(x), h(x)\rangle| > 0.$$

$\Delta > 0$ is a consequence of $\bar{H}^C \setminus H^\epsilon$ being compact and ∇V and h continuous. Hence any trajectory of the ODE starting in H^C reaches H^ϵ in time at most $T := C/\Delta$. By uniform continuity of V on compact sets, we can choose $\delta > 0$ such that for $x \in \bar{H}^C$ and $\|x - y\| < \delta$, we have $|V(x) - V(y)| < \epsilon$. Then by lemma 15.3.1, there is a t_0 such that for all $t \geq t_0$, $\sup_{s\in[t,t+T]} \|\bar{x}(s) - x^t(s)\| < \delta$. Hence for all $t \geq t_0$, we have $|V(\bar{x}(t + T)) - V(x^t(t + T))| < \epsilon$. Since $x^t(t + T) \in H^\epsilon$, we obtain $\bar{x}(t + T) \in H^{2\epsilon}$. So for all $t \geq t_0 + T$, $\bar{x}(t) \in H^{2\epsilon}$. Since ϵ can be chosen arbitrarily small, it follows that $\bar{x}(t) \to H$ as $t \to \infty$. □

The following corollary is immediate.

Corollary 15.3.4. *If the only internally chain transitive invariant sets for (15.12) are isolated equilibrium points, then $\{x_n\}$ a.s. converges to a possibly sample path dependent equilibrium point.*

Remark on the assumption $\sum_{n=0}^{\infty} \gamma_n^2 < \infty$

Consider the cumulative noise term

$$\zeta_n = \sum_{m=0}^{n-1} \gamma_m D_{m+1}, n \geq 1.$$

in (15.10). We want to show that the effect of noise becomes negligible asymptotically, as this is a basic ingredient to prove lemma 15.3.1. Note that ζ_n is a (zero mean) *martingale*, i.e.

$$E[\zeta_{n+1}|\mathcal{F}_n] = \zeta_n,\ n \geq 1,$$

which follows immediately from assumption 3. The definition of a martingale also requires ζ_n to be $\mathcal{F}_n$ measurable, which is immediate, and integrable. In fact in this case ζ_n is even square integrable, i.e., $E[\|\zeta_n\|^2] < \infty$ for all n, which is a consequence of assumptions 1 and 3. Moreovever,

$$\begin{aligned}\sum_{n\geq 0} E[\|\zeta_{n+1} - \zeta_n\|^2 | \mathcal{F}_n] &= \sum_{n\geq 0} \gamma_n^2 E[\|M_{n+1}\|^2 | \mathcal{F}_n] \\ &\leq \sum_{n\geq 0} \gamma_n^2 K(1 + \|x_n\|^2), \quad \text{a.s.} \\ &\leq K(1+B^2)\left(\sum_{n\geq 0} \gamma_n^2\right) < \infty, \quad \text{a.s.}\end{aligned}$$

where $B = \sup_n \|x_n\| < \infty$ from assumption 4. We can then apply the Martingale convergence theorem to conclude that ζ_n converges almost surely as $n \to \infty$. In particular, the noise entering in the iterations after time K, i.e., $\sum_{n=K}^{\infty} \gamma_n D_{n+1}$, vanishes as $K \to \infty$. This ensures that the effect of the noise indeed becomes asymptotically negligible. Note here that this property relies on the assumption $\sum_{n\geq 0} \gamma_n^2 < \infty$, which is important to obtain a general theorem[5].

[5]The case of constant step sizes, which does not satisfy this assumption, is also well studied