

## Chapter 3

# Stochastic Approximation, Exogenous Noise Model

### 3.1 Motivation

The previous chapters dealt with the analysis of algorithms of the form

$$\theta_{n+1} = \theta_n + \epsilon_n G(\theta_n)$$

designed to “find the zeroes” of a function  $G$ , under appropriate assumptions. However, it is often the case that the function  $G(\theta)$  is not known with precision, and must be *estimated* or generated via *computer simulation*. This is the situation when optimising real systems in telecommunications, transportation, and robotics; or finding optimal inventory and maintenance policies; or finding economic equilibria, and a number of other important applications.

**Definition 3.1** A recursion of the form  $\theta_{n+1} = \theta_n + \epsilon_n Y_n$ , where  $\{Y_n\}$  is a stochastic process, is called a stochastic approximation.

In this chapter and the next we will present an overview of the theory of stochastic approximations, which establishes the conditions under which such algorithms converge as  $n \rightarrow \infty$ ; how to characterize the limits; in which sense do they converge, and at what rate, when the sequence  $\{Y_n\}$  is a random process, possibly interdependent and dependent on the sequence of values  $\theta_m, m \leq n$ . For each  $n$ ,  $Y_n$  should be built as an estimator of the (unknown) function  $G(\theta_n)$ , in a manner which we will make precise later on. Observe that given the information up to time  $n$ , i.e.,  $\mathfrak{F}_n$ , the value of  $\theta_{n+1}$  is known and so is  $G(\theta_{n+1})$ .

**EXAMPLE 3.1.** To illustrate the framework for stochastic approximation, suppose that we wish to solve the problem treated in Example 2.2, namely  $\min_{\theta \in \mathbb{R}^2} J(\theta)$ , where  $J(\theta) = 2\theta_1^2 + \theta_2^2$ , for  $\theta = (\theta_1, \theta_2)^\top$ . To solve this problem we would like to follow the iterative algorithm  $\theta_{n+1} = \theta_n - \epsilon \nabla J(\theta_n)^\top$ , with  $\theta_n \in \mathbb{R}^2$ , which has a limit behavior as the ODE

$$\begin{aligned}\frac{dx(t)}{dt} &= -4x(t) \\ \frac{dy(t)}{dt} &= -2y(t).\end{aligned}$$

However, suppose that we do not know  $J$  or  $\nabla J$  in closed form. Rather, what we can observe are measurements of the gradient with additive noise. As a simple example, suppose that for each value of  $\theta = (\theta_1, \theta_2)^\top \in \mathbb{R}^2$  we can somehow measure the noisy observations

$$Y_{n,1}(\theta) = -4\theta_1 + \xi_{n,1}$$

$$Y_{n,2}(\theta) = -2\theta_2 + \xi_{n,2},$$

where  $(\xi_{n,1}, \xi_{n,2}), n \geq 0$ , are independent normally distributed, zero-mean random variables (white noise model). To illustrate the situation, we have made simulations where  $\epsilon = 0.01$  and  $\xi_{n,i} \sim \mathcal{N}(0, 15)$ , independent of the value of  $\theta_n$  and of previous values of the observations. While this model is known by the simulator, it is assumed that the “controller” who builds the estimates does not know it, but only has access to the observations, or samples. The sequence  $\{\xi_n\}$  is called an *underlying process*. There are two “reasonable” approaches to follow:

- Because we would like to follow the ODE, we may attempt at sampling “a lot” of observations at each value of  $\theta_n$  so that the variance of the estimator of  $\nabla J(\theta_n)$  is “small”.
- More simply, we may replace the desired quantity  $\nabla J(\theta_n)$  at iteration  $n$  by its estimated value using only one observation per update.

Figure 3.1(a) shows the limit ODE as we have seen in Chapter 2. The first approach is illustrated in Figure 3.1(b), where each update is built using sample averages with 10 observations of the noise.

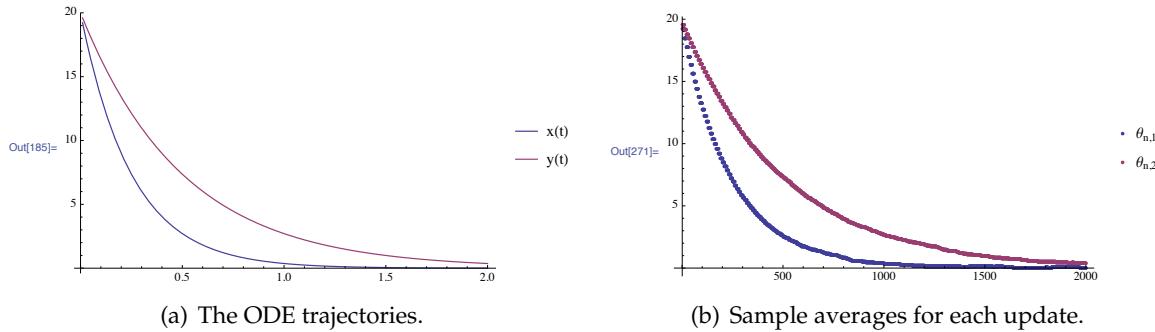


Figure 3.1: Approximations with noise.

Figure 3.2(a) illustrates a typical trajectory when using only one observation per update. It looks clearly worse as an approximation to the ODE than when several consecutive observations are used per update. However, the sample average approach requires slowing down the algorithm considerably, which affects particularly the first iterations of the algorithm, when it stays at “wrong” values of the estimates for longer than necessary. To compare, Figure 3.2(b) shows the same trajectory as in Figure 3.1(b) that averages 10 consecutive values, but only for the same amount of iterations of the algorithm as in Figure 3.2(a). Note that convergence occurs faster in Figure 3.2(a) than in Figure 3.1(b).

\*\*\*

Chapters 3 and 4 will present the necessary theoretical methods in order to understand the behavior of such iterative algorithms in the presence of noise. Emphasis is on understanding under which conditions do the noisy iterations “converge” (and in what sense) to the target limit ODE. In

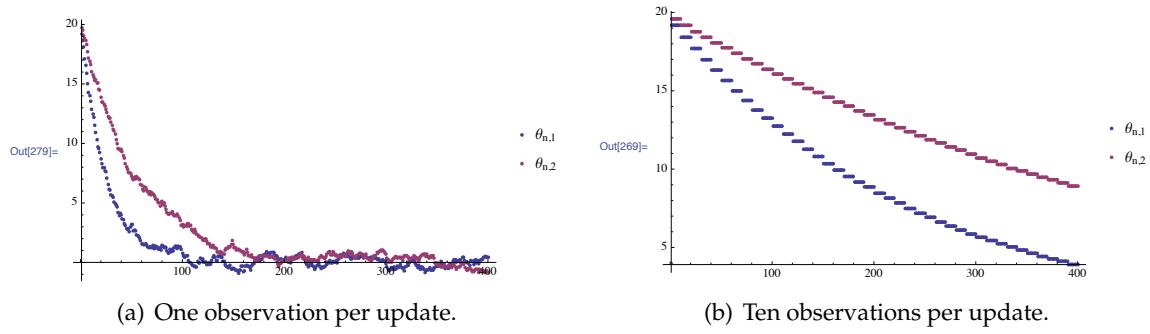


Figure 3.2: Approximations using one or several samples per update

particular we will address the exogenous noise model in Chapter 3 and the endogenous noise model in Chapter 4. As the name implies, exogenous noise deals with the situation when the noise in the observations comes “from outside” (as in Example 3.1), rather than from the sequence  $\theta_n$  itself. The endogenous noise model comprises situations that are more general, when there is feedback in the estimation and update procedure. In this case the noise of observation  $n$  is affected by the previous history of the trajectories and a more complex mathematical model for the underlying processes has to be used.

### 3.2 The Robbins Monro Procedure

Tracking a target's position is an important problem with applications in robotics, defence, estimation, etc. Suppose that we wish to estimate the location  $\theta^* \in \mathbb{R}^3$  of a target. Denoting the distance from the target by

$$J(\theta) = \frac{1}{2}(\theta - \theta^*)^\top(\theta - \theta^*),$$

the problem can be stated as the unconstrained optimisation problem

$$\min_{\theta} J(\theta).$$

The gradient-based algorithm from (1.42) would require updating the current estimate via the recursion  $\theta_{n+1} = \theta_n - \epsilon_n \nabla J(\theta_n) = \theta_n - \epsilon_n(\theta_n - \theta^*)$ , but we don't know  $\theta^*$ . Suppose that we can obtain noisy measurements (through radar, satellite, statistics, etc) of  $\theta^*$ .

The  $n$ -th observation is denoted  $\xi_n$ . In the simplest setting,  $\mathbb{E}[\xi_n] = \theta^*$ , so that the observations are unbiased. Using these estimates, one obtains:

$$\theta_{n+1} = \theta_n - \epsilon_n(\theta_n - \xi_n),$$

where  $\mathbb{E}[\theta_n - \xi_n] = \nabla J(\theta_n)$ . The above equation describes the familiar model called “regression to the mean”. Notice that for this particular model  $G(\theta) = -(\theta - \theta^*)$  which has a unique zero at the point  $\theta^*$ , as desired.

An important problem related to target tracking is that of *supervised learning*. Here, the random input  $X$  is transformed using a known parametrized transformation  $v(\theta, X)$ . The problem is to find

the value  $\theta^*$  such that  $L(\theta^*) = \mathbb{E}[v(\theta^*, x)] = \alpha$ , where  $\alpha$  is a known target. This is known as an “inverse” problem.

**EXAMPLE 3.2.** A thermostat is to be set up to control a heating device in order to reach a temperature of  $\alpha$  degrees in the room. The amount of power  $\theta$  used by the heating element yields a room temperature  $L(\theta)$  in the long run, so we seek  $\theta^*$  such that  $L(\theta^*) = \alpha$ , but the function  $L(\cdot)$  is unknown (it depends on the room characteristics, the environment, time of day, etc).

A first approach to this problem would be to “train” the thermostat. In a controlled environment, the device would be set to work under various settings for an *input vector* consisting of the room size, humidity, and some outdoors conditions. A series of statistical tests would estimate the inverse problem, that is, the values of  $\theta^*$  for each value of  $\alpha$  within a reasonable range, as a function of the input vector. This approach would then require the user to choose values for the input vector as well as for the desired room temperature, and the device would simply read the corresponding entry from a pre-set table. While this static estimation may be accurate for some problems, it is often the case that many important variables that affect the function  $L(\theta)$  may not be easy to measure by the regular household user. He or she may prefer a more “intelligent” device that requires exclusively the choice of the target room temperature. Because the thermostat is to be designed to work on different environments and different rooms, a “learning algorithm” can be programmed to adjust the power with minimal input information from the user. One such scheme is precisely a gradient-based optimization algorithm as we will describe shortly.

Realistically, we must assume that the measurements of the thermometer are imprecise for the actual room temperature  $L(\theta)$ , so that at value  $\theta$ , the  $n$ -th reading is  $\xi_n(\theta) = L(\theta) + w_n$ , where the sequence  $\{w_n, n \in \mathbb{N}\}$  models the “noise”. A common assumption (although not realistic in many cases) is that they are zero mean independent random variables.

\*\*\*

Let

$$J(\theta) = \frac{1}{2}(L(\theta) - \alpha)^\top(L(\theta) - \alpha),$$

for  $L: \mathbb{R}^d \rightarrow \mathbb{R}^k$  and  $\alpha \in \mathbb{R}^k$ . A general formulation of the problem of Example 3.2 as an optimisation problem is:

$$\min_{\theta} J(\theta),$$

with  $k$  targets for the tracking problem, and it is assumed that there exists  $\theta^* \in \mathbb{R}^d$  such that  $L(\theta^*) = \alpha$ . The gradient-based algorithm has the deterministic formulation:

$$\theta_{n+1} = \theta_n - \epsilon_n \nabla J(\theta)^\top = \theta_n - \epsilon_n ((L(\theta_n) - \alpha)^\top \nabla L(\theta_n))^\top.$$

In a number of cases, the function  $L(\cdot)$  is monotone non-decreasing (or non-increasing) in each of its components. As we will explain in the following, in this case a stochastic approximation can be built without any gradient information. In Example 3.2, where  $d = k = 1$ , more power for the heating element yields a higher room temperature. For cooling elements, more energy is required to decrease the room temperature, so that the function  $L(\cdot)$  is monotonic. Under monotonicity, the matrix  $e = \text{sign}(\nabla L(\theta))$  is independent of  $\theta$  and has deterministic fixed values. Because neither  $L(\theta)$  nor its gradient are known, one can use instead the recursion:

$$\theta_{n+1} = \theta_n - \epsilon_n ((\widehat{L}(\theta_n) - \alpha)^T e)^\top, \quad (3.1)$$

where  $\widehat{L}(\theta)$  is an estimator of  $L(\theta)$ . Relating back to Chapters 1 and 2, under strict monotonicity, the chosen direction, though not the negative gradient, is always a descent direction. The stationary point  $\theta^*$  is unique and it is also the only stable point of the vector field driving the learning algorithm.

In their important paper [26], Robbins and Monro studied the convergence of this algorithm to  $\theta^*$ , for the one dimensional case. The following result summarises their findings.

**Theorem 3.1 (Robbins-Monro)** *Assume that there is a unique root  $\theta^*$  to the equation  $f(\theta) = \alpha$ , and that  $f(\cdot)$  is non-decreasing, with  $f'(\theta^*) > 0$ . Let  $\xi_n = \widehat{f}(\theta_n)$  denote the  $n$ -th measurement, and assume that  $\{\xi_n\}$  are independent, that  $\mathbb{E}[\xi_n | \theta_n = \theta] = f(\theta)$ , and that  $|\xi_n| \leq C < \infty$  is bounded a.s. Let  $\epsilon_n$  be a sequence satisfying (3.2):*

$$\sum_{n=1}^{\infty} \epsilon_n = +\infty, \quad \sum_{n=1}^{\infty} \epsilon_n^2 < \infty. \quad (3.2)$$

*Then the stochastic approximation:*

$$\theta_{n+1} = \theta_n - \epsilon_n(\xi_n - \alpha) \quad (3.3)$$

*converges in probability as  $n \rightarrow \infty$ . Specifically,  $\mathbb{E}[(\theta_n - \theta^*)^2] \rightarrow 0$ .*

In their paper, Robbins and Monro establish more general conditions under which the result holds, including the case where the function  $f(\cdot)$  has a jump at  $\theta^*$ . Let  $b_n = \mathbb{E}[(\theta_n - \theta^*)^2]$ . The method of proof is based on the expression:

$$b_{n+1} = b_n + \epsilon_n^2 \mathbb{E}[f(\theta_n)] - 2\epsilon_n \mathbb{E}[(\theta_n - \theta^*)(f(\theta_n) - \alpha)],$$

which they study using the method of Lemma 1.1 to establish convergence of  $b_n$  to zero.

In Example 3.2, it is assumed that *accurate* measurements of the steady temperature  $L(\theta)$  are available. Specifically, it is assumed that  $\mathbb{E}[\xi | \theta_n] = L(\theta_n)$ . The function  $L(\theta)$  here represents the long term equilibrium temperature reached in the room when the heater has been working at a fixed power setting  $\theta$ . Suppose now that we use power  $\theta_n$  for, say, 2 minutes, during which the thermometer re-adjusts its measurement. After the 2 minute period, we adjust the power with (3.3) and leave the power setting at  $\theta_{n+1}$  for another 2 minutes. A problem with the classical (or “Robbins-Monro”) model is the assumption that the temperature reading is unbiased for  $L$ : is the 2 minute period sufficient for the room temperature to reach its equilibrium state? Naturally this depends on room size, but also on the actual power  $\theta_{n+1}$ , the initial room temperature  $\xi_n$  of the room and perhaps also on previous values of  $\xi_m, m < n$ . This measurement problem is an important one: “temperature” is a quantity that cannot be measured as an instantaneous response, because it is an equilibrium quantity. Not only is the temperature of the room not well defined (until it reaches equilibrium, which may take longer than the updating time), but there may be inaccuracies in the thermometer as well.

To account for more realistic models, we will study the theory of stochastic approximations in more detail, following the approach of [21]. The remainder of this chapter presents stronger convergence results for stochastic approximation for more general models for the consecutive measurements.

### 3.3 Exogenous noise model, decreasing stepsize

Let  $G(\theta)$  be a vector field and consider the problem of finding the zeros of  $G$ . It is assumed that the function  $G$  is not known in closed form, but for every value of the control  $\theta$  a noisy estimate

of the vector  $G(\theta)$  can be built (as will be made precise shortly). We will often refer to  $G(\theta)$  as the *target vector field*. In this section we present the stochastic approximation algorithm with decreasing stepsize:

$$\theta_{n+1} = \theta_n + \epsilon_n Y_n, \quad (3.4)$$

where the stepsize sequence satisfies (3.2).

## Exogenous Noise Model

The initial point of the algorithm  $\theta_0$  is assumed to be well defined on a probability space  $(\Omega, \mathfrak{F}_0, \mathbb{P})$ , where  $\mathfrak{F}_0$  contains all  $\mathbb{P}$ -null sets. The updates  $Y_n$  are random variables defined on the common probability space  $(\Omega, \mathfrak{F}, \mathbb{P})$ .  $Y_0$  is measurable wrt  $\mathfrak{F}_0 \subset \mathfrak{F}$  and  $Y_n, n \geq 1$  are assumed to be measurable with respect to

$$\mathfrak{F}_{n-1} \stackrel{\text{def}}{=} \sigma(\theta_0; Y_0, \dots, Y_{n-1}) \subset \mathfrak{F}.$$

This formulation models the algorithm as a stochastic process, where the updates may be dependent on the past, thus generalising the common models where consecutive feedback measurements are supposed to be iid. Define the *error* terms as

$$\beta_n = \mathbb{E}[Y_n | \mathfrak{F}_{n-1}] - G(\theta_n), \quad (3.5)$$

which are random variables, measurable wrt  $\mathfrak{F}_{n-1}$ . Given the information available up to “time”  $n-1$ , the expected outcome of  $Y_n$  is given by  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}]$ . Put differently,  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}]$  is the trend of  $Y_n$  in the next iteration. The difference between the trend and the realization is denoted by

$$\delta M_n \stackrel{\text{def}}{=} Y_n - \mathbb{E}[Y_n | \mathfrak{F}_{n-1}], \quad (3.6)$$

and  $\delta M_n$  thus measures the deviation of the realization  $Y_n$  from its (conditional) mean value. Notice that by construction,  $\mathbb{E}[\delta M_n] = 0$ . Put differently,  $\delta M_n$  is the noise in the measurement of  $G(\theta_n) + \beta_n$ . In the same way

$$V_n = \mathbb{E}[(\delta M)^2] = \mathbb{E}[(Y_n - \mathbb{E}[Y_n | \mathfrak{F}_{n-1}])^2] \quad (3.7)$$

expresses a measure of the variation of  $Y_n$  around its conditional mean given information  $\mathfrak{F}_{n-1}$ . The cumulative process  $\sum_{k=1}^n \delta M_k$  is, by construction, a martingale and  $\{\delta M_n\}$  is called *martingale difference process*<sup>1</sup>, see Definition B.34 in Appendix B.6. With the above notation  $Y_n$  can be decomposed into the following three components:

$$Y_n = \underbrace{G(\theta_n)}_{\text{vector field}} + \underbrace{\delta M_n}_{\text{noise}} + \underbrace{\beta_n}_{\text{error of vector field}}. \quad (3.8)$$

The noise is separated from vector field and the above model is called the exogenous noise model. In particular, in the proof of our main theorem we will elaborate on the fact that the limit of  $Y_n$  can be studied through the analysis of the limit of  $G(\theta_n)$  whereby the noise component can be treated separately. The above model is called the “martingale difference” model in [21] and the material presented in this chapter follows the approach therein.

<sup>1</sup>Let  $\{\delta M_n\}$  be adapted to filtration  $\{\mathfrak{F}_n\}$ . Then,  $\{\delta M_n\}$  is called a martingale difference process if  $\mathbb{E}[\delta M_n | \mathcal{F}_{n-1}] = 0$  a.s.

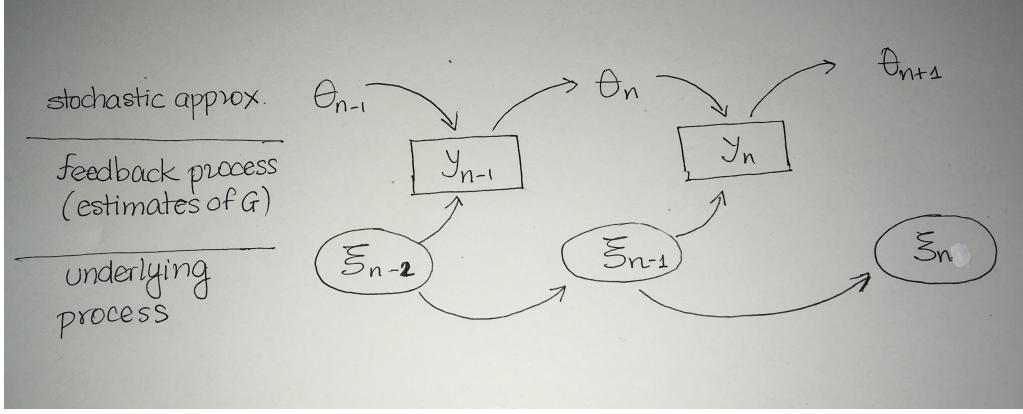


Figure 3.3: Schematic representation of processes in exogenous noise model.

**EXAMPLE 3.3.** [Linear Regression] Optimization techniques arise naturally in data analysis. Here, a vector of *inputs* ( $x_i, i = 1, \dots, N$ ) is associated with the corresponding observed *outputs* ( $\xi_i, i = 1, \dots, N$ ). The model for the system's response is given by a mapping  $h(\theta_1, \theta_2, x)$  and the errors  $\xi_i - h(\theta_1, \theta_2, x_i)$  are assumed to be zero mean, independent random variables. Not knowing the true parameters  $\theta_1^*, \theta_2^*$ , one uses the observations to find the best estimates. In the following, we will discuss linear regression and least squares approximation in detail.

In a linear regression problem,  $h$  is assumed to be given by

$$h(\theta_1, \theta_2, x) = \theta_1 + \theta_2 x, \quad (3.9)$$

and the *linear regression* problem is that of “fitting” the best values of  $\theta = (\theta_1, \theta_2)^\top \in \mathbb{R}^2$  for a specific set of observed values. Let  $x_1, \dots, x_N$  be a chosen set of design points, and  $\xi_1, \dots, \xi_N$  the corresponding (noisy) observations of the system's response  $h(\theta_1, \theta_2, x_i), i = 1, \dots, N$ . Let

$$J(\theta) = \frac{1}{2} \sum_{i=1}^N (\xi_i - \theta_1 - \theta_2 x_i)^2.$$

Then seek  $\bar{\theta}$  as the solution to the *deterministic problem*:

$$\min_{\theta \in \mathbb{R}^2} J(\theta).$$

In this problem, the data are seen as fixed (static estimation approach), and the *estimate*  $\bar{\theta}$  is the point that minimises the above expression. It depends on the design points  $X$  and the random vector of output responses  $\xi$ , and it is usually not equal to  $\theta^*$ , although for many models it can be shown that  $\mathbb{E}[\bar{\theta}] = \theta^*$ . This is commonly referred to as a “training” algorithm.

\*\*\*

**EXAMPLE 3.4.** [Learning and Dynamic Model Fitting] Let  $Z(X)$  denote the system response to input  $X$  where  $X$  is a random variable distributed in the experimentation range:  $X \in S$ . Like in linear regression, we model the system response by the mapping  $h$  given in (3.9). For  $\theta = (\theta_1, \theta_2)^\top \in \mathbb{R}^2$ , let

$$J(\theta) = \frac{1}{2} \mathbb{E}[(Z(X) - (\theta_1 + \theta_2 X))^2].$$

The problem of *least squares* is to find the parameter values that solve the optimization problem:

$$\min_{\theta \in \mathbb{R}^2} J(\theta).$$

It is left as an exercise to verify that

$$\begin{aligned}\frac{\partial}{\partial \theta_1} J(\theta) &= -\mathbb{E}[Z(X) - \theta_1 - \theta_2 X)] \\ \frac{\partial}{\partial \theta_2} J(\theta) &= -\mathbb{E}[X Z(X) - \theta_1 X - \theta_2 X^2]\end{aligned}$$

To minimize this function, Theorem 1.3 suggests to use a gradient procedure of the form

$$\theta_{n+1} = \theta_n - \epsilon_n \nabla_{\theta} J(\theta_n)^T = \theta_n + \epsilon_n \mathbb{E}[(Z(X) - \theta_{n,1} - \theta_{n,2} X)(1, X)^T],$$

but we do not know the (random) function  $Z$ . We can use estimates of  $Z(x)$  for different (possibly random) values of the input  $x$  to construct unbiased estimators of the gradient. This dynamical optimization problem arises naturally when using streaming data for reinforcement learning. Suppose that consecutive design points  $\{x_n\}$  are randomly distributed over  $S$ . For each  $x_n$  we obtain a corresponding random observation  $\xi_n = Z(x_n)$ , and notice that  $Y_n = (\xi_n - \theta_{n,1} - \theta_{n,2} x_n)(1, x_n)^T$  satisfies  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}] = -\nabla_{\theta} J(\theta_n)$  (the proof is left as an exercise, see Exercise 3.3). Then the martingale difference model has zero bias because  $G(\theta_n) = -\nabla_{\theta} J(\theta_n)$  and

$$\delta M_n = (\xi_n - h(\theta_{n,1}, \theta_{n,2}, x_n))(1, x_n)^T + \nabla_{\theta} J(\theta_n),$$

recall  $h$  from (3.9). Algorithm (3.4) then becomes:

$$\begin{aligned}\theta_{n+1,1} &= \theta_{n,1} + \epsilon_n (\xi_n - \theta_{n,1} - \theta_{n,2} x_n), \\ \theta_{n+1,2} &= \theta_{n,2} + \epsilon_n x_n (\xi_n - \theta_{n,1} - \theta_{n,2} x_n).\end{aligned}$$

This method is often referred to as the *least mean square (LMS)* or *recursive mean square* algorithm. Because the algorithm is designed to improve its estimates as it gathers more and more information, it is often referred to as a *reinforcement learning* algorithm and  $\{\epsilon_n\}$  is called the learning rate sequence (a constant step size is often used).

\*\*\*

We will now apply the basic ideas of the ODE method used in the proof of Theorem 2.3, incorporating a bias term as in Exercise 1.9. Before we do that we present a technical result that will be used in the proof of the main theorem below. This result establishes sufficient conditions for the noise in the tail of  $\delta M_n + \beta_n$  to tend to zero.

**Proposition 3.1** *Let  $G: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be Lipschitz continuous. Suppose that the feedback sequence  $\{Y_n\}$  satisfies the Martingale model, so  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}] = G(\theta_n) + \beta_n$ . Let  $\{\epsilon_n\}$  be a deterministic step size sequence and assume:*

- in case of decreasing step sizes

- (a1)  $\sum_{i=0}^{\infty} \epsilon_i = \infty$ ,
- (a2) the error terms are asymptotically negligible, in the sense that  $\sum_{i=0}^{\infty} \epsilon_i \|\beta_i\| < \infty$  w.p.1,

### 3.3. EXOGENOUS NOISE MODEL, DECREASING STEPSIZE

---

- or, in case of fixed step size  $\epsilon_i = \epsilon$

(a12) the error terms are asymptotically negligible, in the sense that  $\sum_{i=0}^{\infty} \|\beta_i\| < \infty$  a.s.,

and

(a3) the variance term (3.7) satisfies:  $\sum_{i=0}^{\infty} \epsilon_i^2 V_i < \infty$ , w.p.1.

Define the processes

$$Z_m^{(n)} = M_{n+m} - M_n = \sum_{i=n}^{n+m} \epsilon_i \delta M_i$$

$$B_m^{(n)} = \sum_{i=n}^{n+m} \epsilon_i \beta_i.$$

Then

$$\lim_{n \rightarrow \infty} \left( \sup_{m \geq 0} \|Z_m^{(n)}\| \right) = 0 \quad a.s, \quad (3.10)$$

$$\lim_{n \rightarrow \infty} \left( \sup_{m \geq 0} \|B_m^{(n)}\| \right) = 0 \quad a.s. \quad (3.11)$$

**Proof:** Result (3.11) follows directly from assumption (a1) and (a2) for decreasing stepsizes, and from (a12) for fixed step size. We now show (3.10). The goal is to show that the shifted martingale noise converges to 0 with probability 1 as  $n \rightarrow \infty$ . By construction, for each  $n$  the process  $Z_m^{(n)}$  for  $m \geq 0$ , is also a martingale. From Proposition B.3 in Appendix B.6, for non-negative convex function  $q(\cdot)$  a martingale  $W_m$  satisfies

$$\mathbb{P} \left( \sup_{n \leq \ell \leq n+m} |W_\ell| \geq \Delta \right) \leq \frac{1}{q(\Delta)} \mathbb{E}[q(W_{n+m})],$$

for  $m > 0$ . We now apply this property to the martingale  $Z_m^{(n)}$  for  $q(x) = x^2$  noticing that  $\mathbb{E}[Z_n^{(n)}] = 0$ , which yields for any  $\Delta > 0$  and  $\ell \geq 0$

$$\begin{aligned} \mathbb{P} \left( \sup_{0 \leq \ell \leq m} \|Z_\ell^{(n)}\| \geq \Delta \right) &\leq \frac{1}{\Delta^2} \mathbb{E} \left[ \left( \sum_{i=n}^{n+m} \epsilon_i \delta M_i \right)^2 \right] \\ &= \frac{1}{\Delta^2} \sum_{i=n}^{n+m} \epsilon_i^2 \mathbb{E}[(\delta M_i)^2] \\ &= \frac{1}{\Delta^2} \sum_{i=n}^{n+m} \epsilon_i^2 V_i \leq \frac{1}{\Delta^2} \sum_{i=n}^{\infty} \epsilon_i^2 V_i =: \frac{K_n}{\Delta^2}, \end{aligned}$$

where the first equality follows from the martingale property  $\mathbb{E}[\delta M_n \delta M_m] = 0$  (see Proposition B.1 in Appendix B.6) and the second equality follows from (3.7). This implies that for every  $n$ ,

$$\lim_{m \rightarrow \infty} \mathbb{P} \left( \sup_{\ell \leq m} \|Z_\ell^{(n)}\| \geq \Delta \right) \leq \frac{K_n}{\Delta^2},$$

and we now show that this implies

$$\mathbb{P} \left( \sup_{0 \leq m} \|Z_m^{(n)}\| > \Delta \right) \leq \frac{K_n}{\Delta^2}. \quad (3.12)$$

Let  $A_m$  be the event  $A_m = \{\sup_{\ell \leq m} \|Z_\ell^{(n)}\| \geq \Delta\}$ . By construction,  $A_m \subset A_{m+1}$ . Use now the continuity Theorem for increasing events (see Theorem B.1 in Appendix B.1) to establish that  $\mathbb{P}(\lim_{m \rightarrow \infty} A_m) = \lim_{m \rightarrow \infty} \mathbb{P}(A_m) \leq K_n/\Delta^2$ , which establishes (3.12).

Assumption (a3) implies that  $\lim_{n \rightarrow \infty} K_n = 0$ , so that

$$\lim_{n \rightarrow \infty} \mathbb{P} \left( \sup_{0 \leq m} \|Z_m^{(n)}\| > \Delta \right) = 0.$$

Call  $A'_n$  the event  $A'_n = \{\sup_{0 \leq m} \|Z_m^{(n)}\| > \Delta\}$ . What we have shown so far is that  $\lim_{n \rightarrow \infty} \mathbb{P}(A_n) = 0$ . What we wish to show is that  $\mathbb{P}(\lim_{n \rightarrow \infty} A_n) = 0$  in order to establish the desired convergence w.p.1. By construction, the sequence  $\{A_n\}$  is monotone decreasing. Use again Theorem B.1 and the fact that this holds for any  $\Delta > 0$  to establish the desired result:

$$\mathbb{P} \left( \lim_{n \rightarrow \infty} \sup_{n < m} \|M_{n+m} - M_n\| = 0 \right) = 1.$$

QED

**Definition 3.2** The shifted interpolation processes  $\vartheta^n(\cdot)$  are defined for each  $n \in \mathbb{N}$  by:

$$\begin{aligned} \vartheta^0(t) &= \theta_n, \quad t \in [t_n, t_{n+1}), \quad t_n = \sum_{i=0}^n \epsilon_i \\ \vartheta^n(t) &= \vartheta^0(t_n + t), \end{aligned}$$

and the iteration counting process  $m(t)$  is defined by:

$$m(t) = \max \{n : t_n \leq t\}. \quad (3.13)$$

Refer to Figure 2.6 for an illustration of  $\vartheta^0(\cdot)$ .

**Theorem 3.2** Let  $G: \mathbb{R}^d \rightarrow \mathbb{R}^d$  be a Lipschitz continuous vector field. Suppose that the feedback sequence  $\{Y_n\}$  satisfies the Martingale model, so  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}] = G(\theta_n) + \beta_n$ . Consider the stochastic approximation (3.4)

$$\theta_{n+1} = \theta_n + \epsilon_n Y_n.$$

Assume (a1), (a2), (a3) of Proposition 3.1, and that:

(a4) The ODE

$$\frac{d\vartheta(t)}{dt} = G(\vartheta(t)) \quad (3.14)$$

has bounded trajectories and a unique limit for each initial condition. Let  $S$  denote the set of stable points of this ODE and assume that  $S \neq \emptyset$ .

### 3.3. EXOGENOUS NOISE MODEL, DECREASING STEPSIZE

---

Then  $\{\theta_n\}$  converges a.s. and all accumulation points are in  $S$ . In particular, if  $\theta^*$  is the only asymptotically stable point of the ODE, then  $\theta_n \rightarrow \theta^*$  a.s.

**Proof:** This proof parallels the method of proof of Theorem 2.3 and is done in three parts.

*Telescopic sum and integral representation:* We define the shifted interpolated processes  $\vartheta^n(t)$ ,  $M^n(t)$ ,  $B^n(t)$ :

$$\begin{aligned}\vartheta^n(t) &= \theta_n + \sum_{i=n}^{m(t_n+t)-1} \epsilon_i Y_i \\ M^n(t) &= M_n + \sum_{i=n}^{m(t_n+t)-1} \epsilon_i \delta M_i \\ B^n(t) &= B_n + \sum_{i=n}^{m(t_n+t)-1} \epsilon_i \beta_i\end{aligned}$$

Following (3.4) we have

$$\vartheta^n(t) = \theta_n + \sum_{i=n}^{m(t_n+t)-1} \epsilon_i G(\theta_i) + M^n(t) + B^n(t),$$

where  $M^n(t) \rightarrow 0$ ,  $B^n(t) \rightarrow 0$  a.s., which follows directly from the proof of Proposition 3.1. Because  $\vartheta^n(\cdot)$  is piecewise constant, then the sum of the terms  $\epsilon_i G(\theta_i)$  is an integral and

$$\vartheta^n(t) - \vartheta^n(0) = \int_0^t G(\vartheta^n(u)) du + M^n(t) + B^n(t) + \rho^n(t),$$

where  $\rho^n(t)$  accounts for “end point” errors in the integral approximation and  $\rho^n(t) \rightarrow 0$  a.s. uniformly in  $t$  as  $n \rightarrow \infty$  (see Exercise 3.5).

*Characterising the limit ODE:* From the a.s. convergence of  $M^n(\cdot)$ ,  $B^n(\cdot)$  and  $\rho^n(\cdot)$  to zero, there exists a null set  $N$  such that for all  $\omega \notin N$  these sequences converge to zero as  $n \rightarrow \infty$ , uniformly in  $t$ . Pick  $\omega \notin N$ . Given fixed  $\omega$ , the rest of the proof is for a deterministic sequence and we can apply Theorem 2.6 to establish that there is a subsequence of functions  $\{\vartheta^{n_k}(\cdot, \omega)\}$  with a continuous limit and call  $\vartheta(t, \omega)$  the particular limit (we now make it explicit that  $\omega$  is a fixed trajectory). Then  $\vartheta(t, \omega)$  satisfies:

$$\vartheta(t, \omega) = \vartheta(0, \omega) + \int_0^t G(\vartheta(u, \omega)) du,$$

which means that each convergent subsequence (in  $n$ ) has a limit that satisfies the ODE (4.4). By Assumption (a4), however, there is a unique solution for each initial condition so all subsequences share the same limit.

*Asymptotic behaviour:* The final part of the proof uses stability of the limiting ODE to characterise the limit behaviour (in  $t$ ) of the algorithm. Indeed, as  $t \rightarrow \infty$ , for each  $\omega \notin N$ ,

$$\lim_{t \rightarrow \infty} \vartheta(t, \omega) = \lim_{t \rightarrow \infty} \left( \lim_{n \rightarrow \infty} \vartheta^n(t, \omega) \right) = \lim_{n \rightarrow \infty} \theta_n(\omega) \in S,$$

which follows because  $S$  is the (invariant) set of stable points of the ODE. Therefore all limits of the algorithm will be asymptotically stable points. If, in particular,  $\theta^* \in S$  is unique, then  $\theta_n \rightarrow \theta^*$  a.s.

QED

REMARK. Condition (a3) in Proposition 3.1 is somewhat elusive. In the following we provide an argument called *variance control scheme* that allows sometimes to enforce (a3), albeit changing the actual algorithm. Let  $\{Y_n(i) : 1 \leq i \leq k\}$  be collection of iid copies of  $Y_n$ , and denote by

$$\bar{Y}_n^k = \frac{1}{k} \sum_{i=1}^k Y_n(i)$$

the sample average of  $\bar{Y}_n$ . Note that  $\mathbb{E}[\bar{Y}_n^k | \mathfrak{F}_{n-1}] = \mathbb{E}[Y_n | \mathfrak{F}_{n-1}]$  and

$$\frac{1}{\sqrt{k}} V_n = \frac{1}{\sqrt{k}} \mathbb{E}[(Y_n - \mathbb{E}[Y_n | \mathfrak{F}_{n-1}])^2] = V_n^k = \mathbb{E}[(\bar{Y}_n^k - \mathbb{E}[\bar{Y}_n^k | \mathfrak{F}_{n-1}])^2].$$

Hence, if we replace  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}] = G(\theta_n) + \beta_n$  by  $\mathbb{E}[\bar{Y}_n^k | \mathfrak{F}_{n-1}] = G(\theta_n) + \beta_n$ , we reduce the variance by a factor  $1/\sqrt{k}$ . Therefore, by increasing the sample size from 1 to  $k$ , the variance can be reduced. This type of argument may be useful theoretically, but it is not practical for programming, because it forces the values of  $\theta_n$  to remain in undesirable areas of potential numerical instabilities for a long time while generating or evaluating large samples with controlled variance.

**EXAMPLE 3.5.** We revisit Example 3.4 to illustrate how Theorem 3.2 can be applied to show that the procedure converges. First we must impose some conditions on the problem itself so we have a well posed optimization problem and a coercive target vector field. Assume that  $S \subset \mathbb{R}$  is an interval that contains more than one point. Indeed, were  $S$  to contain only one point, say  $x_0$ , then  $\mathbb{E}[Z(x_0)] = y_0$  would be the only corresponding output value, and an infinite number of straight lines could be drawn to pass over this point. In addition, and the slopes of the lines would not be bounded. At least two different points are required to uniquely define a straight line. Assume also that the observations  $X$  have a continuous distribution over  $S$ . Finally, assume that consecutive observations of the streaming data  $\{(x_n, \xi_n)\}$  are iid. and have finite variance. This implies that there is a unique point  $\theta^* \in \mathbb{R}^2$  with  $\|\theta\| < \infty$  that minimizes  $J(\theta)$ . Following the results in Example 2.6, because the field  $-\nabla J$  is coercive, then for any bounded initial set  $\Theta_0$ , there is a box  $H(\Theta_0) = [-M(\Theta_0), M(\Theta_0)]^d$  such that the solution to the ODE

$$\frac{dx}{dt} = -\nabla J(x(t)); \quad x(0) \in \Theta_0 \tag{3.15}$$

has bounded trajectories inside the box  $H(\Theta_0)$ .

Now consider the stochastic approximation of Example 3.4:

$$\begin{aligned} \theta_{n+1,1} &= \theta_{n,1} + \epsilon_n(\xi_n - \theta_{n,1} - \theta_{n,2}x_n), \\ \theta_{n+1,2} &= \theta_{n,2} + \epsilon_n x_n(\xi_n - \theta_{n,1} - \theta_{n,2}x_n), \end{aligned}$$

for which  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}] = -\nabla J(\theta_n)$ . Because we assume that samples are iid, then  $\text{Var}(x_n)$  and  $\text{Var}(\xi_n)$  are (finite) constants and  $V_n$  is a second degree polynomial in  $\theta_1, \theta_2$ . Assumptions (a1), (a2) and (a4) of Theorem 3.2 follow if we choose  $\epsilon_n = \mathcal{O}(n^{-1})$  for example. The elusive Assumption (a3) is much

### 3.3. EXOGENOUS NOISE MODEL, DECREASING STEPSIZE

---

harder to show, because  $V_n = \mathcal{O}(\|\theta_n\|^2)$  is not necessarily uniformly bounded, unless we can show that  $\theta_n$  remains in a compact set almost surely, or other similar conditions.

Let  $\Theta_0$  be a bounded set and use initial points  $\theta_0 \in \Theta_0$  a.s. The following method of proof uses a standard technique: let  $M >> M(\Theta_0)$  and let  $\tilde{\theta}_n$  be the truncated algorithm that follows the recursion  $\tilde{\theta}_{n+1} = \Pi_{[-M,M]^d}(\tilde{\theta}_n + \epsilon_n Y_n)$ , so that  $\tilde{\theta}_n \in [-M, M]^d$  a.s. Then (a3) is satisfied for the truncated algorithm. Applying Theorem 3.2 to the truncated sequence  $\tilde{\theta}_n$  implies that the interpolated truncated process converges to the solution of (3.15). This limit is independent of the truncation, so we can make  $M \rightarrow \infty$  to conclude that the original algorithm also converges to (3.15).

\*\*\*

**EXAMPLE 3.6.** In [20], the Robbins-Monro procedure was modified to approximate the solution to

$$\min_{\theta \in \mathbb{R}} J(\theta),$$

when (independent) experiments can be used to estimate the value of the cost function  $J$  at any design point  $\theta$ , but the function may not be differentiable, or an estimate of its gradient may not be available. Because only the cost function can be estimated, and not its gradient, a finite difference approach leads to the scheme:

$$Y_n(\theta_n) = \frac{\xi_n(\theta_n + c_n) - \xi_n(\theta_n - c_n)}{2c_n},$$

where  $\xi_n(\theta)$  is the  $n$ -th observation of the cost at design parameter value  $\theta$ . It is assumed that these observations are unbiased, that is:  $\mathbb{E}[\xi_n(\theta_n) | \mathfrak{F}_{n-1}] = J(\theta_n)$ , and we assume that the observations of the “experiments” at  $\theta_n + c_n$  and  $\theta_n - c_n$  are statistically independent, that is,  $\xi(\theta_n + c_n)$  is independent of  $\xi(\theta_n - c_n)$ . This procedure is now known as the “Kiefer-Wolfowitz” procedure, and we sometimes will refer to it as the K-W method for short. Recall from Example 1.2 in Chapter 1 that approximations via finite differences can lead to the correct limit, provided that the finite difference scheme uses  $c_n = n^{-c}$  for a positive constant  $c$ .

When  $J(\cdot) \in \mathcal{C}^3$  admits a Taylor approximation, the bias term  $\beta_n$  (see (a3) in Proposition 3.1) can be calculated:

$$\begin{aligned} \mathbb{E}[Y_n | \mathfrak{F}_{n-1}] &= \frac{J(\theta_n + c_n) - J(\theta_n)}{2c_n} - \frac{J(\theta_n - c_n) - J(\theta_n)}{2c_n} \\ &= \frac{1}{2c_n} \left( J'(\theta_n)c_n + \frac{1}{2}J''(\theta_n)c_n^2 + \frac{1}{6}J'''(x)c_n^3 \right. \\ &\quad \left. - J'(\theta_n)(-c_n) - \frac{1}{2}J''(\theta_n)c_n^2 - \frac{1}{6}J'''(y)(-c_n^3) \right) \\ &= J'(\theta_n) + \mathcal{O}(c_n^2), \end{aligned}$$

so that  $\beta_n = \mathbb{E}[Y_n | \mathfrak{F}_{n-1}] - J'(\theta_n) = \mathcal{O}(c_n^2)$ . The term  $V_n$  is given by

$$V_n = \frac{v(\theta_n + c_n) + v(\theta_n - c_n)}{4c_n^2},$$

where  $v(\theta)$  is the variance of the observations at parameter value  $\theta$ . If the variance of the observations is independent of  $\theta$  or uniformly bounded, then  $v(\cdot) \leq V < \infty$  and  $V_n = \mathcal{O}(c_n^{-2})$ . In order for the assumptions (a2) and (a3) to hold, we need to choose the stepsize sequence and the finite difference

sequence  $\{c_n\}$  simultaneously. First, because of (a2) and (a3), necessarily  $c_n \rightarrow 0$ , which immediately implies that  $V_n \rightarrow \infty$ . For instance, if  $\epsilon = n^{-1}$  and  $c_n = n^{-c}$ , then  $\beta_n = \mathcal{O}(n^{-2c})$  and  $V_n = \mathcal{O}(n^{+2c})$  (the variance blows up), so that we need convergence of the series:

$$\sum_n \frac{1}{n^{2c-2}} < \infty \implies 2c - 2 > 1,$$

so that  $c > 3/2$  for the algorithm to converge almost surely to the correct value. Later in Chapter 5 we will see how to tune the various parameters in order to achieve the best convergence rates.

\*\*\*

The method of proof put forward in Example 3.6 above can be further generalised to other schemes that do not necessarily update using the gradient direction. As mentioned in Section 2.3, iterative numerical methods can be used in general to find the zeroes of functions, which can approximate the solution to optimization problems, or can be used to approximate equilibrium solutions, or to approximate fixed points of mappings.

**EXAMPLE 3.7.** Consider a continuously differentiable contraction mapping  $T: \mathbb{R}^d \rightarrow \mathbb{R}^d$  with unique fixed point  $\theta^*$ . Suppose that, given a value of  $\theta$ , a noisy observation or simulation can be performed to produce a random variable  $Z(\theta)$  such that  $\mathbb{E}[Z(\theta)] = T(\theta)$  is Lipschitz continuous, and assume that  $\sup_\theta \text{Var}[Z(\theta)] < \infty$ . Consider the stochastic approximation procedure:

$$\theta_{n+1} = \theta_n + \epsilon_n(Z(\theta_n) - \theta_n), \quad (3.16)$$

where  $Z(\theta_n)$  is independent of previous observations  $Z(\theta_0), \dots, Z(\theta_{n-1})$ . Then this procedure fits the exogenous noise model, where  $Y_n = Z(\theta_n) - \theta_n$  and  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}] = T(\theta_n) - \theta_n$ , and the usual conditions on  $\epsilon_n$ . We will use Theorem 2.5 to show that the ODE (3.16)

$$\frac{dx(t)}{dt} = T(x(t)) - x(t),$$

is coercive for this problem. By assumption, there is only one stationary point of this ODE. We now show that it is asymptotically stable, which in turn establishes that  $\theta_n \rightarrow \theta^*$  a.s.

To show stability, using the method of Section 2.2, it suffices to show that  $\mathbb{A} = \nabla T(\theta^*)^\top - I$  is a Hurwitz matrix. Call  $\mathbb{B} = \nabla T(\theta^*)^\top$ . Then every eigenvalue of  $\mathbb{A}$  has the form  $\rho - 1$ , where  $\rho$  is an eigenvalue of  $\mathbb{B}$ . We now show that if  $\rho$  is any eigenvalue of  $\mathbb{B}$  then its real part (denoted by  $\Re(\rho)$ ) satisfies  $\Re(\rho) < 1$ , which will complete the proof.

Let  $\rho$  be an eigenvalue of  $\mathbb{B}$  with corresponding eigenvector  $v$ , and use a Taylor approximation of  $T(\theta)$  for  $\theta = \theta^* + \delta v$ ,  $\delta \approx 0$  to establish that:

$$\frac{T(\theta) - T(\theta^*)}{\|\theta - \theta^*\|} = \frac{\delta \mathbb{B} v}{\delta \|v\|} + \mathcal{O}(\delta) = \rho \frac{v}{\|v\|} + \mathcal{O}(\delta).$$

Because  $T$  is a contraction mapping, then there is a number  $0 < \tau < 1$  independent of  $\delta$  such that

$$\frac{\|T(\theta) - T(\theta^*)\|}{\|\theta - \theta^*\|} < \tau < 1.$$

Choosing a small enough  $\delta$ , this implies that  $|\rho| < 1$ , which proves the claim.

\*\*\*

### 3.4 Summary for the Exogenous Noise Case

In this section we wrap up the results of this chapter. Consider the optimization problem

$$\min_{\theta \in \Theta} J(\theta), \quad (3.17)$$

Furthermore, for  $\theta_0$  given, let

$$\theta_{n+1} = \theta_n + \epsilon_n Y_n, \quad (3.18)$$

where  $Y_n$  can be decomposed according to the Martingal noise model (3.19) as follows:

$$Y_n = G(\theta_n) + \delta M_n + \beta_n, \quad (3.19)$$

with definitions as in (3.19).

We group the conditions according to their domain. The first set of conditions requires the actual optimization to be well-posed. The second set requires the driving vector field to be appropriate for the optimization problem. Finally, we present the set of conditions that impose convergence of the iterates  $\theta_n$  towards local minima of the original optimization problem. For ease of reference we list all basic definitions

- **The Problem:** the problem is well posed, that is,

**P1:**  $J \in \mathcal{C}^1$

**P2:** in case  $\Theta \neq \mathbb{R}^d$  it holds that  $\Theta = \{\theta \in \mathbb{R}^d : g(\theta) \leq 0\} \subset \mathbb{R}$ , with  $g \in \mathcal{C}^1$

**P3:** the problem has a solution, which is characterized by the KKT conditions, and all KKT points are in a compact set

- **The Target Vector Field:**  $G$  is coercive for the optimization problem, that is,

**VF1:**  $G$  is Lipschitz continuous

**VF2:** the KKT points of (3.17) are the only asymptotically stable points of the ODE

$$\frac{x(t)}{dt} = G(x(t)), \quad (3.20)$$

**VF3:** and for any initial point  $x_0$ , the vector field  $G$  is bounded along  $\{x(t) : t \geq 0\}$ .

- **The Algorithm:**

**A1:** the variance term (3.7) satisfies:  $\sum_{i=0}^{\infty} \epsilon_i^2 V_i < \infty$  a.s.,

**A2D1:**  $\sum_{i=0}^{\infty} \epsilon_i = \infty$  and  $\epsilon_n \rightarrow 0$ ,

**A2D2:** the error terms satisfy  $\sum_{i=0}^{\infty} \epsilon_i \|\beta_i\| < \infty$  a.s.,

**A2F:** the error terms satisfy  $\sum_{i=0}^{\infty} \|\beta_i\| < \infty$  a.s.,

Then, it holds

- provided that **P1** to **P3**, **VF1** to **VF3**, and **A1**, **A2D1** and **A2D2** hold,  $\{\theta_n\}$  converges a.s. in case of a decreasing step size and all accumulation points of  $\{\theta_n\}$  are local minima of the optimization problem;

- provided that **P1** to **P3**, **VF1** to **VF3**, and **A1** and **A2F** hold, we have for fixed  $\epsilon$ , i.e.,  $\epsilon_i = \epsilon$  that the accumulation points  $\theta^\epsilon$  of the sequence  $\theta_n$  obtained from (3.18) approximate local minima of the optimization problem. Letting  $\epsilon$  tend to zero, the accumulation points  $\theta^\epsilon$  provably converge to local minima of the optimization problem.

In case of  $G$  being the negative gradient and  $Y_n$  an unbiased estimator for  $G(\theta_n)$ , the bias term  $\beta_n$  is zero. The negative gradient is coercive for a well-posed problem and one only has to make sure that the variance condition is met. Elaborating on the variance reduction scheme, averaging iid samples of  $Y_n$  to  $\bar{Y}_n^k$ , the variance condition **A1** becomes an issue of efficient implementation of the algorithm rather than a mathematical challenge. Put differently, by increasing the batch size  $k$ ,  $\bar{Y}_n^k$  gets closer to the target field and the "target vector field takes over." Coerciveness of the target field then guarantees convergences of  $\theta_n$  towards a local minimum. With this being said, unbiasedness of  $Y_n$  becomes a desirable property and in Part II we will explain how unbiased gradient estimators can be established for a wide class of problems which are of relevance in practice.

Suppose that we want to use (3.4) to minimise a function  $J(\theta)$  using  $G(\cdot) = -\nabla_\theta J(\cdot)^\top$ , but that there are several local minima. Then the ODE method can be applied to establish that for each trajectory, the iterations converge to local minima, which may depend both on the initial conditions as well as on the trajectory  $\omega$ . Here, the probability that the limit point is actually  $\theta^* = \arg \min J(\theta)$  can be bounded in terms of the probability that the algorithm remains within the domain of attraction of the global minimum  $\theta^*$ .

### 3.5 Exercises

**EXERCISE 3.1.** Use RM procedure in Theorem 3.1 to adjust the power control variables of Example 3.2 with stepsize condition (3.2) so as to attain the target temperature. Pose the model, specify which assumptions you must make about  $M$  and  $w_n$  for convergence. Does your result imply that  $\theta_n \rightarrow \theta^*$  almost surely?

**EXERCISE 3.2.** Show that the process  $M_n \stackrel{\text{def}}{=} \sum_{i=0}^n \epsilon_i \delta M_i$  is a martingale process on  $(\Omega, \mathbb{P}, \{\mathfrak{F}_n\})$ ; for the basic definition and properties of martingale processes we refer to the Appendix. Show that  $\mathbb{E}[\delta M_n \delta M_m] = 0$ .

**EXERCISE 3.3.** Refer to the model in Example 3.4. Show that for a random variable  $X$  with finite variance,

$$\nabla J(\theta) = \begin{pmatrix} -\mathbb{E}[Z(X) - \theta_1 - \theta_2 X] \\ -\mathbb{E}[X Z(X) - \theta_1 X - \theta_2 X^2] \end{pmatrix}.$$

Use this to show the claim that  $\mathbb{E}[Y_n | \mathfrak{F}_{n-1}] = -\nabla J(\theta_n)$ .

**EXERCISE 3.4.** Explain why the continuity theorem for decreasing sets implies that as  $n \rightarrow \infty, m > n$ ,  $|M_m - M_n| \rightarrow 0$  with probability 1.

**EXERCISE 3.5.** Refer to the proof of Theorem 3.2. Assume that  $G$  is bounded and show that  $\rho^n(\cdot)$  converges a.s., uniformly in  $t$ , and that for each  $\omega \notin N$ , the sequence  $\{\vartheta^n(\cdot)\}$  is equicontinuous.

**EXERCISE 3.6.** Consider the following well known Nash equilibrium problem in transportation. The transportation time along path  $i$  is called  $T_i(\theta)$  and it depends on the vector of traffic flow  $\theta \in \mathbb{R}^3$ , where  $\theta_i$  is the total traffic (total number of cars) on path  $i$ . In Figure 3.4 there are three possible paths,  $\{(ABD), (ABCD), (ACD)\}$  and the total number of cars is  $N = 100$ .

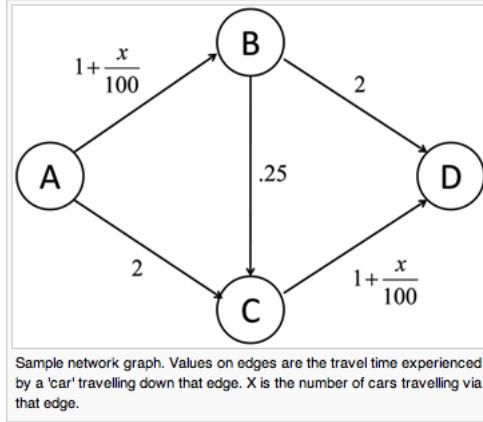


Figure 3.4: Traffic flow equilibrium

A Nash equilibrium will be an allocation such that if a driver on path  $i$  decides to change his/her path, then he/she will experience a larger delay than staying at equilibrium. Here,

$$T(\theta) = \begin{pmatrix} 3 + \frac{\theta_1 + \theta_2}{N} \\ 2.25 + \frac{\theta_1 + 2\theta_2 + \theta_3}{N} \\ 3 + \frac{\theta_2 + \theta_3}{N} \end{pmatrix}$$

- (a) Show that there is a unique Nash equilibrium, by showing that there is a unique value  $\theta^*$  such that  $T_i(\theta^*) = \text{constant}$ , is independent of  $i$ .
- (b) In order to attain equilibrium, the flow on path  $i$  should aim to "equalize" the delay. Suppose now that we do not know the various constants in the delay function, but can only just estimate the delays by running a simulation, which yields an unbiased estimator  $\widehat{T}(\theta)$ . Consider using the following algorithm to approximate the solution:

$$\theta_{n+1,i} = \theta_{n,i} - \epsilon_n \left( \widehat{T}_i(\theta_n) - \frac{1}{3} \sum_k \widehat{T}_k(\theta_n) \right); \quad i = 1, 2, 3. \quad (3.21)$$

Show first that the algorithm is mass preserving, that is,  $\sum_i \theta_{0,i} = N$  then  $\sum_i \theta_{n,i} = N$ .

- (c) Characterize the behavior of the stochastic approximation (3.21) and specify your assumptions.

By the way, this example is classic to show that the Nash equilibrium does not minimise overall travel time. For the specific model,  $\theta^* = (25, 50, 25)^T$ , and  $T_i(\theta^*) = 3.75$  for all  $i$ . However, for  $\theta = (50, 0, 50)$ , then  $T_1(\theta) = T_3(\theta) = 3.5$ .

