Eidgenössische Technische Hochschule Zürich
Swiss Federal Institute of Technology Zurich

Seminar for Applied Mathematics SAM

Weak error analysis for stochastic gradient descent optimization algorithms

A. Bercher and L. Gonon and A. Jentzen and D. Salimova

Research Report No. 2020-50
July 2020

Seminar für Angewandte Mathematik
Eidgenössische Technische Hochschule
CH-8092 Zürich
Switzerland

Weak error analysis for stochastic gradient descent optimization algorithms

Aritz Bercher[1], Lukas Gonon[2,3],
Arnulf Jentzen[4,5], and Diyora Salimova[6,7]

[1] Department of Mathematics, ETH Zurich,
Switzerland, e-mail: abercher@outlook.com

[2] Faculty of Mathematics and Statistics, University of
St. Gallen, Switzerland, e-mail: lukas.gonon@unisg.ch

[3] Department of Mathematics, ETH Zurich,
Switzerland, e-mail: lukas.gonon@math.ethz.ch

[4] Department of Mathematics, ETH Zurich,
Switzerland, e-mail: arnulf.jentzen@sam.math.ethz.ch

[5] Faculty of Mathematics and Computer Science, University of
Münster, Germany, e-mail: ajentzen@uni-muenster.de

[6] Department of Mathematics, ETH Zurich,
Switzerland, e-mail: diyora.salimova@sam.math.ethz.ch

[7] Department of Information Technology and Electrical Engineering,
ETH Zurich, Switzerland, e-mail: sdiyora@mins.ee.ethz.ch

July 31, 2020

Abstract

Stochastic gradient descent (SGD) type optimization schemes are fundamental ingredients in a large number of machine learning based algorithms. In particular, SGD type optimization schemes are frequently employed in applications involving natural language processing, object and face recognition, fraud detection, computational advertisement, and numerical approximations of partial differential equations. In mathematical convergence results for SGD type optimization schemes there are usually two types of error criteria studied in the scientific literature, that is, the error in the strong sense and the error with respect to the objective function. In applications one is often not only interested in the size of the error with respect to the objective function but also in the size of the error with respect to a test function which is possibly different from the objective function. The analysis of the size of this error is the subject of this article. In particular, the main result of this article proves under suitable assumptions that the size of this error decays at the same speed as in the special case where the test function coincides with the objective function.

Contents

1 Introduction **7**

2 Existence results for solutions of first-order Kolmogorov backward partial differential equations (PDEs) **11**

2.1 Gronwall-type inequalities 12

2.2 Sufficient conditions for interchanging derivatives and integrals 14

2.3 Existence, uniqueness, and regularity results for solutions of ordinary differential equations (ODEs) 23

2.4 Existence results for solutions of first-order Kolmogorov backward PDEs 26

3 Weak error estimates for stochastic approximation algorithms (SAAs) in the case of general learning rates **29**

3.1 Mathematical description for SAAs in the case of general learning rates 30

3.2 Sufficient conditions for interchanging derivatives and expectations . . 30

3.3 Spatial regularity results for flows of deterministic ODEs 31

3.4 Upper bounds for second-order spatial derivatives of certain deterministic flows 38

3.5 Temporal regularity results for SAAs in the case of general learning rates 46

3.6 A priori estimates for SAAs in the case of general learning rates . . . 46

3.7 Weak error estimates for SAAs in the case of general learning rates with mini-batches 51

3.8 Upper bounds for integrals of certain exponentially decaying functions 65

3.9 Weak error estimates for SAAs in the case of polynomially decaying learning rates with mini-batches 72

4 Weak error estimates for SAAs in the case of polynomially decaying learning rates **75**

4.1 Mathematical description for SAAs in the case of polynomially decaying learning rates . . . 76
4.2 On a sequence of uniformly bounded functions . . . 76
4.3 A priori estimates for SAAs in the case of polynomially decaying learning rates . . . 79
4.4 A posteriori estimates for conditional variances associated to SAAs . 85
4.5 A priori estimates for suitable approximation error constants associated to SAAs . . . 87
4.6 Weak convergence rates for SAAs in the case of polynomially decaying learning rates with mini-batches . . . 95
4.7 Weak convergence rates for SAAs in the case of polynomially decaying learning rates without mini-batches . . . 98
4.8 SAAs for random rotation problems . . . 99

5 Weak error estimates for stochastic gradient descent (SGD) optimization methods 103

5.1 Weak error estimates for SGD optimization methods in the case of coercive objective functions . . . 103
5.2 Weak error estimates for SGD optimization methods in the case of objective functions with linearly growing derivatives . . . 108

Chapter 1

Introduction

Stochastic gradient descent (SGD) type optimization schemes are fundamental ingredients in a large number of machine learning based algorithms. In particular, SGD type optimization schemes are frequently employed in applications involving natural language processing (cf., e.g., [27,44,47,49,57,106]), object and face recognition (cf., e.g., [50, 62, 95, 100, 104]), fraud detection (cf., e.g., [24, 88]), computational advertisement (cf., e.g., [103, 110]), price formation (cf., e.g., [96]), portfolio hedging (cf., e.g., [21]), financial model calibration (cf., e.g., [6, 71]), and numerical approximations of partial differential equations (PDEs) (cf., e.g., [8, 9, 39, 40, 45, 46, 74, 77, 97]). In view of the success of the SGD type optimization schemes in the above sketched applications, SGD type optimization schemes have also been intensively studied in the scientific literature. In particular, we refer, e.g., to [14,18,89] for overview articles on SGD type optimization schemes, we refer, e.g., to [13,15,29,30,32,36,37,58,64,65, 69,72,73,79,82,83,84,92,93,94,98,108,109,111] and the references mentioned therein for the proposal and the derivation of SGD type optimization schemes, we refer, e.g., to [4, 5, 16, 19, 28, 35, 51, 56, 67, 75, 78, 80, 81, 85, 86, 99, 107, 112] and the references mentioned therein for numerical simulations for SGD type optimization schemes, and we refer, e.g., to [7, 10, 11, 12, 26, 31, 33, 43, 44, 47, 48, 52, 62, 66, 90, 91, 107, 112] and the references mentioned therein for applications involving neural networks and SGD type optimization schemes. There are also a number of rigorous mathematical results on SGD type optimization schemes which aim to contribute to an understanding toward the success and the limitations of SGD type optimization schemes (cf., e.g., [35, 53, 55, 56, 63, 78, 81, 85, 101] for mathematical results in case of strongly convex objective functions, cf., e.g., [3, 4, 5, 17, 105] for mathematical results in case of convex but possibly non-strongly convex objective functions, and cf., e.g., [2, 20, 22, 23, 41, 42, 68, 70, 72] for mathematical results in case of possibly non-convex objective functions). In mathematical convergence results for SGD type optimization schemes there are usually two types of error criteria studied in the scientific literature, that is, (I) the error in the strong sense (cf., e.g., [3, 4, 19, 35, 53, 55, 78, 78]) and (II) the error with respect to the objective function (cf., e.g., [3, 4, 5, 35, 55, 56, 63, 81, 85, 101, 105]). More specifically, suppose that the objective function $f\colon \mathbb{R}^d \to \mathbb{R}$ which we intend to minimize by means of an SGD type optimization scheme satisfies for all $x \in \mathbb{R}^d$ that $f(x) = \mathbb{E}[F(x, Z)]$, where $d \in \mathbb{N} = \{1, 2, 3, \ldots\}$, where $Z\colon \Omega \to S$ is a random variable on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values in a measurable space $(S, \mathcal{S})$, and where $F\colon \mathbb{R}^d \times S \to \mathbb{R}$ is a sufficiently regular function (cf., e.g., [34, Section 1], [53, Theorem 1.1], and [55, Theorem 1.1]). Moreover, suppose that $\Xi \in \mathbb{R}^d$ is a minimum point of the objective function $f\colon \mathbb{R}^d \to \mathbb{R}$ and suppose that $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ is the stochastic process induced by the considered SGD type optimization scheme (cf. (1.5) in Theorem 1.0.1 below). Then in the case of (I) one is interested in the size of the strong L^2-error between the minimum point Ξ and Θ_n as $n \to \infty$ and in the case of (II) one is interested in the size of the error between the objective function f evaluated at the minimum point Ξ and the expectation of the objective function f evaluated at Θ_n as $n \to \infty$. In the case of (II) the error is in some sense weaker but in many situations one can establish quicker convergence rates for (II), namely, twice the convergence rate in (I) (see, e.g., [55, items (ii) and (iii) in Theorem 1.1]). In applications one is usually not only interested in the objective function f evaluated at the minimum point Ξ but also in some other functional evaluated at the minimum point Ξ and the analysis of the error corresponding to this approximation problem is the subject of this article. More formally, the main contribution of this work is to study an error criteria which is different from (I) and (II) and which essentially generalizes (II), that is, in this work we study the size of the error between $\psi(\Xi)$ and $\mathbb{E}[\psi(\Theta_n)]$ as $n \to \infty$ for any sufficiently regular function $\psi\colon \mathbb{R}^d \to \mathbb{R}$ (in particular, including the objective function $f\colon \mathbb{R}^d \to \mathbb{R}$ as a special case). More specifically, the main result of this article, Theorem 4.6.2 below, establishes that under suitable convexity type assumptions the convergence rate of this error is the same convergence rate as in the special case (II) where the sufficiently regular function $\psi\colon \mathbb{R}^d \to \mathbb{R}$ coincides with the objective function $f\colon \mathbb{R}^d \to \mathbb{R}$. To illustrate the findings of Theorem 4.6.2 we now present a special case of the main result of this article.

Theorem 1.0.1. *Let $d \in \mathbb{N}$, $\xi, \Xi \in \mathbb{R}^d$, $\varepsilon \in (0,1)$, $\eta, L, c \in (0,\infty)$, $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$, let $(S, \mathcal{S})$ be a measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $F = (F(\theta, s))_{(\theta,s)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}$ be $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{S})/\mathcal{B}(\mathbb{R})$-measurable, let $Z_n\colon \Omega \to S$, $n \in \mathbb{N}$, be i.i.d. random variables, assume for all $s \in S$ that $(\mathbb{R}^d \ni \theta \mapsto F(\theta, s) \in$*

$\mathbb{R}) \in C^3(\mathbb{R}^d, \mathbb{R})$, assume for all $\theta, \vartheta \in \mathbb{R}^d$ that

$$\mathbb{E}\big[\|(\nabla_\theta F)(\theta, Z_1)\|_{\mathbb{R}^d}^2\big] \leq c\big[1 + \|\theta\|_{\mathbb{R}^d}\big]^2, \tag{1.1}$$

$$\textstyle\sum_{i=2}^3 \inf_{\delta\in(0,\infty)} \sup_{u\in[-\delta,\delta]^d} \mathbb{E}\big[|F(\theta, Z_1)| + \|(\frac{\partial^i}{\partial\theta^i} F)(\theta + u, Z_1)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})}^{1+\delta}\big] < \infty, \tag{1.2}$$

$$\langle \theta - \vartheta, \mathbb{E}[(\nabla_\theta F)(\theta, Z_1)] - \mathbb{E}[(\nabla_\theta F)(\vartheta, Z_1)]\rangle_{\mathbb{R}^d} \geq L\|\theta - \vartheta\|_{\mathbb{R}^d}^2, \tag{1.3}$$

$$\big\|\mathbb{E}\big[(\tfrac{\partial^3}{\partial\theta^3} F)(\theta, Z_1)\big]\big\|_{L^{(3)}(\mathbb{R}^d,\mathbb{R})} + \max\nolimits_{i\in\{1,2\}} \|\psi^{(i)}(\theta)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})} < \infty, \tag{1.4}$$

and $\|\mathbb{E}[(\nabla_\theta F)(\theta, Z_1)]\|_{\mathbb{R}^d} \leq c\|\theta - \Xi\|_{\mathbb{R}^d}$, and let $\Theta \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ satisfy for all $n \in \mathbb{N}$ that $\Theta_0 = \xi$ and

$$\Theta_n = \Theta_{n-1} - \tfrac{\eta}{n^{1-(\varepsilon/2)}}(\nabla_\theta F)(\Theta_{n-1}, Z_n). \tag{1.5}$$

Then

(i) we have that $\{\theta \in \mathbb{R}^d \colon (\mathbb{E}[F(\theta, Z_1)] = \inf_{\vartheta\in\mathbb{R}^d} \mathbb{E}[F(\vartheta, Z_1)])\} = \{\Xi\}$ and

(ii) there exists $C \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have that

$$|\psi(\Xi) - \mathbb{E}[\psi(\Theta_n)]| \leq C n^{\varepsilon-1}. \tag{1.6}$$

Theorem 1.0.1 is an immediate consequence of Corollary 5.2.1 below. Corollary 5.2.1, in turn, follows from Theorem 4.6.2 which is the main result of this article. We now introduce some of the notation which we have used in Theorem 1.0.1 above and which we will use in the later part of this article. For every $d \in \mathbb{N}$ we denote by $\|\cdot\|_{\mathbb{R}^d} : \mathbb{R}^d \to [0, \infty)$ the standard norm on $\mathbb{R}^d$, for every $d \in \mathbb{N}$ we denote by $\langle\cdot, \cdot\rangle_{\mathbb{R}^d} \colon \mathbb{R}^d \times \mathbb{R}^d \to \mathbb{R}$ the standard scalar product on $\mathbb{R}^d$, for every $k, m, n \in \mathbb{N}$ we denote by $L^{(k)}(\mathbb{R}^m, \mathbb{R}^n)$ the set of all continuous k-linear functions from $\mathbb{R}^m \times \mathbb{R}^m \times \ldots \times \mathbb{R}^m = (\mathbb{R}^m)^k$ to $\mathbb{R}^n$, for every $k, m, n \in \mathbb{N}$ we denote by $\|\cdot\|_{L^{(k)}(\mathbb{R}^m,\mathbb{R}^n)} : L^{(k)}(\mathbb{R}^m, \mathbb{R}^n) \to [0, \infty)$ the function which satisfies for all $A \in L^{(k)}(\mathbb{R}^m, \mathbb{R}^n)$ that

$$\|A\|_{L^{(k)}(\mathbb{R}^m,\mathbb{R}^n)} = \sup_{u_1,u_2,\ldots,u_k\in\mathbb{R}^m\setminus\{0\}} \frac{\|A(u_1, u_2, \ldots, u_k)\|_{\mathbb{R}^n}}{\|u_1\|_{\mathbb{R}^m}\|u_2\|_{\mathbb{R}^m} \cdots \|u_k\|_{\mathbb{R}^m}}, \tag{1.7}$$

for every $m, n \in \mathbb{N}$ we denote by $L^{(0)}(\mathbb{R}^m, \mathbb{R}^n)$ the set given by $L^{(0)}(\mathbb{R}^m, \mathbb{R}^n) = \mathbb{R}^n$, and for every $m, n \in \mathbb{N}$ we denote by $\|\cdot\|_{L^{(0)}(\mathbb{R}^m,\mathbb{R}^n)} : \mathbb{R}^n \to [0, \infty)$ the function which satisfies for all $x \in \mathbb{R}^n$ that $\|x\|_{L^{(0)}(\mathbb{R}^m,\mathbb{R}^n)} = \|x\|_{\mathbb{R}^n}$. Note that for all $m, n \in \mathbb{N}$, $A \in L(\mathbb{R}^m, \mathbb{R}^n)$ we have that $L(\mathbb{R}^m, \mathbb{R}^n) = L^{(1)}(\mathbb{R}^m, \mathbb{R}^n)$ and $\|A\|_{L(\mathbb{R}^m,\mathbb{R}^n)} = \|A\|_{L^{(1)}(\mathbb{R}^m,\mathbb{R}^n)}$. Let us also add a few further comments on some of the mathematical objects appearing appearing in Theorem 1.0.1 above. In Theorem 1.0.1 above we intend to approximately solve the stochastic optimization problem in item (i) above. More specifically, in Theorem 1.0.1 above we intend to weakly approximate the global minimizer $\Xi \in \mathbb{R}^d$ of the function $\mathbb{R}^d \ni \theta \mapsto \mathbb{E}[F(\theta, Z_1)] \in \mathbb{R}$, where $F\colon \mathbb{R}^d \times S \to \mathbb{R}$ is a sufficiently regular function and where $Z_1\colon \Omega \to S$ is a random variable on the probability space $(\Omega, \mathcal{F}, \mathbb{P})$ with values on the measurable space $(S, \mathcal{S})$. In Theorem 1.0.1 above we intend to accomplish this by means of the stochastic gradient descent process $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ defined recursively in (1.5). In (1.6) in item (ii) in Theorem 1.0.1 above we establish that for every sufficiently regular function $\psi\colon \mathbb{R}^d \to \mathbb{R}$ and every arbitrarily small $\varepsilon \in (0,1)$ we have that the weak error $|\psi(\Xi) - \mathbb{E}[\psi(\Theta_n)]|$ converges with convergence rate $1-\varepsilon$ to 0 as $n \to \infty$. The weak error analysis which we use in our proof of Theorem 1.0.1 above is strongly based on employing first-order Kolmogorov backward PDEs associated to ordinary differential equations (ODEs). In that aspect our strategy of our proof of Theorem 1.0.1 is inspired by the weak error analysis for numerical approximations of stochastic differential equations (SDEs). In particular, the weak error analysis for numerical approximations of SDEs is often based on employing second-order Kolmogorov PDEs associated to SDEs; see, e.g., Kloeden & Platen [60, Chapter 14], Rößler [87, Subsection 2.2.1], Müller-Gronbach & Ritter [76, Section 4] and the references mentioned therein for weak error analyses for numerical approximations of SDEs.

The rest of this article is structured in the following way. As we mentioned earlier, the weak error analysis which we use in our proof of Theorem 1.0.1 above is strongly based on employing first-order Kolmogorov backward PDEs associated to ODEs. To this end, we recall in Chapter 2 existence and regularity properties for solutions of such first-order Kolmogorov backward PDEs. In Chapter 3 we use the analysis for first-order Kolmogorov backward PDEs from Chapter 2 to study weak approximation errors for stochastic approximation algorithms (SAAs) in the case of general learning rates. In Chapter 4 we specialize the weak error analysis for SAAs in the case of general learning rates from Chapter 3 to accomplish weak error estimates for SAAs in the case of polynomially decaying learning rates. In Chapter 5 we apply the weak error analysis results for SAAs from Chapter 4 to establish weak error estimates for SGD optimization methods.

Chapter 2

Existence results for solutions of first-order Kolmogorov backward partial differential equations (PDEs)

The weak error analysis which we use in our proof of Theorem 1.0.1 above is strongly based on employing first-order Kolmogorov backward PDEs associated to ODEs. In this chapter we present in Proposition 2.4.1 in Section 2.4 below an elementary existence result for solutions of such first-order Kolmogorov backward PDEs. In our proof of Proposition 2.4.1 we use the well-known regularity result for solutions of ODEs in Lemma 2.3.2 in Section 2.3 below and we use the elementary uniqueness result for solutions of ODEs in Lemma 2.3.4 in Section 2.3 below. Our proof of Lemma 2.3.4, in turn, employs the well-known result for continuous functions on compact topological spaces in Lemma 2.2.2 in Section 2.2 below and the well-known Gronwall integral inequality in Lemma 2.1.2 in Section 2.1 below. In addition, our proof of Proposition 2.4.1 also uses the essentially well-known result on the possibility of interchanging derivatives and integrals in Lemma 2.2.6 in Section 2.2 below. A slightly modified version of Lemma 2.2.6 can, e.g., be found in Durrett [38, Theorem A.5.1]. In order to formulate the statement of Lemma 2.2.6 we employ the essentially well-known measurability result for derivatives of sufficiently regular functions in Corollary 2.2.5 in Section 2.2 below. Corollary 2.2.5 follows directly from the elementary measurability results in Lemmas 2.2.1, 2.2.3, and 2.2.4 in Section 2.2 below. Moreover, in this chapter we present in Lemma 2.1.1 a well-known Gronwall-type differential inequality, we present in Lemma 2.2.7 a direct generalization of the result on the possibility of interchanging derivatives and integrals in Lemma 2.2.6, we present in Proposition 2.3.1 an essentially well-known existence and uniqueness result for solutions of ODEs, and we present in Corollary 2.3.3 a direct generalization of the regularity result for solutions of ODEs in Lemma 2.3.2. In Chapter 3 below we employ Lemma 2.1.1, Lemma 2.1.2, Lemma 2.2.2, Corollary 2.2.5, Lemma 2.2.6, Lemma 2.2.7, Proposition 2.3.1, Lemma 2.3.2, Corollary 2.3.3, and Proposition 2.4.1 to study weak approximation errors for SAAs.

2.1 Gronwall-type inequalities

Lemma 2.1.1. *Let $t \in \mathbb{R}$, $T \in (t,\infty)$, $b \in C([t,T],\mathbb{R})$, $f \in C^1([t,T],\mathbb{R})$ satisfy for all $s \in [t,T]$ that $f'(s) \leq b(s)f(s)$. Then we have for all $s \in [t,T]$ that*

$$f(s) \leq f(t)\exp\!\left(\int_t^s b(u)\,du\right). \tag{2.1}$$

Proof of Lemma 2.1.1. Throughout this proof let $v\colon [t,T] \to (0,\infty)$ satisfy for all $s \in [t,T]$ that

$$v(s) = \exp\!\left(\int_t^s b(u)\,du\right) \tag{2.2}$$

and let $g\colon [t,T] \to \mathbb{R}$ satisfy for all $s \in [t,T]$ that

$$g(s) = \frac{f(s)}{v(s)}. \tag{2.3}$$

Observe that for all $s \in [t,T]$ we have that $v'(s) = b(s)v(s)$. This implies that for all $s \in [t,T]$ we have that

$$\begin{aligned} g'(s) &= \frac{f'(s)v(s) - f(s)v'(s)}{v(s)^2} \\ &= \frac{f'(s)v(s) - f(s)b(s)v(s)}{v(s)^2} \\ &\leq \frac{b(s)f(s)v(s) - f(s)b(s)v(s)}{v(s)^2} = 0. \end{aligned} \tag{2.4}$$

This assures that g is non-increasing. This reveals that for all $s \in [t,T]$ it holds that

$$\frac{f(s)}{v(s)} = g(s) \leq g(t) = \frac{f(t)}{v(t)} = f(t). \tag{2.5}$$

This establishes (2.1). The proof of Lemma 2.1.1 is thus completed. □

Lemma 2.1.2. *Let $T \in (0,\infty)$, $a, b \in [0,\infty)$, let $f\colon [0,T] \to [0,\infty)$ be $\mathcal{B}([0,T])$/ $\mathcal{B}([0,\infty))$-measurable, and assume for all $t \in [0,T]$ that*

$$\int_0^T |f(s)|\,ds < \infty \qquad \text{and} \qquad f(t) \le a + b\int_0^t f(s)\,ds. \tag{2.6}$$

Then we have for all $t \in [0,T]$ that $f(t) \le a\exp(bt)$.

Proof of Lemma 2.1.2. We claim that for all $n \in \mathbb{N}_0$, $t \in [0,T]$ we have that

$$f(t) \le a\left(\sum_{k=0}^{n} \frac{(bt)^k}{k!}\right) + b^{n+1}\int_0^t \frac{(t-s)^n}{n!} f(s)\,ds. \tag{2.7}$$

We now establish (2.7) by induction on $n \in \mathbb{N}_0$. The base case $n = 0$ is an immediate consequence of (2.6). For the induction step $\mathbb{N}_0 \ni n \to n+1 \in \mathbb{N}_0$ assume that (2.7) holds for a given $n \in \mathbb{N}_0$. Observe that the induction hypothesis and (2.6) ensure that for all $t \in [0,T]$ we have that

$$\begin{aligned} f(t) &\le a\left(\sum_{k=0}^{n} \frac{(bt)^k}{k!}\right) + b^{n+1}\int_0^t \frac{(t-s)^n}{n!} f(s)\,ds \\ &\le a\left(\sum_{k=0}^{n} \frac{(bt)^k}{k!}\right) + b^{n+1}\int_0^t \frac{(t-s)^n}{n!}\left(a + b\int_0^s f(v)\,dv\right)ds. \end{aligned} \tag{2.8}$$

Moreover, note that for all $t \in [0,T]$ we have that

$$\int_0^t \frac{(t-s)^n}{n!}\,ds = \frac{1}{n!}\left[\frac{-(t-s)^{n+1}}{n+1}\right]_{s=0}^{s=t} = \frac{t^{n+1}}{(n+1)!}. \tag{2.9}$$

Furthermore, observe that Tonelli's theorem implies that for all $t \in [0,T]$ we have that

$$\begin{aligned} \int_0^t (t-s)^n \int_0^s f(v)\,dv\,ds &= \int_0^t \int_0^t (t-s)^n f(v) \mathbb{1}_{\{0\le v\le s\le t\}}\,dv\,ds \\ &= \int_0^t f(v)\int_v^t (t-s)^n\,ds\,dv \\ &= \int_0^t f(v)\left[-\frac{(t-s)^{n+1}}{n+1}\right]_{s=v}^{s=t} dv \\ &= \int_0^t f(v)\frac{(t-v)^{n+1}}{n+1}\,dv. \end{aligned} \tag{2.10}$$

Combining this, (2.8), and (2.9) establishes that for all $t \in [0,T]$ we have that

$$\begin{aligned} f(t) &\leq a\left(\sum_{k=0}^{n} \frac{(bt)^k}{k!}\right) + b^{n+1}\int_0^t \frac{(t-s)^n}{n!}\left(a + b\int_0^s f(v)\,dv\right)ds \\ &= a\left(\sum_{k=0}^{n} \frac{(bt)^k}{k!}\right) + a\frac{(bt)^{n+1}}{(n+1)!} + b^{n+2}\int_0^t \frac{(t-s)^{n+1}}{(n+1)!} f(s)\,ds \\ &= a\left(\sum_{k=0}^{n+1} \frac{(bt)^k}{k!}\right) + b^{n+2}\int_0^t \frac{(t-s)^{n+1}}{(n+1)!} f(s)\,ds. \end{aligned} \tag{2.11}$$

This proves (2.7) in the case $n+1$. This finishes the proof of the induction step. Induction hence establishes (2.7). Next observe that (2.7) implies that for all $t \in [0,T]$, $n \in \mathbb{N}_0$ we have that

$$f(t) \leq a\,e^{bt} + b^{n+1}\int_0^t \frac{(t-s)^n}{n!} f(s)\,ds \leq a\,e^{bt} + b^{n+1}\frac{t^n}{n!}\int_0^t f(s)\,ds. \tag{2.12}$$

Moreover, note that (2.6) ensures that for all $t \in [0,T]$ we have that

$$\limsup_{n\to\infty}\left[b^{n+1}\frac{t^n}{n!}\int_0^t f(s)\,ds\right] = 0. \tag{2.13}$$

Combining this and (2.12) establishes that for all $t \in [0,T]$ we have that $f(t) \leq a\exp(bt)$. The proof of Lemma 2.1.2 is thus completed. □

2.2 Sufficient conditions for interchanging derivatives and integrals

Lemma 2.2.1. *Let $(S,\mathcal{S})$ be a measurable space, let (X,d_X) be a compact metric space, let (Y,d_Y) be a separable metric space, let $C(X,Y)$ be the space of continuous functions endowed with the topology of d_Y-uniform convergence, let $f\colon X\times S \to Y$ be a function, assume for all $x \in X$ that $(S \ni s \mapsto f(x,s) \in Y)$ is $\mathcal{S}/\mathcal{B}(Y)$-measurable, and assume for all $s \in S$ that $(X \ni x \mapsto f(x,s) \in Y) \in C(X,Y)$. Then we have that*

$$(S \ni s \mapsto (X \ni x \mapsto f(x,s) \in Y) \in C(X,Y)) \tag{2.14}$$

is $\mathcal{S}/\mathcal{B}(C(X,Y))$-measurable.

Lemma 2.2.2. *Let $(X,\mathcal{X})$ be a compact topological space, let (M,d) be a metric space, and let $u \in M$, $f \in C(X,M)$. Then we have that*

$$\sup\bigl(\{d(f(x),u) \in \mathbb{R}\colon x \in X\} \cup \{0\}\bigr) < \infty. \tag{2.15}$$

Lemma 2.2.3. *Let $(X, \|\cdot\|_X)$ and $(Y, \|\cdot\|_Y)$ be finite dimensional normed vector spaces, let $(S, \mathcal{S})$ be a measurable space, let $F = (F(x,s))_{(x,s)\in X\times S} \colon X \times S \to Y$ be $(\mathcal{B}(X)\otimes\mathcal{S})/\mathcal{B}(Y)$-measurable, and assume for all $s \in S$ that $(X \ni x \mapsto F(x,s) \in Y) \in C^1(X,Y)$. Then we have for all $x \in X$ that*

$$\big(S \ni s \mapsto (\tfrac{\partial}{\partial x}F)(x,s) \in L(X,Y)\big) \tag{2.16}$$

is $\mathcal{S}/\mathcal{B}(L(X,Y))$-measurable.

Proof of Lemma 2.2.3. Throughout this proof let $V = C(\{w \in X \colon \|w\|_X \leq 1\}, Y)$, let $\|\cdot\|_V : V \to [0,\infty)$ satisfy for all $f \in V$ that

$$\|f\|_V = \sup\nolimits_{h\in\{w\in X\colon \|w\|_X\leq 1\}} \|f(h)\|_Y \tag{2.17}$$

(cf. Lemma 2.2.2), let $\iota \colon C(X,Y) \to V$ satisfy for all $\varphi \in C(X,Y)$ that

$$\iota(\varphi) = (\{w \in X \colon \|w\|_X \leq 1\} \ni h \mapsto \varphi(h) \in Y), \tag{2.18}$$

and let $\psi \colon \{f \in V \colon f \text{ is linear}\} \to L(X,Y)$ satisfy for all $\mathcal{A} \in \{f \in V \colon f \text{ is linear}\}$, $h \in X\backslash\{0\}$ that

$$\psi(\mathcal{A})(h) = \|h\|_X \mathcal{A}\Big(\tfrac{h}{\|h\|_X}\Big). \tag{2.19}$$

Observe that the assumption that $\forall\, s \in S \colon (X \ni x \mapsto F(x,s) \in Y) \in C^1(X,Y)$ implies that for all $x \in X$, $s \in S$, $\varepsilon \in (0,\infty)$ there exists $r \in (0,\infty)$ such that for all $h \in X\backslash\{0\}$ with $\|h\|_X \leq r$ we have that

$$\frac{\|F(x+h,s) - F(x,s) - (\frac{\partial}{\partial x}F)(x,s)h\|_Y}{\|h\|_X} \leq \varepsilon. \tag{2.20}$$

This reveals that for all $x \in X$, $s \in S$, $\varepsilon \in (0,\infty)$ there exists $r \in (0,\infty)$ such that

$$\sup_{h\in\{w\in X\colon 0<\|w\|_X\leq r\}} \frac{\|F(x+h,s) - F(x,s) - (\frac{\partial}{\partial x}F)(x,s)h\|_Y}{\|h\|_X} \leq \varepsilon. \tag{2.21}$$

This ensures that for all $x \in X$, $s \in S$, $\varepsilon \in (0,\infty)$ there exists $r \in (0,\infty)$ such that for all $\delta \in (0,r]$ we have that

$$\sup_{h\in\{w\in X\colon 0<\|w\|_X\leq \delta\}} \frac{\|F(x+h,s) - F(x,s) - (\frac{\partial}{\partial x}F)(x,s)h\|_Y}{\|h\|_X} \leq \varepsilon. \tag{2.22}$$

This reveals that for all $x \in X$, $s \in S$ it holds that

$$\limsup_{(0,\infty)\ni r\to 0} \sup_{h\in\{w\in X\colon 0<\|w\|_X\leq r\}} \left[\frac{\|F(x+h,s)-F(x,s)-(\frac{\partial}{\partial x}F)(x,s)h\|_Y}{\|h\|_X}\right] = 0. \quad (2.23)$$

This assures that for all $x \in X$, $s \in S$ we have that

$$\limsup_{(0,\infty)\ni r\to 0} \sup_{h\in\{w\in X\colon 0<\|w\|_X\leq 1\}} \left[\frac{\|F(x+rh,s)-F(x,s)-(\frac{\partial}{\partial x}F)(x,s)rh\|_Y}{r\|h\|_X}\right] = 0. \quad (2.24)$$

This reveals that for all $x \in X$, $s \in S$ it holds that

$$\limsup_{(0,\infty)\ni r\to 0} \sup_{h\in\{w\in X\colon 0<\|w\|_X\leq 1\}} \left\|\tfrac{F(x+rh,s)-F(x,s)}{r} - (\tfrac{\partial}{\partial x}F)(x,s)h\right\|_Y = 0. \quad (2.25)$$

This and (2.17) demonstrate that for all $x \in X$, $s \in S$ we have that

$$\begin{aligned} &\limsup_{(0,\infty)\ni r\to 0} \left\|\Big(\{w\in X\colon \|w\|_X\leq 1\}\ni h\mapsto \tfrac{F(x+rh,s)-F(x,s)}{r}\in Y\Big) - \iota\Big(\big(\tfrac{\partial}{\partial x}F\big)(x,s)\Big)\right\|_V \\ &= \limsup_{(0,\infty)\ni r\to 0} \sup_{h\in\{w\in X\colon 0<\|w\|_X\leq 1\}} \left\|\tfrac{F(x+rh,s)-F(x,s)}{r} - (\tfrac{\partial}{\partial x}F)(x,s)h\right\|_Y = 0. \end{aligned} \quad (2.26)$$

Next observe that Lemma 2.2.1 (with $S = S$, $\mathcal{S} = \mathcal{S}$, $X = \{w \in X\colon \|w\|_X \leq 1\}$, $d_X = (\{w \in X\colon \|w\|_X \leq 1\} \times \{w \in X\colon \|w\|_X \leq 1\} \ni (y,z) \mapsto \|y-z\|_X \in [0,\infty))$, $Y = Y$, $d_Y = (Y \times Y \ni (y,z) \mapsto \|y-z\|_Y \in [0,\infty))$, $f = (\{w \in X\colon \|w\|_X \leq 1\} \times S \ni (h,s) \mapsto F(x+rh,s) \in Y)$ for $x \in X$, $r \in (0,\infty)$ in the notation of Lemma 2.2.1) implies that for all $x \in X$, $r \in (0,\infty)$ we have that

$$(S \ni s \mapsto (\{w \in X\colon \|w\|_X \leq 1\} \ni h \mapsto F(x+rh,s) \in Y) \in V) \quad (2.27)$$

is $\mathcal{S}/\mathcal{B}(V)$-measurable. This and (2.26) prove that for all $x \in X$ we have that

$$\Big(S \ni s \mapsto \iota\big((\tfrac{\partial}{\partial x}F)(x,s)\big) \in V\Big) \quad (2.28)$$

is $\mathcal{S}/\mathcal{B}(V)$-measurable. Moreover, note that for all $f_n \in \{f \in V\colon f \text{ is linear}\}$, $n \in \mathbb{N}$, and all functions $g\colon \{w \in X\colon \|w\|_X \leq 1\} \to Y$ with

$$\limsup_{n\to\infty} \sup_{h\in\{w\in X\colon \|w\|_X\leq 1\}} \|f_n(h) - g(h)\|_Y = 0 \quad (2.29)$$

we have that

$$g \in \{f \in V\colon f \text{ is linear}\}. \quad (2.30)$$

This ensures that

$$\{f \in V : f \text{ is linear}\} \in \mathcal{B}(V). \tag{2.31}$$

Combining this, the fact that $\forall\, A \in L(X,Y) : \iota(A) \in \{f \in V : f \text{ is linear}\}$, and (2.28) proves that

$$\big(S \ni s \mapsto \iota\big((\tfrac{\partial}{\partial x}F)(x,s)\big) \in \{f \in V : f \text{ is linear}\}\big) \tag{2.32}$$

is $\mathcal{S}/\mathcal{B}(\{f \in V : f \text{ is linear}\})$-measurable. Furthermore, observe that for all $A \in L(X,Y)$, $x \in X \backslash \{0\}$ we have that

$$\psi(\iota(A))x = \|x\|_X(\iota(A))\Big(\tfrac{x}{\|x\|_X}\Big) = \|x\|_X A\Big(\tfrac{x}{\|x\|_X}\Big) = Ax. \tag{2.33}$$

This implies that for all $A \in L(X,Y)$ we have that

$$\psi(\iota(A)) = A. \tag{2.34}$$

Next note that for all $f_1, f_2 \in \{f \in V : f \text{ is linear}\}$ we have that

$$\begin{aligned} \|\psi(f_1) - \psi(f_2)\|_{L(X,Y)} &= \sup_{h \in X\backslash\{0\}} \frac{\|\psi(f_1)h - \psi(f_2)h\|_Y}{\|h\|_X} \\ &= \sup_{h \in X\backslash\{0\}} \frac{\|h\|_X \big\| f_1\big(\frac{h}{\|h\|_X}\big) - f_2\big(\frac{h}{\|h\|_X}\big)\big\|_Y}{\|h\|_X} \\ &= \sup\nolimits_{h \in \{w \in X : \|w\|_X \le 1\}} \|(f_1 - f_2)(h)\|_Y. \end{aligned} \tag{2.35}$$

Combining this and (2.17) establishes that

$$\psi \in C(\{f \in V : f \text{ is linear}\}, L(X,Y)). \tag{2.36}$$

This, (2.31), (2.32), and (2.34) demonstrate that for all $x \in X$ we have that

$$\big(S \ni s \mapsto (\tfrac{\partial}{\partial x}F)(x,s) \in L(X,Y)\big) \tag{2.37}$$

is $\mathcal{S}/\mathcal{B}(L(X,Y))$-measurable. The proof of Lemma 2.2.3 is thus completed. □

Lemma 2.2.4. *Let $(S, \mathcal{S})$ be a measurable space, let (X, d_X) be a separable metric space, let (Y, d_Y) be a metric space, let $F : X \times S \to Y$ satisfy for all $s \in S$, $x \in X$ that $(X \ni y \mapsto F(y,s) \in Y) \in C(X,Y)$ and $(S \ni w \mapsto F(x,w) \in Y)$ is $\mathcal{S}/\mathcal{B}(Y)$-measurable. Then F is $(\mathcal{B}(X) \otimes \mathcal{S})/\mathcal{B}(Y)$-measurable.*

Proof of Lemma 2.2.4. This is a direct consequence of, e.g., Aliprantis & Border [1, Lemma 4.51]. The proof of Lemma 2.2.4 is thus completed. □

Corollary 2.2.5. *Let $d, m, n \in \mathbb{N}$, let $(S, \mathcal{S})$ be a measurable space, let $F = (F(x,s))_{(x,s)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}^m$ be $(\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{S})/\mathcal{B}(\mathbb{R}^m)$-measurable, and assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto F(x,s) \in \mathbb{R}^m) \in C^n(\mathbb{R}^d, \mathbb{R}^m)$. Then we have for all $k \in \{1, 2, \ldots, n\}$, $x \in \mathbb{R}^d$ that*

$$\big(S \ni s \mapsto (\tfrac{\partial^k}{\partial x^k} F)(x,s) \in L^{(k)}(\mathbb{R}^d, \mathbb{R}^m)\big) \tag{2.38}$$

is $\mathcal{S}/\mathcal{B}(L^{(k)}(\mathbb{R}^d, \mathbb{R}^m))$-measurable.

Proof of Corollary 2.2.5. This is a direct consequence of Lemma 2.2.3 and Lemma 2.2.4. The proof of Corollary 2.2.5 is thus completed. □

Lemma 2.2.6. *Let $d, m, n \in \mathbb{N}$, let $(S, \mathcal{S}, \mu)$ be a finite measure space, let $F = (F(x,s))_{(x,s)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d\times S \to \mathbb{R}^m$ be $(\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{S})/\mathcal{B}(\mathbb{R}^m)$-measurable, let $f\colon \mathbb{R}^d \to \mathbb{R}^m$ be $(n-1)$-times differentiable, assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto F(x,s) \in \mathbb{R}^m) \in C^n(\mathbb{R}^d, \mathbb{R}^m)$, and assume for all $x \in \mathbb{R}^d$ that*

$$\begin{aligned}\inf_{\delta\in(0,\infty)} \sup_{u\in[-\delta,\delta]^d} \int_S \|(\tfrac{\partial^{n-1}}{\partial x^{n-1}} F)(x,z)\|_{L^{(n-1)}(\mathbb{R}^d,\mathbb{R}^m)} \\ + \|(\tfrac{\partial^n}{\partial x^n} F)(x+u,z)\|^{1+\delta}_{L^{(n)}(\mathbb{R}^d,\mathbb{R}^m)}\,\mu(dz) < \infty\end{aligned} \tag{2.39}$$

(cf. Corollary 2.2.5) and

$$f^{(n-1)}(x) = \int_S (\tfrac{\partial^{n-1}}{\partial x^{n-1}} F)(x,s)\,\mu(ds). \tag{2.40}$$

Then

(i) we have that $f \in C^n(\mathbb{R}^d, \mathbb{R}^m)$ and

(ii) we have for all $x \in \mathbb{R}^d$ that

$$f^{(n)}(x) = \int_S (\tfrac{\partial^n}{\partial x^n} F)(x,s)\,\mu(ds). \tag{2.41}$$

Proof of Lemma 2.2.6. Throughout this proof let $f_1, f_2, \ldots, f_m\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that

$$f(x) = (f_1(x), f_2(x), \ldots, f_m(x)), \tag{2.42}$$

let $F_1, F_2, \ldots, F_m\colon \mathbb{R}^d \times S \to \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$, $s \in S$ that

$$F(x,s) = (F_1(x,s), F_2(x,s), \ldots, F_m(x,s)), \tag{2.43}$$

let $\delta_x \in (0,\infty)$, $x \in \mathbb{R}^d$, satisfy for all $x \in \mathbb{R}^d$ that

$$\sup_{v \in [-\delta_x,\delta_x]^d} \int_S \|(\tfrac{\partial^n}{\partial x^n} F)(x+v,s)\|_{L^{(n)}(\mathbb{R}^d,\mathbb{R}^m)}^{1+\delta_x} \, \mu(ds) < \infty, \tag{2.44}$$

and let $e_1 = (1,0,\dots,0)$, $e_2 = (0,1,0,\dots,0)$, $\dots$, $e_d = (0,0,\dots,0,1) \in \mathbb{R}^d$. Note that, e.g., Coleman [25, pages 93-94, Section 4.5] assures that for all $x \in \mathbb{R}^d$, $i_1, i_2, \dots, i_n \in \{1,2,\dots,d\}$, $j \in \{1,2,\dots,m\}$, $s \in S$ we have that

$$(\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_n}} F_j)(x,s) = (\tfrac{\partial^n}{\partial x^n} F_j)(x,s)(e_{i_1}, e_{i_2}, \dots, e_{i_n}). \tag{2.45}$$

This ensures that for all $x \in \mathbb{R}^d$, $i_1, i_2, \dots, i_n \in \{1,2,\dots,d\}$, $j \in \{1,2,\dots,m\}$ we have that

$$\begin{aligned}
&\sup_{v \in [-\delta_x,\delta_x]^d} \int_S \big|(\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_n}} F_j)(x+v,s)\big|^{1+\delta_x} \, \mu(ds) \\
&= \sup_{v \in [-\delta_x,\delta_x]^d} \int_S \big|(\tfrac{\partial^n}{\partial x^n} F_j)(x+v,s)(e_{i_1}, e_{i_2}, \dots, e_{i_n})\big|^{1+\delta_x} \, \mu(ds) \\
&\leq \sup_{v \in [-\delta_x,\delta_x]^d} \int_S \big(\|(\tfrac{\partial^n}{\partial x^n} F_j)(x+v,s)\|_{L^{(n)}(\mathbb{R}^d,\mathbb{R})} \|e_{i_1}\|_{\mathbb{R}^d} \|e_{i_2}\|_{\mathbb{R}^d} \cdots \|e_{i_n}\|_{\mathbb{R}^d}\big)^{1+\delta_x} \, \mu(ds) \\
&= \sup_{v \in [-\delta_x,\delta_x]^d} \int_S \|(\tfrac{\partial^n}{\partial x^n} F_j)(x+v,s)\|_{L^{(n)}(\mathbb{R}^d,\mathbb{R})}^{1+\delta_x} \, \mu(ds).
\end{aligned} \tag{2.46}$$

In addition, note that, e.g., Coleman [25, Proposition 4.6] and (2.43) demonstrate that for all $x \in \mathbb{R}^d$, $s \in S$ we have that

$$(\tfrac{\partial^n}{\partial x^n} F)(x,s) = \big((\tfrac{\partial^n}{\partial x^n} F_1)(x,s), (\tfrac{\partial^n}{\partial x^n} F_2)(x,s), \dots, (\tfrac{\partial^n}{\partial x^n} F_m)(x,s)\big). \tag{2.47}$$

This reveals that for all $x \in \mathbb{R}^d$, $s \in S$, $j \in \{1,2,\dots,m\}$ it holds that

$$\begin{aligned}
\|(\tfrac{\partial^n}{\partial x^n} F_j)(x,s)\|_{L^{(n)}(\mathbb{R}^d,\mathbb{R})} &= \sup_{y_1,y_2,\dots,y_n \in \mathbb{R}^d \setminus \{0\}} \frac{|(\frac{\partial^n}{\partial x^n} F_j)(x,s)(y_1,y_2,\dots,y_n)|}{\|y_1\|_{\mathbb{R}^d} \|y_2\|_{\mathbb{R}^d} \dots \|y_n\|_{\mathbb{R}^d}} \\
&\leq \sup_{y_1,y_2,\dots,y_n \in \mathbb{R}^d \setminus \{0\}} \frac{\|(\frac{\partial^n}{\partial x^n} F)(x,s)(y_1,y_2,\dots,y_n)\|_{\mathbb{R}^m}}{\|y_1\|_{\mathbb{R}^d} \|y_2\|_{\mathbb{R}^d} \dots \|y_n\|_{\mathbb{R}^d}} = \|(\tfrac{\partial^n}{\partial x^n} F)(x,s)\|_{L^{(n)}(\mathbb{R}^d,\mathbb{R}^m)}.
\end{aligned} \tag{2.48}$$

Combining this, (2.46), and (2.44) implies that for all $x \in \mathbb{R}^d$, $i_1, i_2, \dots, i_n \in \{1,2,\dots,d\}$, $j \in \{1,2,\dots,m\}$ we have that

$$\begin{aligned}
&\sup_{v \in [-\delta_x,\delta_x]^d} \int_S \big|(\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \dots \partial x_{i_n}} F_j)(x+v,s)\big|^{1+\delta_x} \, \mu(ds) \\
&\leq \sup_{v \in [-\delta_x,\delta_x]^d} \int_S \|(\tfrac{\partial^n}{\partial x^n} F)(x+v,s)\|_{L^{(n)}(\mathbb{R}^d,\mathbb{R}^m)}^{1+\delta_x} \, \mu(ds) < \infty.
\end{aligned} \tag{2.49}$$

Next observe that the assumption that $\forall\, s \in S\colon (\mathbb{R}^d \ni y \mapsto F(y,s) \in \mathbb{R}^m) \in C^n(\mathbb{R}^d, \mathbb{R}^m)$ and the fundamental theorem of calculus imply that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, $h \in \mathbb{R}$ we have that

$$\begin{aligned}
&(\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} f_j)(x + h e_{i_n}) - (\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} f_j)(x) \\
&= \int_S (\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} F_j)(x + h e_{i_n}, s)\, \mu(ds) - \int_S (\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} F_j)(x, s)\, \mu(ds) \\
&= \int_S (\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} F_j)(x + h e_{i_n}, s) - (\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} F_j)(x, s)\, \mu(ds) \\
&= \int_S \int_0^h (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + u e_{i_n}, s)\, du\, \mu(ds).
\end{aligned} \tag{2.50}$$

Moreover, note that Tonelli's theorem, Hölder's inequality, and (2.49) prove that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, $h \in [-\delta_x, \delta_x]$ we have that

$$\begin{aligned}
&\left| \int_S \int_0^h \left| (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + u e_{i_n}, s) \right| du\, \mu(ds) \right| \\
&= \left| \int_0^h \int_S \left| (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + u e_{i_n}, s) \right| \mu(ds)\, du \right| \\
&\leq |h| \sup_{v \in [-\delta_x, \delta_x]^d} \int_S \left| (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + v, s) \right| \mu(ds) \\
&\leq |h| \sup_{v \in [-\delta_x, \delta_x]^d} \left(\int_S \left| (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + v, s) \right|^{1+\delta_x} \mu(ds) \right)^{\frac{1}{1+\delta_x}} \cdot |\mu(S)|^{(1 - \frac{1}{1+\delta_x})} \\
&= |h| \left(\sup_{v \in [-\delta_x, \delta_x]^d} \int_S \left| (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + v, s) \right|^{1+\delta_x} \mu(ds) \right)^{\frac{1}{1+\delta_x}} \cdot |\mu(S)|^{(1 - \frac{1}{1+\delta_x})} < \infty.
\end{aligned} \tag{2.51}$$

This, Fubini's theorem, and (2.50) assure that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, $h \in [-\delta_x, \delta_x]$ we have that

$$\begin{aligned}
&(\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} f_j)(x + h e_{i_n}) - (\tfrac{\partial^{n-1}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n-1}}} f_j)(x) \\
&= \int_S \int_0^h (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + u e_{i_n}, s)\, du\, \mu(ds) \\
&= \int_0^h \int_S (\tfrac{\partial^{n}}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_{n}}} F_j)(x + u e_{i_n}, s)\, \mu(ds)\, du.
\end{aligned} \tag{2.52}$$

In addition, observe that, e.g., Klenke [59, Corollary 6.21] and (2.49) ensure that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ we have that

$$\big(S \ni s \mapsto (\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x + v, s) \in \mathbb{R}\big), v \in [-\delta_x, \delta_x]^d, \tag{2.53}$$

is a uniformly integrable family of functions. This reveals that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, and all functions $u = (u_k)_{k \in \mathbb{N}} \colon \mathbb{N} \to [-\delta_x, \delta_x]^d$ it holds that

$$\big(S \ni s \mapsto (\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x + u_k, s) \in \mathbb{R}\big), k \in \mathbb{N}, \tag{2.54}$$

is a uniformly integrable sequence of functions. This and the Vitali convergence theorem (see, e.g., Klenke [59, Theorem 6.25]) assure that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, and all functions $u = (u_k)_{k \in \mathbb{N}} \colon \mathbb{N} \to [-\delta_x, \delta_x]^d$ with $\limsup_{k \to \infty} \|u_k\|_{\mathbb{R}^d} = 0$ we have that

$$\limsup_{k \to \infty} \int_S \big|(\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x + u_k, s) - (\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x, s)\big| \, \mu(ds) = 0. \tag{2.55}$$

This reveals that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ it holds that

$$\limsup_{v \to 0} \int_S \big|(\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x + v e_{i_n}, s) - (\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x, s)\big| \, \mu(ds) = 0. \tag{2.56}$$

This implies that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, $\varepsilon \in (0, \infty)$ there exists $\delta \in (0, \infty)$ such that

$$\forall \, v \in (-\delta, \delta) \colon \int_S \big|(\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x + v e_{i_n}, s) - (\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x, s)\big| \, \mu(ds) < \varepsilon. \tag{2.57}$$

Combining this and (2.52) ensures that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, $\varepsilon \in (0, \infty)$ there exists $\delta \in (0, \infty)$ such that for all $h \in (-\delta, \delta) \backslash \{0\}$

we have that

$$\begin{aligned} &\Big|\frac{1}{h}\Big(\big(\tfrac{\partial^{n-1}}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_{n-1}}} f_j\big)(x+he_{i_n}) - \big(\tfrac{\partial^{n-1}}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_{n-1}}} f_j\big)(x)\Big) \\ &\quad - \int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\,\mu(ds)\Big| \\ &= \Big|\frac{1}{h}\int_0^h \int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x+ue_{i_n},s)\,\mu(ds)\,du \\ &\quad - \frac{1}{h}\int_0^h \int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\,\mu(ds)\,du\Big| \\ &\le \frac{1}{h}\int_0^h \int_S \big|\big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x+ue_{i_n},s) - \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\big|\,\mu(ds)\,du < \varepsilon. \end{aligned} \tag{2.58}$$

This demonstrates that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ we have that

$$\begin{aligned} \limsup_{\substack{h\to 0\\ h\in\mathbb{R}\setminus\{0\}}} &\Big|\frac{1}{h}\Big(\big(\tfrac{\partial^{n-1}}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_{n-1}}} f_j\big)(x+he_{i_n}) - \big(\tfrac{\partial^{n-1}}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_{n-1}}} f_j\big)(x)\Big) \\ &\qquad - \int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\,\mu(ds)\Big| = 0. \end{aligned} \tag{2.59}$$

This reveals that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ it holds that

$$\big(\tfrac{\partial}{\partial x_{i_n}}\big(\tfrac{\partial^{n-1}}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_{n-1}}} f_j\big)\big)(x) = \int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\,\mu(ds). \tag{2.60}$$

Next observe that (2.55) proves that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$, and all functions $u = (u_k)_{k\in\mathbb{N}} \colon \mathbb{N} \to \mathbb{R}^d$ with $\limsup_{k\to\infty} \|u_k\|_{\mathbb{R}^d} = 0$ we have that

$$\begin{aligned} \limsup_{k\to\infty} &\Big|\int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x+u_k,s)\,\mu(ds) - \int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\,\mu(ds)\Big| \\ &\le \limsup_{k\to\infty} \int_S \big|\big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x+u_k,s) - \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\big|\,\mu(ds) = 0. \end{aligned} \tag{2.61}$$

This reveals that for all $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ it holds that

$$\Big(\mathbb{R}^d \ni x \mapsto \int_S \big(\tfrac{\partial^n}{\partial x_{i_1}\partial x_{i_2}\ldots\partial x_{i_n}} F_j\big)(x,s)\,\mu(ds) \in \mathbb{R}\Big) \in C(\mathbb{R}^d, \mathbb{R}). \tag{2.62}$$

Combining this, (2.60) and, e.g., Coleman [25, Corollary 2.2] demonstrates that for all $x \in \mathbb{R}^d$, $i_1, i_2, \ldots, i_n \in \{1, 2, \ldots, d\}$, $j \in \{1, 2, \ldots, m\}$ we have that $f \in C^n(\mathbb{R}^d, \mathbb{R}^m)$ and

$$(\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} f_j)(x) = \int_S (\tfrac{\partial^n}{\partial x_{i_1} \partial x_{i_2} \ldots \partial x_{i_n}} F_j)(x, s)\, \mu(ds). \tag{2.63}$$

This establishes items (i)–(ii). The proof of Lemma 2.2.6 is thus completed. □

Lemma 2.2.7. *Let $d, m, n \in \mathbb{N}$, let $(S, \mathcal{S}, \mu)$ be a finite measure space, let $F = (F(x, s))_{(x,s) \in \mathbb{R}^d \times S} \colon \mathbb{R}^d \times S \to \mathbb{R}^m$ be $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{S})/\mathcal{B}(\mathbb{R}^m)$-measurable, let $f \colon \mathbb{R}^d \to \mathbb{R}^m$ be a function, assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto F(x, s) \in \mathbb{R}^m) \in C^n(\mathbb{R}^d, \mathbb{R}^m)$, and assume for all $x \in \mathbb{R}^d$ that*

$$\inf_{\delta \in (0,\infty)} \sup_{u \in [-\delta,\delta]^d} \int_S \|F(x, z)\|_{\mathbb{R}^m} + \sum_{k=1}^{n} \|(\tfrac{\partial^k}{\partial x^k} F)(x + u, z)\|^{1+\delta}_{L^{(k)}(\mathbb{R}^d, \mathbb{R}^m)}\, \mu(dz) < \infty \tag{2.64}$$

(cf. Corollary 2.2.5) and

$$f(x) = \int_S F(x, s)\, \mu(ds). \tag{2.65}$$

Then

(i) we have that $f \in C^n(\mathbb{R}^d, \mathbb{R}^m)$ and

(ii) we have for all $k \in \{1, 2, \ldots, n\}$, $x \in \mathbb{R}^d$ that

$$f^{(k)}(x) = \int_S (\tfrac{\partial^k}{\partial x^k} F)(x, s)\, \mu(ds). \tag{2.66}$$

Proof of Lemma 2.2.7. This is a direct consequence of Lemma 2.2.6. The proof of Lemma 2.2.7 is thus completed. □

2.3 Existence, uniqueness, and regularity results for solutions of ordinary differential equations (ODEs)

Proposition 2.3.1. *Let $d \in \mathbb{N}$, $L, T \in [0, \infty)$, $f \in C([0, T] \times \mathbb{R}^d, \mathbb{R}^d)$ satisfy for all $t \in [0, T]$, $x, y \in \mathbb{R}^d$ that*

$$\|f(t, x) - f(t, y)\|_{\mathbb{R}^d} \leq L\|x - y\|_{\mathbb{R}^d}. \tag{2.67}$$

Then there exists a unique $\chi \in C(\{(s,t) \in [0,T]^2 \colon s \leq t\} \times \mathbb{R}^d, \mathbb{R}^d)$ *which satisfies for all* $x \in \mathbb{R}^d$, $s \in [0,T]$, $t \in [s,T]$ *that*

$$\chi(s,t,x) = x + \int_s^t f(u, \chi(s,u,x))\, du. \tag{2.68}$$

Proof of Proposition 2.3.1. Throughout this proof let $g \colon (-1, T+1) \times \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $s \in (-1,0)$, $t \in [0,T]$, $u \in (T, T+1)$, $x \in \mathbb{R}^d$ that $g(s,x) = f(0,x)$, $g(t,x) = f(t,x)$, and $g(u,x) = f(T,x)$. Note that the hypothesis that $f \in C([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$ and (2.67) imply that for all $t \in (-1, T+1)$, $x, y \in \mathbb{R}^d$ we have that $g \in C((-1,T+1) \times \mathbb{R}^d, \mathbb{R}^d)$ and

$$\|g(t,x) - g(t,y)\|_{\mathbb{R}^d} \leq L\|x-y\|_{\mathbb{R}^d}. \tag{2.69}$$

This and, e.g., Teschl [102, Corollary 2.6] ensure that there exists a unique $\chi \colon \{(s,t) \in [0,T]^2 \colon s \leq t\} \times \mathbb{R}^d \to \mathbb{R}^d$ which satisfies for all $x \in \mathbb{R}^d$, $s \in [0,T]$, $t \in [s,T]$ that $([s,T] \ni u \mapsto \chi(s,u,x) \in \mathbb{R}^d) \in C([s,T], \mathbb{R}^d)$ and

$$\chi(s,t,x) = x + \int_s^t g(u, \chi(s,u,x))\, du = x + \int_s^t f(u, \chi(s,u,x))\, du. \tag{2.70}$$

Moreover, observe that, e.g., Teschl [102, Theorem 2.9] assures that $\chi \in C(\{(s,t) \in [0,T]^2 \colon s \leq t\} \times \mathbb{R}^d, \mathbb{R}^d)$. The proof of Proposition 2.3.1 is thus completed. □

Lemma 2.3.2. *Let* $d, n \in \mathbb{N}$, $T \in (0,\infty)$, $f \in C^n(\mathbb{R}^d, \mathbb{R}^d)$ *and let* $\theta^\vartheta \in C([0,T], \mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, *satisfy for all* $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ *that*

$$\theta^\vartheta_t = \vartheta + \int_0^t f(\theta^\vartheta_s)\, ds. \tag{2.71}$$

Then we have that $([0,T] \times \mathbb{R}^d \ni (t,\vartheta) \mapsto \theta^\vartheta_t \in \mathbb{R}^d) \in C^n([0,T] \times \mathbb{R}^d, \mathbb{R}^d)$.

Proof of Lemma 2.3.2. This a direct consequence of, e.g., Coleman [25, Theorem 10.3]. The proof of Lemma 2.3.2 is thus completed. □

Corollary 2.3.3. *Let* $d, n \in \mathbb{N}$, $f \in C^n(\mathbb{R}^d, \mathbb{R}^d)$ *and let* $\theta^\vartheta \in C([0,\infty), \mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, *satisfy for all* $\vartheta \in \mathbb{R}^d$, $t \in [0,\infty)$ *that*

$$\theta^\vartheta_t = \vartheta + \int_0^t f(\theta^\vartheta_s)\, ds. \tag{2.72}$$

Then we have that $([0,\infty) \times \mathbb{R}^d \ni (t,\vartheta) \mapsto \theta^\vartheta_t \in \mathbb{R}^d) \in C^n([0,\infty) \times \mathbb{R}^d, \mathbb{R}^d)$.

Proof of Corollary 2.3.3. This is a direct consequence of Lemma 2.3.2. The proof of Corollary 2.3.3 is thus completed. □

Lemma 2.3.4. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, $\vartheta \in \mathbb{R}^d$, $A \in C([0,T], L(\mathbb{R}^d, \mathbb{R}^d))$, let $y_1, y_2 \colon [0,T] \to \mathbb{R}^d$ be $\mathcal{B}([0,T])/\mathcal{B}(\mathbb{R}^d)$-measurable, and assume for all $t \in [0,T]$, $i \in \{1,2\}$ that*

$$\int_0^T \|y_i(s)\|_{\mathbb{R}^d}\, ds < \infty \qquad \text{and} \qquad y_i(t) = \vartheta + \int_0^t A(s) y_i(s)\, ds. \tag{2.73}$$

Then we have that $y_1 = y_2$.

Proof of Lemma 2.3.4. First, note that (2.73) and the triangle inequality ensure that

$$\int_0^T \|y_1(s) - y_2(s)\|_{\mathbb{R}^d}\, ds \leq \int_0^T \|y_1(s)\|_{\mathbb{R}^d} + \|y_2(s)\|_{\mathbb{R}^d}\, ds < \infty. \tag{2.74}$$

Next observe that (2.73) and the triangle inequality for the Bochner integral prove that for all $t \in [0,T]$ we have that

$$\begin{aligned} \|y_1(t) - y_2(t)\|_{\mathbb{R}^d} &= \left\| \int_0^t A(s)(y_1(s) - y_2(s))\, ds \right\|_{\mathbb{R}^d} \\ &\leq \int_0^t \|A(s)(y_1(s) - y_2(s))\|_{\mathbb{R}^d}\, ds \\ &\leq \int_0^t \|A(s)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} \|y_1(s) - y_2(s)\|_{\mathbb{R}^d}\, ds \\ &\leq \left[\sup\nolimits_{v \in [0,T]} \|A(v)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} \right] \int_0^t \|y_1(s) - y_2(s)\|_{\mathbb{R}^d}\, ds. \end{aligned} \tag{2.75}$$

Moreover, note that the fact that $[0,T]$ is a compact set, the assumption that $A \in C([0,T], L(\mathbb{R}^d, \mathbb{R}^d))$, and Lemma 2.2.2 establish that

$$\sup_{v \in [0,T]} \|A(v)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} < \infty. \tag{2.76}$$

Combining (2.74), (2.75), and the Gronwall integral inequality in Lemma 2.1.2 hence assures that for all $t \in [0,T]$ we have that

$$y_1(t) = y_2(t). \tag{2.77}$$

The proof of Lemma 2.3.4 is thus completed. □

2.4 Existence results for solutions of first-order Kolmogorov backward PDEs

Proposition 2.4.1. *Let $d \in \mathbb{N}$, $T \in (0,\infty)$, $\psi \in C^1(\mathbb{R}^d,\mathbb{R})$, $f \in C^1(\mathbb{R}^d,\mathbb{R}^d)$, let $\theta^\vartheta = (\theta^\vartheta_t)_{t\in[0,T]} \in C([0,T],\mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, satisfy for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ that*

$$\theta^\vartheta_t = \vartheta + \int_0^t f(\theta^\vartheta_s)\, ds, \tag{2.78}$$

and let $u = (u(t,\vartheta))_{(t,\vartheta)\in[0,T]\times\mathbb{R}^d} \colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ that $u(t,\vartheta) = \psi(\theta^\vartheta_t)$. Then

(i) we have that $u \in C^1([0,T]\times\mathbb{R}^d,\mathbb{R})$ and

$$\big([0,T]\times\mathbb{R}^d \ni (t,\vartheta) \mapsto \theta^\vartheta_t \in \mathbb{R}^d\big) \in C^1([0,T]\times\mathbb{R}^d,\mathbb{R}^d), \tag{2.79}$$

(ii) we have for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ that $f(\theta^\vartheta_t) = (\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_t)f(\vartheta)$, and

(iii) we have for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ that

$$(\tfrac{\partial}{\partial t}u)(t,\vartheta) = (\tfrac{\partial}{\partial\vartheta}u)(t,\vartheta)f(\vartheta) = \langle(\nabla_\vartheta u)(t,\vartheta), f(\vartheta)\rangle_{\mathbb{R}^d}. \tag{2.80}$$

Proof of Proposition 2.4.1. First, observe that Lemma 2.3.2, the assumption that $f \in C^1(\mathbb{R}^d,\mathbb{R}^d)$, and the assumption that $\psi \in C^1(\mathbb{R}^d,\mathbb{R})$ establish item (i). Next let $y_\vartheta \colon [0,T] \to \mathbb{R}^d$, $\vartheta \in \mathbb{R}^d$, satisfy for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ that

$$y_\vartheta(t) = (\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_t)f(\vartheta), \tag{2.81}$$

and let $z_\vartheta \colon [0,T] \to \mathbb{R}^d$, $\vartheta \in \mathbb{R}^d$, satisfy for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ that

$$z_\vartheta(t) = f(\theta^\vartheta_t). \tag{2.82}$$

Observe that the assumption that $f \in C^1(\mathbb{R}^d,\mathbb{R}^d)$ and (2.79) ensure that for all $\vartheta \in \mathbb{R}^d$ we have that $y_\vartheta, z_\vartheta \in C([0,T],\mathbb{R}^d)$. This and Lemma 2.2.2 prove that for all $\vartheta \in \mathbb{R}^d$ we have that

$$\int_0^T \|y_\vartheta(s)\|_{\mathbb{R}^d} + \|z_\vartheta(s)\|_{\mathbb{R}^d}\, ds \le T \sup_{s\in[0,T]} \|y_\vartheta(s)\|_{\mathbb{R}^d} + T \sup_{s\in[0,T]} \|z_\vartheta(s)\|_{\mathbb{R}^d} < \infty. \tag{2.83}$$

Next note that the assumption that $f \in C^1(\mathbb{R}^d,\mathbb{R}^d)$ and (2.79) demonstrate that for all $\vartheta \in \mathbb{R}^d$ we have that

$$\big([0,T]\times\mathbb{R}^d \ni (t,h) \mapsto f'(\theta^{\vartheta+h}_t)(\tfrac{\partial}{\partial\vartheta}\theta^{\vartheta+h}_t) \in L(\mathbb{R}^d,\mathbb{R}^d)\big) \in C([0,T]\times\mathbb{R}^d, L(\mathbb{R}^d,\mathbb{R}^d)). \tag{2.84}$$

Lemma 2.2.2 hence assures that for all $\vartheta \in \mathbb{R}^d$ we have that

$$\sup_{(s,h)\in[0,T]\times[-1,1]^d} \|f'(\theta_s^{\vartheta+h})(\tfrac{\partial}{\partial\vartheta}\theta_s^{\vartheta+h})\|_{L(\mathbb{R}^d,\mathbb{R}^d)} < \infty. \tag{2.85}$$

This ensures that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$\sup_{h\in[-1,1]^d} \int_0^t \|f'(\theta_s^{\vartheta+h})(\tfrac{\partial}{\partial\vartheta}\theta_s^{\vartheta+h})\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^2 \, ds < \infty. \tag{2.86}$$

Lemma 2.2.6 and (2.78) hence imply that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$\tfrac{\partial}{\partial\vartheta}\theta_t^\vartheta = \mathrm{Id}_{\mathbb{R}^d} + \int_0^t f'(\theta_s^\vartheta)(\tfrac{\partial}{\partial\vartheta}\theta_s^\vartheta)\, ds. \tag{2.87}$$

This reveals that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ it holds that

$$(\tfrac{\partial}{\partial\vartheta}\theta_t^\vartheta)f(\vartheta) = f(\vartheta) + \int_0^t f'(\theta_s^\vartheta)(\tfrac{\partial}{\partial\vartheta}\theta_s^\vartheta)f(\vartheta)\, ds. \tag{2.88}$$

This and (2.81) assure that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$y_\vartheta(t) = f(\vartheta) + \int_0^t f'(\theta_s^\vartheta) y_\vartheta(s)\, ds. \tag{2.89}$$

Moreover, observe that the assumption that $f \in C^1(\mathbb{R}^d,\mathbb{R}^d)$ and (2.79) prove that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$z_\vartheta \in C^1([0,T],\mathbb{R}^d). \tag{2.90}$$

The fundamental theorem of calculus hence demonstrates that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$z_\vartheta(t) = z_\vartheta(0) + \int_0^t z_\vartheta'(s)\, ds. \tag{2.91}$$

Combining this and (2.78) ensures that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$z_\vartheta(t) = z_\vartheta(0) + \int_0^t f'(\theta_s^\vartheta)f(\theta_s^\vartheta)\, ds = z_\vartheta(0) + \int_0^t f'(\theta_s^\vartheta) z_\vartheta(s)\, ds. \tag{2.92}$$

Furthermore, note that the assumption that $f \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and (2.79) imply that for all $\vartheta \in \mathbb{R}^d$ we have that

$$\big([0,T] \ni s \mapsto f'(\theta_s^\vartheta) \in L(\mathbb{R}^d, \mathbb{R}^d)\big) \in C([0,T], L(\mathbb{R}^d, \mathbb{R}^d)). \tag{2.93}$$

This, (2.89), (2.92), the fact that $\forall\, \vartheta \in \mathbb{R}^d \colon z_\vartheta(0) = f(\vartheta)$, (2.83), and Lemma 2.3.4 demonstrate that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$y_\vartheta(t) = z_\vartheta(t). \tag{2.94}$$

Combining this with the chain rule and the assumption that $\forall\, t \in [0,T]$, $\vartheta \in \mathbb{R}^d \colon u(t,\vartheta) = \psi(\theta_t^\vartheta)$ proves that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$\begin{aligned}(\tfrac{\partial}{\partial t} u)(t,\vartheta) &= \psi'(\theta_t^\vartheta)(\tfrac{\partial}{\partial t}\theta_t^\vartheta) = \psi'(\theta_t^\vartheta) f(\theta_t^\vartheta) = \psi'(\theta_t^\vartheta) z_\vartheta(t) \\ &= \psi'(\theta_t^\vartheta) y_\vartheta(t) = \psi'(\theta_t^\vartheta)(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta) f(\vartheta) = (\tfrac{\partial}{\partial \vartheta} u)(t,\vartheta) f(\vartheta).\end{aligned} \tag{2.95}$$

This and (2.94) establish item (ii) and item (iii). The proof of Proposition 2.4.1 is thus completed. □

Chapter 3

Weak error estimates for stochastic approximation algorithms (SAAs) in the case of general learning rates

In this chapter we use the analysis for first-order Kolmogorov backward PDEs from Chapter 2 above to study weak approximation errors for SAAs in the case of general learning rates. In particular, we establish in Proposition 3.7.1 in Section 3.7 below weak error estimates for SAAs in the case of general learning rates with mini-batches. Our proof of Proposition 3.7.1 employs the well-known results on the possibility of interchanging derivatives and expectations in Section 3.2 below, the essentially well-known spatial regularity results for flows of certain deterministic ODEs in Section 3.3 below, the auxiliary intermediate results on upper bounds for second-order spatial derivatives of certain deterministic flows in Section 3.4 below, the elementary temporal regularity result for SAAs in Lemma 3.5.1 in Section 3.5 below, and the auxiliary intermediate results on a priori estimates for SAAs in Section 3.6 below. In Section 3.9 below we combine Proposition 3.7.1 and the elementary auxiliary results on upper bounds for integrals of certain exponentially decaying functions in Section 3.8 below to establish in Corollary 3.9.1 below weak error estimates for SAAs in the case of polynomially decaying learning rates with mini-batches. In Setting 3.1.1 in Section 3.1 below we present a mathematical framework for describing SAAs in the case of general learning rates. In the results of this chapter we frequently employ Setting 3.1.1.

3.1 Mathematical description for SAAs in the case of general learning rates

Setting 3.1.1. *Let $d \in \mathbb{N}$, $\xi \in \mathbb{R}^d$, let $(S, \mathcal{S})$ be a measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $Z_n \colon \Omega \to S$, $n \in \mathbb{N}$, be i.i.d. random variables, for every set A let $\#_A \in \mathbb{N}_0 \cup \{\infty\}$ be the number of elements of A, let $\gamma \colon [0,\infty) \to \{A \subseteq \mathbb{N} \colon \#_A < \infty\}$ satisfy for all $t \in [0,\infty)$ that $0 < \#_{\{s \in [0,t] \colon \gamma(s) \neq \emptyset\}} < \infty$, let $G = (G(x,s))_{(x,s) \in \mathbb{R}^d \times S} \colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{S})/\mathcal{B}(\mathbb{R}^d)$-measurable, assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto G(x,s) \in \mathbb{R}^d) \in C^2(\mathbb{R}^d, \mathbb{R}^d)$, assume for all $x \in \mathbb{R}^d$ that*

$$\max_{i \in \{1,2\}} \inf_{\delta \in (0,\infty)} \sup_{u \in [-\delta,\delta]^d} \mathbb{E}\Big[\|G(x,Z_1)\|_{\mathbb{R}^d} + \|(\tfrac{\partial^i}{\partial x^i} G)(x+u, Z_1)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)}^{1+\delta}\Big] < \infty \tag{3.1}$$

(cf. Corollary 2.2.5), let $g \colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $x \in \mathbb{R}^d$ that $g(x) = \mathbb{E}[G(x,Z_1)]$, let $\theta^\vartheta = (\theta^\vartheta_t)_{t \in [0,\infty)} \in C([0,\infty), \mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, satisfy for all $t \in [0,\infty)$, $\vartheta \in \mathbb{R}^d$ that

$$\theta^\vartheta_t = \vartheta + \int_0^t g(\theta^\vartheta_s)\, ds \tag{3.2}$$

(cf. item (i) in Lemma 2.2.6), let $[\![\cdot]\!] \colon [0,\infty) \to [0,\infty)$ satisfy for all $t \in [0,\infty)$ that

$$[\![t]\!] = \max\{s \in [0,t] \colon \gamma(s) \neq \emptyset\}, \tag{3.3}$$

and let $\Theta = (\Theta_t(\omega))_{(t,\omega) \in [0,\infty) \times \Omega} \colon [0,\infty) \times \Omega \to \mathbb{R}^d$ be the stochastic process with continuous sample paths (w.c.s.p.) which satisfies for all $t \in (0,\infty)$ that $\Theta_0 = \xi$ and

$$\Theta_t = \Theta_{[\![t]\!]} + \tfrac{(t - [\![t]\!])}{\#_{\gamma([\![t]\!])}} \Big[\textstyle\sum_{n \in \gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_n)\Big]. \tag{3.4}$$

3.2 Sufficient conditions for interchanging derivatives and expectations

Lemma 3.2.1. *Assume Setting 3.1.1. Then*

(i) we have that $g \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and

(ii) we have for all $x \in \mathbb{R}^d$ that $g'(x) = \mathbb{E}[(\tfrac{\partial}{\partial x} G)(x, Z_1)]$.

Proof of Lemma 3.2.1. This is a direct consequence of Lemma 2.2.6. The proof of Lemma 3.2.1 is thus completed. □

Lemma 3.2.2. *Assume Setting 3.1.1, let $n \in \mathbb{N}$, assume for all $s \in S$ that $(\mathbb{R}^d \ni y \mapsto G(y,s) \in \mathbb{R}^d) \in C^n(\mathbb{R}^d, \mathbb{R}^d)$, and assume for all $x \in \mathbb{R}^d$ that*

$$\max_{i \in \{1,2,\ldots,n\}} \inf_{\delta \in (0,\infty)} \sup_{u \in [-\delta,\delta]^d} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i} G)(x+u, Z_1)\|^{1+\delta}_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)}\Big] < \infty. \tag{3.5}$$

Then

(i) we have that $g \in C^n(\mathbb{R}^d, \mathbb{R}^d)$ and

(ii) we have for all $x \in \mathbb{R}^d$ that $\mathbb{E}[\|(\frac{\partial^n}{\partial x^n} G)(x, Z_1)\|_{L^{(n)}(\mathbb{R}^d,\mathbb{R}^d)}] < \infty$ and

$$g^{(n)}(x) = \mathbb{E}\big[(\tfrac{\partial^n}{\partial x^n} G)(x, Z_1)\big]. \tag{3.6}$$

Proof of Lemma 3.2.2. Lemma 2.2.7 (with $d = d$, $n = n$, $(S, \mathcal{S}, \mu) = (\Omega, \mathcal{F}, \mathbb{P})$, $F = (\mathbb{R}^d \times \Omega \ni (y,\omega) \mapsto G(y, Z_1(\omega)) \in \mathbb{R}^d)$, $f = g$, $k = n$ in the notation of Lemma 2.2.7) establishes item (i) and item (ii). The proof of Lemma 3.2.2 is thus completed. □

3.3 Spatial regularity results for flows of deterministic ODEs

Lemma 3.3.1. *Let $d \in \mathbb{N}$, $L \in \mathbb{R}$, $T \in (0,\infty)$, let $g \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ satisfy for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y \rangle_{\mathbb{R}^d} \le L\|x - y\|^2_{\mathbb{R}^d}, \tag{3.7}$$

and let $\theta^\vartheta \in C([0,T], \mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, satisfy for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ that

$$\theta^\vartheta_t = \vartheta + \int_0^t g(\theta^\vartheta_s)\, ds. \tag{3.8}$$

Then we have for all $x, y \in \mathbb{R}^d$, $t \in [0,T]$ that $\|\theta^x_t - \theta^y_t\|_{\mathbb{R}^d} \le \|x - y\|_{\mathbb{R}^d} \exp(Lt)$.

Proof of Lemma 3.3.1. Throughout this proof let $E^{x,y} = (E^{x,y}(t))_{t \in [0,T]} = (E^{x,y}_t)_{t \in [0,T]} \colon [0,T] \to [0,\infty)$, $x, y \in \mathbb{R}^d$, satisfy for all $x, y \in \mathbb{R}^d$, $t \in [0,T]$ that

$$E^{x,y}_t = \|\theta^x_t - \theta^y_t\|^2_{\mathbb{R}^d}. \tag{3.9}$$

Note that the assumption that $g \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, (3.8), and Lemma 2.3.2 ensure that for all $x \in \mathbb{R}^d$ we have that $\theta^x \in C^1([0,T], \mathbb{R}^d)$. This and (3.7) prove that for all $x, y \in \mathbb{R}^d$, $t \in [0,T]$ we have that $E^{x,y} \in C^1([0,T], [0,\infty))$ and

$$(E^{x,y})'(t) = 2\langle g(\theta^x_t) - g(\theta^y_t), \theta^x_t - \theta^y_t \rangle_{\mathbb{R}^d} \le 2L\|\theta^x_t - \theta^y_t\|^2_{\mathbb{R}^d} = 2LE^{x,y}_t. \tag{3.10}$$

The Gronwall differential inequality in Lemma 2.1.1 hence assures that for all $x, y \in \mathbb{R}^d$, $t \in [0, T]$ we have that

$$\|\theta_t^x - \theta_t^y\|_{\mathbb{R}^d}^2 = E_t^{x,y} \leq E_0^{x,y}\, e^{2Lt} = \|x - y\|_{\mathbb{R}^d}^2 e^{2Lt}. \tag{3.11}$$

The proof of Lemma 3.3.1 is thus completed. □

Lemma 3.3.2. *Let $d \in \mathbb{N}$, $L \in \mathbb{R}$, $T \in (0, \infty)$, $g \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and let $\theta^\vartheta \in C([0, T], \mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, satisfy for all $t \in [0, T]$, $x, y \in \mathbb{R}^d$ that*

$$\theta_t^x = x + \int_0^t g(\theta_s^x)\, ds \tag{3.12}$$

and $\|\theta_t^x - \theta_t^y\|_{\mathbb{R}^d} \leq \|x - y\|_{\mathbb{R}^d} \exp(Lt)$. Then we have for all $x, y \in \mathbb{R}^d$ that $\langle g(x) - g(y), x - y\rangle_{\mathbb{R}^d} \leq L\|x - y\|_{\mathbb{R}^d}^2$.

Proof of Lemma 3.3.2. Throughout this proof let $E^{x,y} = (E^{x,y}(t))_{t \in [0,T]} = (E_t^{x,y})_{t \in [0,T]} \colon [0, T] \to [0, \infty)$, $x, y \in \mathbb{R}^d$, satisfy for all $x, y \in \mathbb{R}^d$, $t \in [0, T]$ that

$$E_t^{x,y} = \|\theta_t^x - \theta_t^y\|_{\mathbb{R}^d}^2. \tag{3.13}$$

Observe that for all $x, y \in \mathbb{R}^d$, $h \in (0, T]$ we have that

$$\begin{aligned} \frac{\|\theta_h^x - \theta_h^y\|_{\mathbb{R}^d}^2 - \|\theta_0^x - \theta_0^y\|_{\mathbb{R}^d}^2}{h} &= \frac{\|\theta_h^x - \theta_h^y\|_{\mathbb{R}^d}^2 - \|x - y\|_{\mathbb{R}^d}^2}{h} \\ \leq \frac{\|x - y\|_{\mathbb{R}^d}^2 e^{2Lh} - \|x - y\|_{\mathbb{R}^d}^2 e^{2L \cdot 0}}{h} &= \|x - y\|_{\mathbb{R}^d}^2 \left[\frac{e^{2Lh} - e^{2L \cdot 0}}{h}\right]. \end{aligned} \tag{3.14}$$

Next note that the assumption that $g \in C^1(\mathbb{R}^d, \mathbb{R}^d)$, (3.12), and Lemma 2.3.2 imply that for all $x \in \mathbb{R}^d$ we have that $\theta^x \in C^1([0, T], \mathbb{R}^d)$. This and (3.14) prove that for all $x, y \in \mathbb{R}^d$, $t \in [0, T]$ we have that $E^{x,y} \in C^1([0, T], [0, \infty))$ and

$$\begin{aligned} (E^{x,y})'(0) &= \lim_{\substack{h \to 0 \\ h \in (0,\infty)}} \frac{\|\theta_h^x - \theta_h^y\|_{\mathbb{R}^d}^2 - \|\theta_0^x - \theta_0^y\|_{\mathbb{R}^d}^2}{h} \\ &\leq \|x - y\|_{\mathbb{R}^d}^2 \left[\lim_{\substack{h \to 0 \\ h \in (0,\infty)}} \frac{e^{2Lh} - e^{2L \cdot 0}}{h}\right] = 2L\|x - y\|_{\mathbb{R}^d}^2. \end{aligned} \tag{3.15}$$

This reveals that for all $x, y \in \mathbb{R}^d$ it holds that

$$\begin{aligned} 2\langle x - y, g(x) - g(y)\rangle_{\mathbb{R}^d} &= \left[2\langle \theta_t^x - \theta_t^y, g(\theta_t^x) - g(\theta_t^y)\rangle_{\mathbb{R}^d}\right]_{t=0} \\ &= (E^{x,y})'(0) \leq 2L\|x - y\|_{\mathbb{R}^d}^2. \end{aligned} \tag{3.16}$$

The proof of Lemma 3.3.2 is thus completed. □

Corollary 3.3.3. *Let $d \in \mathbb{N}$, $L \in \mathbb{R}$, $T \in (0,\infty)$, $g \in C^1(\mathbb{R}^d,\mathbb{R}^d)$ and let $\theta^\vartheta \in C([0,T],\mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, satisfy for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ that*

$$\theta_t^\vartheta = \vartheta + \int_0^t g(\theta_s^\vartheta)\, ds. \tag{3.17}$$

Then the following two statements are equivalent:

(i) It holds for all $x, y \in \mathbb{R}^d$ that $\langle g(x) - g(y), x - y\rangle_{\mathbb{R}^d} \leq L\|x-y\|_{\mathbb{R}^d}^2$.

(ii) It holds for all $x, y \in \mathbb{R}^d$, $t \in [0,T]$ that $\|\theta_t^x - \theta_t^y\|_{\mathbb{R}^d} \leq \|x-y\|_{\mathbb{R}^d} \exp(Lt)$.

Proof of Corollary 3.3.3. Observe that Lemma 3.3.1 ensures that ((i) $\Rightarrow$ (ii)). Moreover, note that Lemma 3.3.2 establishes that ((ii) $\Rightarrow$ (i)). The proof of Corollary 3.3.3 is thus completed. □

Lemma 3.3.4. *Assume Setting 3.1.1 and let $L \in \mathbb{R}$ satisfy for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y\rangle_{\mathbb{R}^d} \leq L\|x-y\|_{\mathbb{R}^d}^2. \tag{3.18}$$

Then

(i) we have that $([0,\infty) \times \mathbb{R}^d \ni (t,\vartheta) \mapsto \theta_t^\vartheta \in \mathbb{R}^d) \in C^1([0,\infty)\times\mathbb{R}^d,\mathbb{R}^d)$ and

(ii) we have for all $\vartheta \in \mathbb{R}^d$, $t \in [0,\infty)$ that $\|(\frac{\partial}{\partial\vartheta}\theta_t^\vartheta)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} \leq \exp(Lt)$.

Proof of Lemma 3.3.4. First, observe that Corollary 2.3.3 and Lemma 3.2.1 prove item (i). Next note that item (i) and the triangle inequality imply that for all $\vartheta \in \mathbb{R}^d$, $t \in [0,\infty)$ we have that

$$\begin{aligned}
&\limsup_{\substack{h\to 0\\ h\in\mathbb{R}^d\setminus\{0\}}} \left\| \tfrac{1}{\|h\|_{\mathbb{R}^d}} (\tfrac{\partial}{\partial\vartheta}\theta_t^\vartheta)h \right\|_{\mathbb{R}^d} \\
&\leq \limsup_{\substack{h\to 0\\ h\in\mathbb{R}^d\setminus\{0\}}} \left[\left\| \tfrac{1}{\|h\|_{\mathbb{R}^d}} (\tfrac{\partial}{\partial\vartheta}\theta_t^\vartheta)h - \tfrac{1}{\|h\|_{\mathbb{R}^d}}(\theta_t^{\vartheta+h} - \theta_t^\vartheta) \right\|_{\mathbb{R}^d} + \left\| \tfrac{1}{\|h\|_{\mathbb{R}^d}}(\theta_t^{\vartheta+h} - \theta_t^\vartheta) \right\|_{\mathbb{R}^d} \right] \\
&\leq \limsup_{\substack{h\to 0\\ h\in\mathbb{R}^d\setminus\{0\}}} \left\| \tfrac{1}{\|h\|_{\mathbb{R}^d}} (\tfrac{\partial}{\partial\vartheta}\theta_t^\vartheta)h - \tfrac{1}{\|h\|_{\mathbb{R}^d}}(\theta_t^{\vartheta+h} - \theta_t^\vartheta) \right\|_{\mathbb{R}^d} + \limsup_{\substack{h\to 0\\ h\in\mathbb{R}^d\setminus\{0\}}} \left\| \tfrac{1}{\|h\|_{\mathbb{R}^d}}(\theta_t^{\vartheta+h} - \theta_t^\vartheta) \right\|_{\mathbb{R}^d} \\
&= \limsup_{\substack{h\to 0\\ h\in\mathbb{R}^d\setminus\{0\}}} \left\| \tfrac{1}{\|h\|_{\mathbb{R}^d}}(\theta_t^{\vartheta+h} - \theta_t^\vartheta) \right\|_{\mathbb{R}^d}.
\end{aligned} \tag{3.19}$$

This, (3.18), and Corollary 3.3.3 assure that for all $\vartheta \in \mathbb{R}^d$, $t \in [0, \infty)$ we have that

$$\limsup_{\substack{h \to 0 \\ h \in \mathbb{R}^d \setminus \{0\}}} \left\| \tfrac{1}{\|h\|_{\mathbb{R}^d}} \left(\tfrac{\partial}{\partial \vartheta} \theta_t^\vartheta\right) h \right\|_{\mathbb{R}^d} \leq e^{Lt}. \tag{3.20}$$

This reveals that for all $\vartheta \in \mathbb{R}^d$, $v \in \mathbb{R}^d \setminus \{0\}$, $t \in [0, \infty)$ it holds that

$$\begin{aligned} \|(\tfrac{\partial}{\partial \vartheta} \theta_t^\vartheta) v\|_{\mathbb{R}^d} &= \limsup_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left\| \tfrac{1}{\lambda} (\tfrac{\partial}{\partial \vartheta} \theta_t^\vartheta) \lambda v \right\|_{\mathbb{R}^d} \\ &= \|v\|_{\mathbb{R}^d} \limsup_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left\| \tfrac{1}{\lambda} (\tfrac{\partial}{\partial \vartheta} \theta_t^\vartheta) \lambda \tfrac{v}{\|v\|_{\mathbb{R}^d}} \right\|_{\mathbb{R}^d} \leq \|v\|_{\mathbb{R}^d} \, e^{Lt}. \end{aligned} \tag{3.21}$$

This establishes item (ii). The proof of Lemma 3.3.4 is thus completed. □

Lemma 3.3.5. *Assume Setting 3.1.1. Then we have for all $a, b \in [0, \infty)$, $\vartheta \in \mathbb{R}^d$ that $\theta_b^{(\theta_a^\vartheta)} = \theta_{a+b}^\vartheta$.*

Proof of Lemma 3.3.5. Throughout this proof let $a, b \in [0, \infty)$, $\vartheta \in \mathbb{R}^d$ and let $e\colon [0, b] \to \mathbb{R}$ satisfy for all $s \in [0, b]$ that

$$e(s) = \left\|\theta_{a+s}^\vartheta - \theta_s^{(\theta_a^\vartheta)}\right\|_{\mathbb{R}^d}. \tag{3.22}$$

Observe that the fact that $([0, b] \ni t \mapsto \theta_{a+t}^\vartheta \in \mathbb{R}^d)$ and $([0, b] \ni t \mapsto \theta_t^{(\theta_a^\vartheta)} \in \mathbb{R}^d)$ are continuous implies that there exists a non-empty convex compact set $K \subseteq \mathbb{R}^d$ which satisfies for all $s \in [0, b]$ that

$$\theta_{a+s}^\vartheta \in K \qquad \text{and} \qquad \theta_s^{(\theta_a^\vartheta)} \in K. \tag{3.23}$$

Moreover, note that Lemma 2.2.2 and Lemma 3.2.1 assure that

$$\sup_{x \in K} \|g'(x)\|_{L(\mathbb{R}^d, \mathbb{R}^d)} < \infty. \tag{3.24}$$

In the next step we observe that (3.2) proves that for all $s \in [0, b]$ we have that

$$\begin{aligned} e(s) &= \left\|\theta_{a+s}^\vartheta - \theta_s^{(\theta_a^\vartheta)}\right\|_{\mathbb{R}^d} \\ &= \left\| \vartheta + \int_0^{a+s} g\big(\theta_u^\vartheta\big)\, du - \left(\theta_a^\vartheta + \int_0^s g\big(\theta_u^{(\theta_a^\vartheta)}\big)\, du \right) \right\|_{\mathbb{R}^d} \\ &= \left\| \vartheta + \int_0^{a+s} g\big(\theta_u^\vartheta\big)\, du - \left(\vartheta + \int_0^a g\big(\theta_u^\vartheta\big)\, du + \int_0^s g\big(\theta_u^{(\theta_a^\vartheta)}\big)\, du \right) \right\|_{\mathbb{R}^d} \\ &= \left\| \int_a^{a+s} g\big(\theta_u^\vartheta\big)\, du - \int_0^s g\big(\theta_u^{(\theta_a^\vartheta)}\big)\, du \right\|_{\mathbb{R}^d} \\ &= \left\| \int_0^s g(\theta_{a+u}^\vartheta) - g\big(\theta_u^{(\theta_a^\vartheta)}\big)\, du \right\|_{\mathbb{R}^d}. \end{aligned} \tag{3.25}$$

This, the triangle inequality for the Bochner integral, and the mean value inequality demonstrate that for all $s \in [0, b]$ we have that

$$\begin{aligned} e(s) &\leq \int_0^s \big\| g\big(\theta^{\vartheta}_{a+u}\big) - g\big(\theta^{(\theta^{\vartheta}_a)}_u\big) \big\|_{\mathbb{R}^d} \, du \\ &\leq \int_0^s \left[\sup_{x \in K} \| g'(x) \|_{L(\mathbb{R}^d, \mathbb{R}^d)} \right] \big\| \theta^{\vartheta}_{a+u} - \theta^{(\theta^{\vartheta}_a)}_u \big\|_{\mathbb{R}^d} \, du \\ &= \sup_{x \in K} \| g'(x) \|_{L(\mathbb{R}^d, \mathbb{R}^d)} \int_0^s e(u) \, du. \end{aligned} \tag{3.26}$$

The Gronwall integral inequality in Lemma 2.1.2 and (3.24) hence assure that for all $s \in [0, b]$ we have that

$$e(s) = 0. \tag{3.27}$$

The proof of Lemma 3.3.5 is thus completed. □

Lemma 3.3.6. *Assume Setting 3.1.1 and let $L \in (0, \infty)$ satisfy for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y \rangle_{\mathbb{R}^d} \leq -L \| x - y \|^2_{\mathbb{R}^d}. \tag{3.28}$$

Then

(i) we have that there exists a unique $\Xi \in \mathbb{R}^d$ which satisfies that $g(\Xi) = 0$,

(ii) we have for all $t \in [0, \infty)$ that $\theta^{\Xi}_t = \Xi$,

(iii) we have for all $h \in \mathbb{R}^d$, $t \in [0, \infty)$ that $\| \theta^{\Xi + h}_t - \Xi \|_{\mathbb{R}^d} \leq \| h \|_{\mathbb{R}^d} \exp(-Lt)$, and

(iv) we have for all $x \in \mathbb{R}^d$ that $\limsup_{t \to \infty} \| \theta^x_t - \Xi \|_{\mathbb{R}^d} = 0$.

Proof of Lemma 3.3.6. First, observe that Lemma 3.2.1, Corollary 3.3.3, and (3.28) imply that for all $t \in [0, \infty)$, $x, y \in \mathbb{R}^d$ we have that

$$\| \theta^x_t - \theta^y_t \|_{\mathbb{R}^d} \leq \| x - y \|_{\mathbb{R}^d} e^{-Lt}. \tag{3.29}$$

This and the Banach fixed point theorem demonstrate that there exists a unique function $\beta \colon (0, \infty) \to \mathbb{R}^d$ which satisfies for all $t \in (0, \infty)$ that

$$\theta^{\beta_t}_t = \beta_t. \tag{3.30}$$

Next we claim that for all $n \in \mathbb{N}$, $t \in (0, \infty)$ we have that

$$\beta_{nt} = \beta_t. \tag{3.31}$$

We establish this by induction on $n \in \mathbb{N}$. The base case $n = 1$ is clear. For the induction step $\mathbb{N} \ni n \to n+1 \in \mathbb{N}$ observe that Lemma 3.3.5 and the induction hypothesis imply that for all $t \in (0,\infty)$ we have that

$$\theta^{\beta_t}_{(n+1)t} = \theta^{\beta_t}_{nt+t} = \theta^{\theta^{\beta_t}_t}_{nt} = \theta^{\beta_t}_{nt} = \theta^{\beta_{nt}}_{nt} = \beta_{nt} = \beta_t. \tag{3.32}$$

This finishes the proof of the induction step. Induction hence establishes (3.31). Observe that (3.31) implies that for all $m, n \in \mathbb{N}$, $t \in (0,\infty)$ we have that

$$\beta_{\frac{m}{n}t} = \beta_{n\frac{m}{n}t} = \beta_{mt} = \beta_t. \tag{3.33}$$

This reveals that for all $t \in \mathbb{Q} \cap (0,\infty)$ it holds that

$$\beta_t = \beta_1. \tag{3.34}$$

This proves that for all $t \in \mathbb{Q} \cap (0,\infty)$ we have that

$$\theta^{\beta_1}_t = \theta^{\beta_t}_t = \beta_t = \beta_1. \tag{3.35}$$

Therefore, we obtain that for all $t \in [0,\infty)$, $n \in \mathbb{N}$ and all functions $q = (q_k)_{k\in\mathbb{N}} \colon \mathbb{N} \to \mathbb{Q} \cap (0,\infty)$ with $\limsup_{k\to\infty} |q_k - t| = 0$ we have that

$$\theta^{\beta_1}_{q_n} = \beta_1. \tag{3.36}$$

Moreover, observe that Corollary 2.3.3 and Lemma 3.2.1 assure that

$$\big([0,\infty) \ni t \mapsto \theta^{\beta_1}_t \in \mathbb{R}^d\big) \in C^1([0,\infty),\mathbb{R}^d). \tag{3.37}$$

Combining this and (3.36) proves that for all $t \in [0,\infty)$ we have that

$$\theta^{\beta_1}_t = \beta_1. \tag{3.38}$$

This, (3.2), and (3.37) ensure that for all $t \in [0,\infty)$ we have that

$$0 = \tfrac{\partial}{\partial t}(\theta^{\beta_1}_t) = g(\theta^{\beta_1}_t) = g(\beta_1). \tag{3.39}$$

Combining this and (3.28) implies that for all $x \in \{y \in \mathbb{R}^d \colon g(y) = 0\}$ we have that

$$0 = \langle 0, x - \beta_1\rangle_{\mathbb{R}^d} = \langle g(x) - g(\beta_1), x - \beta_1\rangle_{\mathbb{R}^d} \leq -L\|x - \beta_1\|^2_{\mathbb{R}^d}. \tag{3.40}$$

The assumption that $L > 0$ and (3.39) therefore prove item (i). Moreover, note that (3.38) establishes item (ii). Corollary 3.3.3 hence demonstrates that for all $h \in \mathbb{R}^d$, $t \in [0,\infty)$ we have that

$$\|\theta^{\beta_1+h}_t - \beta_1\|_{\mathbb{R}^d} = \|\theta^{\beta_1+h}_t - \theta^{\beta_1}_t\|_{\mathbb{R}^d} \leq \|h\|_{\mathbb{R}^d} e^{-Lt}. \tag{3.41}$$

This establishes item (iii). Next observe that item (iii) implies item (iv). The proof of Lemma 3.3.6 is thus completed. □

Lemma 3.3.7. *Assume Setting 3.1.1, let $L \in \mathbb{R}$, $\psi \in C^1(\mathbb{R}^d, \mathbb{R})$, and assume for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y \rangle_{\mathbb{R}^d} \le L \|x - y\|_{\mathbb{R}^d}^2. \tag{3.42}$$

Then we have for all $t \in [0, \infty)$, $x, y \in \mathbb{R}^d$ that

$$\begin{aligned} |\psi(\theta_t^x) - \psi(\theta_t^y)| \le \sup\bigl\{ \|\psi'(\lambda\theta_t^x + (1-\lambda)\theta_t^y)\|_{L(\mathbb{R}^d,\mathbb{R})} \in \mathbb{R} \colon \lambda \in [0,1] \bigr\} \\ \cdot \|x - y\|_{\mathbb{R}^d} \exp(Lt). \end{aligned} \tag{3.43}$$

Proof of Lemma 3.3.7. Throughout this proof let $M \colon [0, \infty) \times \mathbb{R}^d \times \mathbb{R}^d \to [0, \infty]$ satisfy for all $t \in [0, \infty)$, $x, y \in \mathbb{R}^d$ that

$$M(t, x, y) = \sup\bigl\{ \|\psi'(\lambda\theta_t^x + (1-\lambda)\theta_t^y)\|_{L(\mathbb{R}^d,\mathbb{R})} \in \mathbb{R} \colon \lambda \in [0,1] \bigr\}. \tag{3.44}$$

Note that the fundamental theorem of calculus, Lemma 3.2.1, and Corollary 3.3.3 assure that for all $t \in [0, \infty)$, $x, y \in \mathbb{R}^d$ we have that

$$\begin{aligned} |\psi(\theta_t^x) - \psi(\theta_t^y)| &= \left| \int_0^1 \psi'(\lambda\theta_t^x + (1-\lambda)\theta_t^y)(\theta_t^x - \theta_t^y)\, d\lambda \right| \\ &\le \int_0^1 |\psi'(\lambda\theta_t^x + (1-\lambda)\theta_t^y)(\theta_t^x - \theta_t^y)|\, d\lambda \\ &\le \int_0^1 \|\psi'(\lambda\theta_t^x + (1-\lambda)\theta_t^y)\|_{L(\mathbb{R}^d,\mathbb{R})} \|\theta_t^x - \theta_t^y\|_{\mathbb{R}^d}\, d\lambda \\ &\le M(t,x,y)\|\theta_t^x - \theta_t^y\|_{\mathbb{R}^d} \le M(t,x,y)\|x - y\|_{\mathbb{R}^d} \exp(Lt). \end{aligned} \tag{3.45}$$

The proof of Lemma 3.3.7 is thus completed. □

Lemma 3.3.8. *Assume Setting 3.1.1, let $\psi \in C^1(\mathbb{R}^d, \mathbb{R})$, $T \in (0, \infty)$, and let $u = (u(t, \vartheta))_{(t,\vartheta) \in [0,T] \times \mathbb{R}^d} \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0, T]$, $\vartheta \in \mathbb{R}^d$ that $u(t, \vartheta) = \psi(\theta_{T-t}^\vartheta)$. Then*

(i) we have that $u \in C^1([0, T] \times \mathbb{R}^d, \mathbb{R})$ and

(ii) we have for all $t \in [0, T]$, $\vartheta \in \mathbb{R}^d$ that $(\frac{\partial}{\partial t} u)(t, \vartheta) = -(\frac{\partial}{\partial \vartheta} u)(t, \vartheta) g(\vartheta)$.

Proof of Lemma 3.3.8. Throughout this proof let $v \colon [0, T] \times \mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0, T]$, $\vartheta \in \mathbb{R}^d$ that

$$v(t, \vartheta) = \psi(\theta_t^\vartheta). \tag{3.46}$$

Note that combining Lemma 3.2.1 and item (i) in Proposition 2.4.1 proves item (i). Next observe that item (iii) in Proposition 2.4.1 assures that for all $t \in [0, T]$, $\vartheta \in \mathbb{R}^d$ we have that

$$(\tfrac{\partial}{\partial t} v)(t, \vartheta) = (\tfrac{\partial}{\partial \vartheta} v)(t, \vartheta) g(\vartheta). \tag{3.47}$$

This reveals that for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ it holds that

$$(\tfrac{\partial}{\partial t}u)(t,\vartheta) = -(\tfrac{\partial}{\partial t}v)(T-t,\vartheta) = -(\tfrac{\partial}{\partial \vartheta}v)(T-t,\vartheta)g(\vartheta) = -(\tfrac{\partial}{\partial \vartheta}u)(t,\vartheta)g(\vartheta). \tag{3.48}$$

This establishes item (ii). The proof of Lemma 3.3.8 is thus completed. □

Lemma 3.3.9. *Assume Setting 3.1.1, let $\psi \in C^1(\mathbb{R}^d,\mathbb{R})$, $T \in (0,\infty)$, $L \in \mathbb{R}$, assume for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y\rangle_{\mathbb{R}^d} \le L\|x-y\|^2_{\mathbb{R}^d}, \tag{3.49}$$

and let $u = (u(t,\vartheta))_{(t,\vartheta)\in[0,T]\times\mathbb{R}^d} \colon [0,T]\times\mathbb{R}^d \to \mathbb{R}$ satisfy for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ that $u(t,\vartheta) = \psi(\theta^\vartheta_{T-t})$. Then

(i) we have that $u \in C^1([0,T]\times\mathbb{R}^d,\mathbb{R})$ and

(ii) we have for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ that

$$\|(\tfrac{\partial}{\partial \vartheta}u)(t,\vartheta)\|_{L(\mathbb{R}^d,\mathbb{R})} \le \|\psi'(\theta^\vartheta_{T-t})\|_{L(\mathbb{R}^d,\mathbb{R})}\exp(L(T-t)). \tag{3.50}$$

Proof of Lemma 3.3.9. First, note that item (i) in Lemma 3.3.8 proves item (i). Next observe that item (ii) in Lemma 3.3.4 implies that for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ we have that

$$\begin{aligned}\|(\tfrac{\partial}{\partial \vartheta}u)(t,\vartheta)\|_{L(\mathbb{R}^d,\mathbb{R})} &= \|\psi'(\theta^\vartheta_{T-t})(\tfrac{\partial}{\partial \vartheta}\theta^\vartheta_{T-t})\|_{L(\mathbb{R}^d,\mathbb{R})}\\ &\le \|\psi'(\theta^\vartheta_{T-t})\|_{L(\mathbb{R}^d,\mathbb{R})}\|(\tfrac{\partial}{\partial \vartheta}\theta^\vartheta_{T-t})\|_{L(\mathbb{R}^d,\mathbb{R}^d)}\\ &\le \|\psi'(\theta^\vartheta_{T-t})\|_{L(\mathbb{R}^d,\mathbb{R})}e^{L(T-t)}.\end{aligned} \tag{3.51}$$

This establishes item (ii). The proof of Lemma 3.3.9 is thus completed. □

3.4 Upper bounds for second-order spatial derivatives of certain deterministic flows

Lemma 3.4.1. *Let $d \in \mathbb{N}$, $g \in C^1(\mathbb{R}^d,\mathbb{R}^d)$, $L \in \mathbb{R}$, and assume for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y\rangle_{\mathbb{R}^d} \le L\|x-y\|^2_{\mathbb{R}^d}. \tag{3.52}$$

Then we have for all $x, v \in \mathbb{R}^d$ that $\langle g'(v)x, x\rangle_{\mathbb{R}^d} \le L\|x\|^2_{\mathbb{R}^d}$.

Proof of Lemma 3.4.1. Throughout this proof let $x, v \in \mathbb{R}^d$. Note that

$$\limsup_{\substack{h \to 0 \\ h \in \mathbb{R}^d \setminus \{0\}}} \left\| \frac{1}{\|h\|_{\mathbb{R}^d}} (g(v+h) - g(v) - g'(v)h) \right\|_{\mathbb{R}^d} = 0. \tag{3.53}$$

This demonstrates that

$$\limsup_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left\| \frac{1}{\lambda} (g(v+\lambda x) - g(v) - \lambda g'(v)x) \right\|_{\mathbb{R}^d} = 0. \tag{3.54}$$

This reveals that

$$\lim_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left[\frac{1}{\lambda} (g(v+\lambda x) - g(v)) \right] = g'(v)x. \tag{3.55}$$

This and (3.52) assure that

$$\begin{aligned} &\langle g'(v)x, x \rangle_{\mathbb{R}^d} \\ &= \left\langle \lim_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left[\frac{1}{\lambda} (g(v+\lambda x) - g(v)) \right], x \right\rangle_{\mathbb{R}^d} = \lim_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left[\frac{1}{\lambda} \langle g(v+\lambda x) - g(v), x \rangle_{\mathbb{R}^d} \right] \\ &= \lim_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left[\frac{1}{\lambda^2} \langle g(v+\lambda x) - g(v), \lambda x \rangle_{\mathbb{R}^d} \right] \leq \limsup_{\substack{\lambda \to 0 \\ \lambda \in \mathbb{R} \setminus \{0\}}} \left[\frac{1}{\lambda^2} L \, \|\lambda x\|_{\mathbb{R}^d}^2 \right] = L \|x\|_{\mathbb{R}^d}^2. \end{aligned} \tag{3.56}$$

The proof of Lemma 3.4.1 is thus completed. □

Lemma 3.4.2. *Let $t, L \in \mathbb{R}$, $T \in (t, \infty)$, $d \in \mathbb{N}$, $b \in C([t,T], \mathbb{R}^d)$, let $A \in C([t,T], L(\mathbb{R}^d, \mathbb{R}^d))$ satisfy for all $s \in [t,T]$, $u \in \mathbb{R}^d$ that*

$$\langle A(s)u, u \rangle_{\mathbb{R}^d} \leq L \, \|u\|_{\mathbb{R}^d}^2 , \tag{3.57}$$

and let $y_1, y_2 \in C^1([t,T], \mathbb{R}^d)$ satisfy for all $i \in \{1,2\}$, $s \in [t,T]$ that

$$y_i(t) = 0 \qquad \text{and} \qquad (y_i)'(s) = A(s)y_i(s) + b(s). \tag{3.58}$$

Then we have that $y_1 = y_2$.

Proof of Lemma 3.4.2. Throughout this proof let $\varphi \in C^1([t,T], \mathbb{R})$ satisfy for all $s \in [t,T]$ that

$$\varphi(s) = \|y_1(s) - y_2(s)\|_{\mathbb{R}^d}^2 . \tag{3.59}$$

Observe that (3.59), (3.58), and (3.57) imply that for all $s \in [t, T]$ we have that

$$\begin{aligned}\varphi'(s) &= 2\langle y_1(s) - y_2(s), (y_1)'(s) - (y_2)'(s)\rangle_{\mathbb{R}^d} \\ &= 2\langle y_1(s) - y_2(s), A(s)(y_1(s) - y_2(s))\rangle_{\mathbb{R}^d} \\ &\leq 2L \, \|y_1(s) - y_2(s)\|_{\mathbb{R}^d}^2 = 2L\varphi(s).\end{aligned} \tag{3.60}$$

This and the Gronwall differential inequality in Lemma 2.1.1 prove that for all $s \in [t, T]$ we have that

$$\varphi(s) \leq \varphi(t) e^{2L(s-t)}. \tag{3.61}$$

This and (3.58) assure that for all $s \in [t, T]$ we have that $\varphi(s) = 0$. The proof of Lemma 3.4.2 is thus completed. □

Lemma 3.4.3. *Assume Setting 3.1.1. Then*

(i) we have that $g \in C^2(\mathbb{R}^d, \mathbb{R}^d)$,

(ii) we have that

$$\big([0, \infty) \times \mathbb{R}^d \ni (t, \vartheta) \mapsto \theta_t^\vartheta \in \mathbb{R}^d\big) \in C^2([0, \infty) \times \mathbb{R}^d, \mathbb{R}^d), \tag{3.62}$$

(iii) we have for all $\vartheta \in \mathbb{R}^d$ *that*

$$\big([0, \infty) \ni t \mapsto \tfrac{\partial^2}{\partial \vartheta^2} \theta_t^\vartheta \in L^{(2)}(\mathbb{R}^d, \mathbb{R}^d)\big) \in C^1([0, \infty), L^{(2)}(\mathbb{R}^d, \mathbb{R}^d)), \tag{3.63}$$

and

(iv) we have for all $t \in [0, \infty)$, $\vartheta \in \mathbb{R}^d$ *that*

$$\tfrac{\partial}{\partial t}\big(\tfrac{\partial^2}{\partial \vartheta^2} \theta_t^\vartheta\big) = \tfrac{\partial^2}{\partial \vartheta^2}(g(\theta_t^\vartheta)). \tag{3.64}$$

Proof of Lemma 3.4.3. First, note that Lemma 3.2.2 proves item (i). This and Corollary 2.3.3 demonstrate item (ii). Next observe that for all $t \in [0, \infty)$, $\vartheta \in \mathbb{R}^d$ we have that

$$\tfrac{\partial}{\partial t} \theta_t^\vartheta = g(\theta_t^\vartheta). \tag{3.65}$$

This, item (i), and item (ii) ensure that

$$\big([0, \infty) \times \mathbb{R}^d \ni (t, \vartheta) \mapsto \tfrac{\partial}{\partial t} \theta_t^\vartheta \in \mathbb{R}^d\big) \in C^2([0, \infty) \times \mathbb{R}^d, \mathbb{R}^d). \tag{3.66}$$

This reveals that

$$\begin{aligned}&\big([0, \infty) \times \mathbb{R}^d \ni (t, \vartheta) \mapsto \tfrac{\partial^2}{\partial \vartheta^2}\big(\tfrac{\partial}{\partial t} \theta_t^\vartheta\big) \in L^{(2)}(\mathbb{R}^d, \mathbb{R}^d)\big) \\ &\qquad \in C([0, \infty) \times \mathbb{R}^d, L^{(2)}(\mathbb{R}^d, \mathbb{R}^d)).\end{aligned} \tag{3.67}$$

Schwarz's theorem (cf., e.g., Königsberger [61, Section 2.3]) hence proves that for all $t \in [0,\infty)$, $\vartheta \in \mathbb{R}^d$ we have that

$$\tfrac{\partial^2}{\partial \vartheta^2}\big(\tfrac{\partial}{\partial t}\theta_t^\vartheta\big) = \tfrac{\partial}{\partial \vartheta}\big(\tfrac{\partial^2}{\partial \vartheta \partial t}\theta_t^\vartheta\big) = \tfrac{\partial}{\partial \vartheta}\big(\tfrac{\partial^2}{\partial t \partial \vartheta}\theta_t^\vartheta\big) = \tfrac{\partial^2}{\partial \vartheta \partial t}\big(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta\big). \tag{3.68}$$

This and (3.67) assure that

$$\begin{aligned}\big([0,\infty) \times \mathbb{R}^d \ni (t,\vartheta) \mapsto \tfrac{\partial^2}{\partial \vartheta \partial t}\big(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta\big) \in L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)\big)\\ \in C([0,\infty) \times \mathbb{R}^d, L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)).\end{aligned} \tag{3.69}$$

Schwarz's theorem (cf., e.g., Königsberger [61, Section 2.3]) therefore implies that for all $t \in [0,\infty)$, $\vartheta \in \mathbb{R}^d$ we have that $\frac{\partial^2}{\partial t \partial \vartheta}\big(\frac{\partial}{\partial \vartheta}\theta_t^\vartheta\big)$ exists and

$$\tfrac{\partial^2}{\partial t \partial \vartheta}\big(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta\big) = \tfrac{\partial^2}{\partial \vartheta \partial t}\big(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta\big). \tag{3.70}$$

This and (3.68) establish item (iii) and that for all $t \in [0,\infty)$, $\vartheta \in \mathbb{R}^d$ we have that

$$\tfrac{\partial^2}{\partial \vartheta^2}\big(\tfrac{\partial}{\partial t}\theta_t^\vartheta\big) = \tfrac{\partial^2}{\partial t \partial \vartheta}\big(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta\big) = \tfrac{\partial}{\partial t}\big(\tfrac{\partial^2}{\partial \vartheta^2}\theta_t^\vartheta\big). \tag{3.71}$$

Combining this with (3.65) establishes item (iv). The proof of Lemma 3.4.3 is thus completed. □

Lemma 3.4.4. *Assume Setting 3.1.1, let $T \in (0,\infty)$, $L \in \mathbb{R}$, and assume for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y\rangle_{\mathbb{R}^d} \le L\|x - y\|_{\mathbb{R}^d}^2. \tag{3.72}$$

Then

(i) we have that $g \in C^2(\mathbb{R}^d,\mathbb{R}^d)$,

(ii) we have that $([0,\infty) \times \mathbb{R}^d \ni (t,\vartheta) \mapsto \theta_t^\vartheta \in \mathbb{R}^d) \in C^2([0,\infty) \times \mathbb{R}^d,\mathbb{R}^d)$,

(iii) we have that there exist unique $\chi^\vartheta \in C(\{(s,t) \in [0,T]^2 \colon s \le t\} \times \mathbb{R}^d,\mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, which satisfy for all $\vartheta \in \mathbb{R}^d$, $s \in [0,T]$, $t \in [s,T]$, $x \in \mathbb{R}^d$ that

$$\chi^\vartheta(s,t,x) = x + \int_s^t g'(\theta_u^\vartheta)\,\chi^\vartheta(s,u,x)\,du, \tag{3.73}$$

and

(iv) we have for all $t \in [0,T]$, $\vartheta, v, w \in \mathbb{R}^d$ that

$$\big(\tfrac{\partial^2}{\partial \vartheta^2}\theta_t^\vartheta\big)(v,w) = \int_0^t \chi^\vartheta\big(s,t,g''(\theta_s^\vartheta)\big((\tfrac{\partial}{\partial \vartheta}\theta_s^\vartheta)v,(\tfrac{\partial}{\partial \vartheta}\theta_s^\vartheta)w\big)\big)\,ds. \tag{3.74}$$

Proof of Lemma 3.4.4. First, observe that Lemma 3.4.3 proves item (i) and item (ii). Next note that item (i), item (ii), the fact that the set $[0,T] \subseteq \mathbb{R}$ is compact, and Lemma 2.2.2 assure that there exists $c\colon \mathbb{R}^d \to (0,\infty)$ which satisfies for all $\vartheta \in \mathbb{R}^d$, $t \in [0,T]$ that

$$\|g'(\theta_t^\vartheta)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} \leq c_\vartheta. \tag{3.75}$$

This proves that for all $\vartheta \in \mathbb{R}^d$, $x, y \in \mathbb{R}^d$, $s \in [0,T]$ we have that

$$\|g'(\theta_s^\vartheta)x - g'(\theta_s^\vartheta)y\|_{\mathbb{R}^d} \leq c_\vartheta \|x-y\|_{\mathbb{R}^d}. \tag{3.76}$$

Proposition 2.3.1 hence ensures that there exist unique $\chi^\vartheta \in C(\{(s,t) \in [0,T]^2 \colon s \leq t\} \times \mathbb{R}^d, \mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, which satisfy for all $\vartheta \in \mathbb{R}^d$, $s \in [0,T]$, $t \in [s,T]$, $x \in \mathbb{R}^d$ that

$$\chi^\vartheta(s,t,x) = x + \int_s^t g'(\theta_u^\vartheta)\, \chi^\vartheta(s,u,x)\, du. \tag{3.77}$$

This proves item (iii). In the next step let $\vartheta, v, w \in \mathbb{R}^d$, let $A \in C([0,T], L(\mathbb{R}^d,\mathbb{R}^d))$ satisfy for all $t \in [0,T]$ that

$$A(t) = g'(\theta_t^\vartheta), \tag{3.78}$$

let $b \in C([0,T],\mathbb{R}^d)$ satisfy for all $t \in [0,T]$ that

$$b(t) = g''(\theta_t^\vartheta)\big((\tfrac{\partial}{\partial\vartheta}\theta_t^\vartheta)v, (\tfrac{\partial}{\partial\vartheta}\theta_t^\vartheta)w\big), \tag{3.79}$$

let $y \in C^1([0,T],\mathbb{R}^d)$ satisfy for all $t \in [0,T]$ that

$$y(t) = (\tfrac{\partial^2}{\partial\vartheta^2}\theta_t^\vartheta)(v,w) \tag{3.80}$$

(cf. Lemma 3.4.3), and let $z \in C^1([0,T],\mathbb{R}^d)$ satisfy for all $t \in [0,T]$ that

$$z(t) = \int_0^t \chi^\vartheta\big(s,t,b(s)\big)\, ds. \tag{3.81}$$

Note that (3.80) and (3.2) imply that

$$y(0) = (\tfrac{\partial^2}{\partial\vartheta^2}\theta_0^\vartheta)(v,w) = (\tfrac{\partial^2}{\partial\vartheta^2}\vartheta)(v,w) = 0. \tag{3.82}$$

Moreover, observe that (3.80), Lemma 3.4.3, and the chain rule ensure that for all

$t \in [0, T]$ we have that

$$\begin{aligned} y'(t) &= \tfrac{\partial}{\partial t}\big((\tfrac{\partial^2}{\partial \vartheta^2}\theta_t^\vartheta)(v, w)\big) \\ &= \big(\tfrac{\partial^2}{\partial \vartheta^2}(g(\theta_t^\vartheta))\big)(v, w) \\ &= \big(\tfrac{\partial}{\partial \vartheta}(g'(\theta_t^\vartheta)(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta))\big)(v)(w) \\ &= \big(\tfrac{\partial}{\partial \vartheta}(g'(\theta_t^\vartheta)(\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)v)\big)(w) \\ &= g''(\theta_t^\vartheta)\big((\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)v, (\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)w\big) + g'(\theta_t^\vartheta)\big(\tfrac{\partial}{\partial \vartheta}\big((\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)v\big)(w)\big) \\ &= g''(\theta_t^\vartheta)\big((\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)v, (\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)w\big) + g'(\theta_t^\vartheta)\big((\tfrac{\partial^2}{\partial \vartheta^2}\theta_t^\vartheta)(v, w)\big) \\ &= g'(\theta_t^\vartheta)\big((\tfrac{\partial^2}{\partial \vartheta^2}\theta_t^\vartheta)(v, w)\big) + g''(\theta_t^\vartheta)\big((\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)v, (\tfrac{\partial}{\partial \vartheta}\theta_t^\vartheta)w\big) \\ &= A(t)y(t) + b(t). \end{aligned} \tag{3.83}$$

In addition, note that (3.81) assures that

$$z(0) = 0. \tag{3.84}$$

In the next step we combine (3.81), (3.78), and (3.77) to obtain that for all $t \in [0, T]$ we have that

$$\begin{aligned} z'(t) &= \chi(t, t, b(t)) + \int_0^t A(t)\chi^\vartheta\big(s, t, b(s)\big)\, ds \\ &= b(t) + A(t)z(t) = A(t)z(t) + b(t). \end{aligned} \tag{3.85}$$

Furthermore, observe that Lemma 3.4.1 and (3.72) imply that for all $t \in [0, T]$, $u \in \mathbb{R}^d$ we have that

$$\langle A(t)u, u\rangle_{\mathbb{R}^d} \leq L\|u\|_{\mathbb{R}^d}^2. \tag{3.86}$$

Combining this, (3.82)–(3.85), and Lemma 3.4.2 demonstrates that for all $t \in [0, T]$ we have that

$$y(t) = z(t). \tag{3.87}$$

This establishes item (iv). The proof of Lemma 3.4.4 is thus completed. □

Lemma 3.4.5. *Let $d \in \mathbb{N}$, $a, L \in \mathbb{R}$, $b \in (a, \infty)$, let $f \in C([a, b] \times \mathbb{R}^d, \mathbb{R}^d)$ satisfy for all $x, y \in \mathbb{R}^d$, $s \in [a, b]$ that*

$$\langle f(s, x) - f(s, y), x - y\rangle_{\mathbb{R}^d} \leq L\|x - y\|_{\mathbb{R}^d}^2, \tag{3.88}$$

and let $\chi_{t,\cdot}^x \in C([t, b], \mathbb{R}^d)$, $x \in \mathbb{R}^d$, $t \in [a, b]$, satisfy for all $t \in [a, b]$, $x \in \mathbb{R}^d$, $s \in [t, b]$ that

$$\chi_{t,s}^x = x + \int_t^s f(u, \chi_{t,u}^x)\, du. \tag{3.89}$$

Then we have for all $x, y \in \mathbb{R}^d$, $t \in [a, b]$, $s \in [t, b]$ that

$$\|\chi^x_{t,s} - \chi^y_{t,s}\|_{\mathbb{R}^d} \le \|x - y\|_{\mathbb{R}^d} e^{L(s-t)}. \tag{3.90}$$

Proof of Lemma 3.4.5. Throughout this proof let $t \in [a, b]$, $x, y \in \mathbb{R}^d$, let $E\colon [t, b] \to \mathbb{R}$ satisfy for all $s \in [t, b]$ that

$$E(s) = \|\chi^x_{t,s} - \chi^y_{t,s}\|^2_{\mathbb{R}^d}. \tag{3.91}$$

Note that (3.91), (3.89), and (3.88) assure that for all $s \in [t, b]$ we have that

$$E'(s) = 2\langle f(s, \chi^x_{t,s}) - f(s, \chi^y_{t,s}), \chi^x_{t,s} - \chi^y_{t,s}\rangle_{\mathbb{R}^d} \le 2L\|\chi^x_{t,s} - \chi^y_{t,s}\|^2_{\mathbb{R}^d} = 2LE(s). \tag{3.92}$$

The Gronwall differential inequality in Lemma 2.1.1 hence implies that for all $s \in [t, b]$ we have that

$$\|\chi^x_{t,s} - \chi^y_{t,s}\|_{\mathbb{R}^d} = |E(s)|^{1/2} \le |E(t)|^{1/2} e^{L(s-t)} = \|x - y\|_{\mathbb{R}^d} e^{L(s-t)}. \tag{3.93}$$

The proof of Lemma 3.4.5 is thus completed. □

Lemma 3.4.6. *Assume Setting 3.1.1, let $T \in (0, \infty)$, $L \in \mathbb{R}$, and assume for all $x, y \in \mathbb{R}^d$ that*

$$\langle g(x) - g(y), x - y\rangle_{\mathbb{R}^d} \le L\|x - y\|^2_{\mathbb{R}^d}. \tag{3.94}$$

Then

(i) we have that $g \in C^2(\mathbb{R}^d, \mathbb{R}^d)$,

(ii) we have that $([0, \infty) \times \mathbb{R}^d \ni (t, \vartheta) \mapsto \theta^\vartheta_t \in \mathbb{R}^d) \in C^2([0, \infty) \times \mathbb{R}^d, \mathbb{R}^d)$, and

(iii) we have for all $\vartheta \in \mathbb{R}^d$, $t \in [0, T]$ that

$$\big\|\tfrac{\partial^2}{\partial\vartheta^2}\theta^\vartheta_t\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} \le \int_0^t \exp(L(t + s))\big\|g''(\theta^\vartheta_s)\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\, ds. \tag{3.95}$$

Proof of Lemma 3.4.6. First, observe that Lemma 3.4.4 proves item (i) and item (ii). Next let $\vartheta, v, w \in \mathbb{R}^d$, let $A\colon [0, T] \to L(\mathbb{R}^d, \mathbb{R}^d)$ satisfy for all $t \in [0, T]$ that

$$A(t) = g'(\theta^\vartheta_t), \tag{3.96}$$

let $b\colon [0, T] \to \mathbb{R}^d$ satisfy for all $t \in [0, T]$ that

$$b(t) = g''(\theta^\vartheta_t)\big((\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_t)v, (\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_t)w\big), \tag{3.97}$$

and let $\chi \in C(\{(s,t) \in [0,T]^2 \colon s \leq t\} \times \mathbb{R}^d, \mathbb{R}^d)$ satisfy for all $s \in [0,T]$, $t \in [s,T]$, $x \in \mathbb{R}^d$ that

$$\chi(s,t,x) = x + \int_s^t A(u)\chi(s,u,x)\,du \tag{3.98}$$

(cf. Lemma 3.4.4). Note that Lemma 3.4.1 implies that for all $x, y \in \mathbb{R}^d$, $t \in [0,T]$ we have that

$$\langle A(t)x - A(t)y, x - y\rangle_{\mathbb{R}^d} \leq L\|x-y\|^2_{\mathbb{R}^d}. \tag{3.99}$$

Furthermore, observe that item (iii) in Lemma 3.4.4 assures that for all $s \in [0,T]$, $t \in [s,T]$ we have that

$$\chi(s,t,0) = 0. \tag{3.100}$$

Combining this, (3.99), (3.98), and Lemma 3.4.5 proves that for all $s \in [0,T]$, $t \in [s,T]$ we have that

$$\begin{aligned} \|\chi(s,t,b(s))\|_{\mathbb{R}^d} &= \|\chi(s,t,b(s)) - 0\|_{\mathbb{R}^d} \\ &= \|\chi(s,t,b(s)) - \chi(s,t,0)\|_{\mathbb{R}^d} \leq e^{L(t-s)}\|b(s)\|_{\mathbb{R}^d}. \end{aligned} \tag{3.101}$$

This, Lemma 3.4.4, and the triangle inequality for the Bochner integral ensure that for all $t \in [0,T]$ we have that

$$\begin{aligned} \|(\tfrac{\partial^2}{\partial\vartheta^2}\theta^\vartheta_t)(v,w)\|_{\mathbb{R}^d} &= \left\| \int_0^t \chi\big(s,t,g''(\theta^\vartheta_s)\big((\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_s)v, (\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_s)w\big)\big)\,ds \right\|_{\mathbb{R}^d} \\ &= \left\| \int_0^t \chi(s,t,b(s))\,ds \right\|_{\mathbb{R}^d} \leq \int_0^t \left\|\chi(s,t,b(s))\right\|_{\mathbb{R}^d} ds \\ &\leq \int_0^t e^{L(t-s)}\left\|b(s)\right\|_{\mathbb{R}^d}\,ds. \end{aligned} \tag{3.102}$$

Next observe that for all $s \in [0,T]$ we have that

$$\begin{aligned} \|b(s)\|_{\mathbb{R}^d} &= \left\|g''(\theta^\vartheta_s)((\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_s)v, (\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_s)w)\right\|_{\mathbb{R}^d} \\ &\leq \|(\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_s)v\|_{\mathbb{R}^d}\|(\tfrac{\partial}{\partial\vartheta}\theta^\vartheta_s)w\|_{\mathbb{R}^d}\|g''(\theta^\vartheta_s)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}. \end{aligned} \tag{3.103}$$

Lemma 3.3.4 hence implies that for all $s \in [0,T]$ we have that

$$\|b(s)\|_{\mathbb{R}^d} \leq \|v\|_{\mathbb{R}^d}\,\|w\|_{\mathbb{R}^d}\,\|g''(\theta^\vartheta_s)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}e^{2Ls}. \tag{3.104}$$

Combining this and (3.102) proves that for all $t \in [0,T]$ we have that

$$\begin{aligned} \|(\tfrac{\partial^2}{\partial\vartheta^2}\theta^\vartheta_t)(v,w)\|_{\mathbb{R}^d} &\leq \|v\|_{\mathbb{R}^d}\,\|w\|_{\mathbb{R}^d} \int_0^t e^{L(t-s)}e^{2Ls}\|g''(\theta^\vartheta_s)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,ds \\ &= \|v\|_{\mathbb{R}^d}\,\|w\|_{\mathbb{R}^d} \int_0^t e^{L(t+s)}\|g''(\theta^\vartheta_s)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,ds. \end{aligned} \tag{3.105}$$

This establishes item (iii). The proof of Lemma 3.4.6 is thus completed. □

3.5 Temporal regularity results for SAAs in the case of general learning rates

Lemma 3.5.1. *Assume Setting 3.1.1 and let $T \in (0, \infty)$. Then*

(i) we have for all $\omega \in \Omega$, $t \in [0, T)$ with $\gamma(t) = \emptyset$ that $[0, T] \ni u \mapsto \Theta_u(\omega) \in \mathbb{R}^d$ is differentiable at t and

(ii) we have for all $\omega \in \Omega$, $t \in [0, T)$ with $\gamma(t) = \emptyset$ that

$$\tfrac{\partial}{\partial t}\Theta_t(\omega) = \tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}}\Big[\textstyle\sum_{j \in \gamma(\llbracket t \rrbracket)} G(\Theta_{\llbracket t \rrbracket}(\omega), Z_j(\omega))\Big]. \tag{3.106}$$

Proof of Lemma 3.5.1. Combining the assumption that

$$\forall\, t \in [0, \infty) \colon 0 < \#_{\{s \in [0,t] \colon \gamma(s) \neq \emptyset\}} < \infty \tag{3.107}$$

with (3.4) establishes item (i) and item (ii). The proof of Lemma 3.5.1 is thus completed. □

3.6 A priori estimates for SAAs in the case of general learning rates

Lemma 3.6.1. *Let $n \in \mathbb{N}$, $p \in \{0\} \cup [1, \infty)$, $x_1, x_2, \ldots, x_n \in \mathbb{R}$. Then we have that*

$$\left|\sum_{i=1}^{n} x_i\right|^p \le n^{p-1}\left[\sum_{i=1}^{n} |x_i|^p\right]. \tag{3.108}$$

Proof of Lemma 3.6.1. Throughout this proof assume w.l.o.g. that $p \ge 1$. Observe that the triangle inequality and Hölder's inequality imply that

$$\left|\sum_{i=1}^{n} x_i\right| \le \sum_{i=1}^{n} |x_i| \le \left|\sum_{i=1}^{n} 1\right|^{\frac{p-1}{p}} \left|\sum_{i=1}^{n} |x_i|^p\right|^{\frac{1}{p}} = n^{\frac{p-1}{p}} \left|\sum_{i=1}^{n} |x_i|^p\right|^{\frac{1}{p}}. \tag{3.109}$$

This establishes (3.108). The proof of Lemma 3.6.1 is thus completed. □

Lemma 3.6.2. *Assume Setting 3.1.1, assume for all $v, w \in [0,\infty)$ with $v \neq w$ that $\gamma(v) \cap \gamma(w) = \emptyset$, and let $c \in [0,\infty)$, $p \in [1,\infty)$, $\mathfrak{m} \in \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that*

$$\mathbb{E}\big[\|G(x, Z_1)\|_{\mathbb{R}^d}^p\big] \leq c(1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{m}p}). \tag{3.110}$$

Then we have for all $t \in [0,\infty)$ that

$$\mathbb{E}\Big[\big\|\tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}} \textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} G(\Theta_{\llbracket t \rrbracket}, Z_j)\big\|_{\mathbb{R}^d}^p\Big] \leq c\big(1 + \mathbb{E}\big[\|\Theta_{\llbracket t \rrbracket}\|_{\mathbb{R}^d}^{\mathfrak{m}p}\big]\big). \tag{3.111}$$

Proof of Lemma 3.6.2. Throughout this proof let $t \in [0,\infty)$, $\mathfrak{j} \in \gamma(\llbracket t \rrbracket)$. Note that Lemma 3.6.1 and the triangle inequality assure that

$$\begin{aligned}
&\mathbb{E}\Big[\big\|\tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}} \textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} G(\Theta_{\llbracket t \rrbracket}, Z_j)\big\|_{\mathbb{R}^d}^p\Big] \\
&\leq \tfrac{1}{|\#_{\gamma(\llbracket t \rrbracket)}|^p}\mathbb{E}\Big[\Big(\textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} \|G(\Theta_{\llbracket t \rrbracket}, Z_j)\|_{\mathbb{R}^d}\Big)^p\Big] \\
&\leq \tfrac{1}{|\#_{\gamma(\llbracket t \rrbracket)}|^p}\mathbb{E}\Big[|\#_{\gamma(\llbracket t \rrbracket)}|^{p-1} \textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} \|G(\Theta_{\llbracket t \rrbracket}, Z_j)\|_{\mathbb{R}^d}^p\Big] \\
&= \tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}}\mathbb{E}\Big[\textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} \|G(\Theta_{\llbracket t \rrbracket}, Z_j)\|_{\mathbb{R}^d}^p\Big].
\end{aligned} \tag{3.112}$$

Moreover, observe that combining the assumption that Z_j, $j \in \mathbb{N}$, are i.i.d. random variables, the assumption that $\forall\, (v,w) \in \{(a,b) \in [0,\infty)^2 \colon a \neq b\} \colon \gamma(v) \cap \gamma(w) = \emptyset$, and (3.4) proves that for all $j \in \gamma(\llbracket t \rrbracket)$ we have that Z_j and $\Theta_{\llbracket t \rrbracket}$ are independent. This, (3.112), the assumption that $\mathfrak{j} \in \gamma(\llbracket t \rrbracket)$, and the assumption that Z_j, $j \in \mathbb{N}$, are i.i.d. random variables ensure that

$$\begin{aligned}
&\mathbb{E}\Big[\Big\|\tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}} \textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} G(\Theta_{\llbracket t \rrbracket}, Z_j)\Big\|_{\mathbb{R}^d}^p\Big] \\
&\leq \tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}} \textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} \mathbb{E}\big[\|G(\Theta_{\llbracket t \rrbracket}, Z_{\mathfrak{j}})\|_{\mathbb{R}^d}^p\big] \\
&= \mathbb{E}\big[\|G(\Theta_{\llbracket t \rrbracket}, Z_{\mathfrak{j}})\|_{\mathbb{R}^d}^p\big] \\
&= \int_\Omega \|G(\Theta_{\llbracket t \rrbracket}(\omega), Z_{\mathfrak{j}}(\omega))\|_{\mathbb{R}^d}^p \,\mathbb{P}(d\omega) \\
&= \int_\Omega \int_\Omega \|G(\Theta_{\llbracket t \rrbracket}(\omega), Z_{\mathfrak{j}}(\tilde{\omega}))\|_{\mathbb{R}^d}^p \,\mathbb{P}(d\tilde{\omega})\,\mathbb{P}(d\omega).
\end{aligned} \tag{3.113}$$

Combining this and (3.110) demonstrates that

$$\begin{aligned}
\mathbb{E}\Big[\Big\|\tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}} \textstyle\sum_{j\in\gamma(\llbracket t \rrbracket)} G(\Theta_{\llbracket t \rrbracket}, Z_j)\Big\|_{\mathbb{R}^d}^p\Big] &\leq \int_\Omega c(1 + \|\Theta_{\llbracket t \rrbracket}(\omega)\|_{\mathbb{R}^d}^{\mathfrak{m}p})\,\mathbb{P}(d\omega) \\
&= \mathbb{E}\big[c(1 + \|\Theta_{\llbracket t \rrbracket}\|_{\mathbb{R}^d}^{\mathfrak{m}p})\big] \\
&= c\big(1 + \mathbb{E}\big[\|\Theta_{\llbracket t \rrbracket}\|_{\mathbb{R}^d}^{\mathfrak{m}p}\big]\big).
\end{aligned} \tag{3.114}$$

This establishes (3.111). The proof of Lemma 3.6.2 is thus completed. □

Lemma 3.6.3. *Assume Setting 3.1.1, assume for all $v, w \in [0, \infty)$ with $v \neq w$ that $\gamma(v) \cap \gamma(w) = \emptyset$, let $p \in [1, \infty)$, and assume that*

$$\sup_{x \in \mathbb{R}^d} \left(\frac{\mathbb{E}\big[\|G(x, Z_1)\|_{\mathbb{R}^d}^p\big]}{\big[1 + \|x\|_{\mathbb{R}^d}\big]^p} \right) < \infty. \tag{3.115}$$

Then we have for all $t \in [0, \infty)$ that

$$\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big] < \infty. \tag{3.116}$$

Proof of Lemma 3.6.3. Throughout this proof let $\mathfrak{t} \colon \mathbb{N}_0 \to [0, \infty)$ be a non-decreasing function which satisfies that

$$\{t \in [0, \infty) \colon \gamma(t) \neq \emptyset\} = \{\mathfrak{t}_n \colon n \in \mathbb{N}_0\}. \tag{3.117}$$

Observe that (3.115) implies that there exists $c \in [0, \infty)$ which satisfies for all $x \in \mathbb{R}^d$ that

$$\mathbb{E}\big[\|G(x, Z_1)\|_{\mathbb{R}^d}^p\big] \leq c(1 + \|x\|_{\mathbb{R}^d}^p). \tag{3.118}$$

Next note that the assumption that $\forall\, t \in [0, \infty) \colon 0 < \#_{\{s \in [0,t] \colon \gamma(s) \neq \emptyset\}}$ assures that $\mathfrak{t}_0 = 0$. This and the assumption that $\Theta_0 = \xi$ imply that

$$\mathbb{E}\big[\|\Theta_{\mathfrak{t}_0}\|_{\mathbb{R}^d}^p\big] = \mathbb{E}\big[\|\Theta_0\|_{\mathbb{R}^d}^p\big] = \|\xi\|_{\mathbb{R}^d}^p < \infty. \tag{3.119}$$

Furthermore, observe that the Minkowski inequality and (3.4) ensure that for all $n \in \mathbb{N}_0$ we have that

$$\begin{aligned} \big|\mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n+1}}\|_{\mathbb{R}^d}^p\big]\big|^{1/p} &= \Big|\mathbb{E}\Big[\|\Theta_{\mathfrak{t}_n} + \tfrac{(\mathfrak{t}_{n+1} - \mathfrak{t}_n)}{\#_{\gamma(\mathfrak{t}_n)}} \textstyle\sum_{j \in \gamma(\mathfrak{t}_n)} G(\Theta_{\mathfrak{t}_n}, Z_j)\|_{\mathbb{R}^d}^p\Big]\Big|^{1/p} \\ &\leq \big|\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^p\big]\big|^{1/p} + (\mathfrak{t}_{n+1} - \mathfrak{t}_n)\Big|\mathbb{E}\Big[\big\|\tfrac{1}{\#_{\gamma(\mathfrak{t}_n)}} \textstyle\sum_{j \in \gamma(\mathfrak{t}_n)} G(\Theta_{\mathfrak{t}_n}, Z_j)\big\|_{\mathbb{R}^d}^p\Big]\Big|^{1/p}. \end{aligned} \tag{3.120}$$

This, (3.118), and Lemma 3.6.2 prove that for all $n \in \mathbb{N}_0$ we have that

$$\big|\mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n+1}}\|_{\mathbb{R}^d}^p\big]\big|^{1/p} \leq \big|\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^p\big]\big|^{1/p} + (\mathfrak{t}_{n+1} - \mathfrak{t}_n)\big|c\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^p\big]\big)\big|^{1/p}. \tag{3.121}$$

Induction and (3.119) therefore assure that for all $n \in \mathbb{N}_0$ we have that

$$\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^p\big] < \infty. \tag{3.122}$$

This establishes (3.116). The proof of Lemma 3.6.3 is thus completed. □

Corollary 3.6.4. *Assume Setting 3.1.1, assume for all $v, w \in [0,\infty)$ with $v \neq w$ that $\gamma(v) \cap \gamma(w) = \emptyset$, and let $p \in [1,\infty)$, $c \in [0,\infty)$ satisfy for all $x \in \mathbb{R}^d$ that*

$$\mathbb{E}\big[\|G(x, Z_1)\|_{\mathbb{R}^d}^p\big] \leq c\big(1 + \|x\|_{\mathbb{R}^d}^p\big). \tag{3.123}$$

Then

(i) we have for all $t \in [0,\infty)$ that

$$\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big] \leq 2^{p-1}c(t - [\![t]\!])^p + 2^{p-1}(1 + c(t - [\![t]\!])^p)\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big] \tag{3.124}$$

and

(ii) we have for all $T \in [0,\infty)$ that

$$\sup_{t\in[0,T]} \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big] < \infty. \tag{3.125}$$

Proof of Corollary 3.6.4. First, note that (3.4) and the Minkowski inequality imply that for all $t \in [0,\infty)$ we have that

$$\begin{aligned} \big|\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big]\big|^{1/p} &= \Big|\mathbb{E}\Big[\big\|\Theta_{[\![t]\!]} + \tfrac{(t-[\![t]\!])}{\#_{\gamma([\![t]\!])}} \textstyle\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_j)\big\|_{\mathbb{R}^d}^p\Big]\Big|^{1/p} \\ &\leq \big|\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big]\big|^{1/p} + (t - [\![t]\!])\Big|\mathbb{E}\Big[\big\|\tfrac{1}{\#_{\gamma([\![t]\!])}} \textstyle\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_j)\big\|_{\mathbb{R}^d}^p\Big]\Big|^{1/p}. \end{aligned} \tag{3.126}$$

Lemma 3.6.2 and (3.123) hence prove that for all $t \in [0,\infty)$ we have that

$$\big|\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big]\big|^{1/p} \leq \big|\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big]\big|^{1/p} + (t - [\![t]\!])\big|c\big(1 + \mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big]\big)\big|^{1/p}. \tag{3.127}$$

Lemma 3.6.1 therefore demonstrates that for all $t \in [0,\infty)$ we have that

$$\begin{aligned} \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big] &\leq 2^{p-1}\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big] + (t - [\![t]\!])^p 2^{p-1}c\big(1 + \mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big]\big) \\ &= 2^{p-1}c(t - [\![t]\!])^p + 2^{p-1}(1 + c(t - [\![t]\!])^p)\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big]. \end{aligned} \tag{3.128}$$

This proves item (i). Next note that (3.128) ensures that for all $T \in [0,\infty)$ we have that

$$\sup_{t\in[0,T]} \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big] \leq 2^{p-1}cT^p + 2^{p-1}(1 + cT^p) \sup_{t\in[0,T]} \mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big]. \tag{3.129}$$

The assumption that $\forall\, T \in [0,\infty) \colon 0 < \#_{\{s\in[0,T] \colon \gamma(s)\neq\emptyset\}} < \infty$ and Lemma 3.6.3 hence imply that for all $T \in [0,\infty)$ we have that

$$\sup_{t\in[0,T]} \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big] < \infty. \tag{3.130}$$

This establishes item (ii). The proof of Corollary 3.6.4 is thus completed. □

Lemma 3.6.5. *Assume Setting 3.1.1, assume that* $\#_{\{t\in[0,\infty)\colon \gamma(t)\neq\emptyset\}} = \infty$, *and let* $p \in [1,\infty)$. *Then the following two statements are equivalent:*

(i) *It holds that*
$$\sup_{t\in[0,\infty)} \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big] < \infty. \tag{3.131}$$

(ii) *It holds that*
$$\sup_{t\in[0,\infty)} \mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big] < \infty. \tag{3.132}$$

Proof of Lemma 3.6.5. Throughout this proof let $\lceil\cdot\rceil\colon [0,\infty) \to [0,\infty]$ satisfy for all $t \in [0,\infty)$ that
$$\lceil t\rceil = \inf\big(\{s \in (t,\infty)\colon \gamma(s) \neq \emptyset\} \cup \{\infty\}\big). \tag{3.133}$$
Note that it is clear that ((i) $\Rightarrow$ (ii)). Next we prove that ((ii) $\Rightarrow$ (i)). Observe that the assumption that $\forall\, t \in [0,\infty)\colon \#_{\{s\in[0,t]\colon \gamma(s)\neq\emptyset\}} < \infty$ and the assumption that $\#_{\{t\in[0,\infty)\colon \gamma(t)\neq\emptyset\}} = \infty$ assure that for all $t \in [0,\infty)$ we have that
$$\lceil t\rceil < \infty. \tag{3.134}$$
This and (3.4) prove that for all $t \in [0,\infty)$ we have that
$$\Theta_{\lceil t\rceil} = \Theta_{[\![t]\!]} + (\lceil t\rceil - [\![t]\!])\tfrac{1}{\#_{\gamma([\![t]\!])}} \textstyle\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_j). \tag{3.135}$$
Combining this and (3.4) ensures that for all $t \in [0,\infty)$ we have that
$$\begin{aligned}
&\tfrac{\lceil t\rceil - t}{\lceil t\rceil - [\![t]\!]}\Theta_{[\![t]\!]} + \tfrac{t-[\![t]\!]}{\lceil t\rceil - [\![t]\!]}\Theta_{\lceil t\rceil} \\
&= \tfrac{\lceil t\rceil - t}{\lceil t\rceil - [\![t]\!]}\Theta_{[\![t]\!]} + \tfrac{t-[\![t]\!]}{\lceil t\rceil - [\![t]\!]}\big(\Theta_{[\![t]\!]} + (\lceil t\rceil - [\![t]\!])\tfrac{1}{\#_{\gamma([\![t]\!])}} \textstyle\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_j)\big) \\
&= \tfrac{\lceil t\rceil - t}{\lceil t\rceil - [\![t]\!]}\Theta_{[\![t]\!]} + \tfrac{t-[\![t]\!]}{\lceil t\rceil - [\![t]\!]}\Theta_{[\![t]\!]} + (t - [\![t]\!])\tfrac{1}{\#_{\gamma([\![t]\!])}} \textstyle\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_j) \\
&= \Theta_{[\![t]\!]} + (t - [\![t]\!])\tfrac{1}{\#_{\gamma([\![t]\!])}} \textstyle\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_j) = \Theta_t.
\end{aligned} \tag{3.136}$$
This and the triangle inequality imply that for all $t \in [0,\infty)$ we have that
$$\|\Theta_t\|_{\mathbb{R}^d} \leq \|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d} + \|\Theta_{\lceil t\rceil}\|_{\mathbb{R}^d}. \tag{3.137}$$
Lemma 3.6.1 therefore demonstrates that
$$\begin{aligned}
\sup_{t\in[0,\infty)} \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^p\big] &\leq \sup_{t\in[0,\infty)} \mathbb{E}\big[\big(\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d} + \|\Theta_{\lceil t\rceil}\|_{\mathbb{R}^d}\big)^p\big] \\
&\leq 2^{p-1} \sup_{t\in[0,\infty)} \mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p + \|\Theta_{\lceil t\rceil}\|_{\mathbb{R}^d}^p\big] \\
&\leq 2^p \sup_{t\in[0,\infty)} \mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^p\big].
\end{aligned} \tag{3.138}$$
This reveals that ((ii) $\Rightarrow$ (i)). The proof of Lemma 3.6.5 is thus completed. ∎

3.7 Weak error estimates for SAAs in the case of general learning rates with mini-batches

Proposition 3.7.1. *Assume Setting 3.1.1, assume for all $v, w \in [0,\infty)$ with $v \neq w$ that $\gamma(v) \cap \gamma(w) = \emptyset$, let $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$, $L \in (0,\infty)$, assume for all $y, z \in \mathbb{R}^d$ that*

$$\langle g(y) - g(z), y - z \rangle_{\mathbb{R}^d} \leq -L \|y - z\|^2_{\mathbb{R}^d}, \tag{3.139}$$

assume that

$$\sup_{x \in \mathbb{R}^d} \left(\frac{\mathbb{E}\big[\|G(x, Z_1)\|^2_{\mathbb{R}^d}\big]}{\big[1 + \|x\|_{\mathbb{R}^d}\big]^2} + \frac{\big\|\mathbb{E}\big[(\frac{\partial}{\partial x} G)(x, Z_1)\big]\big\|_{L(\mathbb{R}^d, \mathbb{R}^d)}}{\big[1 + \|x\|_{\mathbb{R}^d}\big]} + \|\psi'(x)\|_{L(\mathbb{R}^d, \mathbb{R})} \right) < \infty, \tag{3.140}$$

and let $Q \colon [0,\infty) \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $t \in [0,\infty)$ that

$$Q_t = \tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}} \Big[\textstyle\sum_{j \in \gamma(\llbracket t \rrbracket)} G(\Theta_{\llbracket t \rrbracket}, Z_j) \Big]. \tag{3.141}$$

Then

(i) we have that $g \in C^2(\mathbb{R}^d, \mathbb{R}^d)$,

(ii) we have that there exists a unique $\Xi \in \mathbb{R}^d$ which satisfies that

$$\limsup_{t \to \infty} \|\theta^\xi_t - \Xi\|_{\mathbb{R}^d} = 0, \tag{3.142}$$

and

(iii) we have for all $T \in (0,\infty)$ that

$$\begin{aligned}
&|\mathbb{E}[\psi(\Theta_T)] - \psi(\Xi)| \leq \sup_{s,v \in [0,T]} \mathbb{E}\Bigg[\|Q_s - g(\Theta_{\llbracket s \rrbracket})\|_{\mathbb{R}^d} \, \|Q_s\|_{\mathbb{R}^d} \\
&\cdot \Bigg(\int_0^1 \exp(-L(T-s)) \big\|\psi''(\theta^{\lambda\Theta_s + (1-\lambda)\Theta_{\llbracket s \rrbracket}}_{T-s})\big\|_{L^{(2)}(\mathbb{R}^d, \mathbb{R})} \\
&+ \big\|\psi'(\theta^{\lambda\Theta_s + (1-\lambda)\Theta_{\llbracket s \rrbracket}}_{T-s})\big\|_{L(\mathbb{R}^d, \mathbb{R})} \int_0^{T-s} \exp(-Lu) \big\|g''(\theta^{\lambda\Theta_s + (1-\lambda)\Theta_{\llbracket s \rrbracket}}_u)\big\|_{L^{(2)}(\mathbb{R}^d, \mathbb{R}^d)} \, du \, d\lambda \Bigg) \\
&+ \big\|\psi'(\theta^{\Theta_s}_{T-s})\big\|_{L(\mathbb{R}^d, \mathbb{R})} \big\|g'(\Theta_v) Q_v\big\|_{\mathbb{R}^d} \Bigg] \int_0^T \exp(-L(T-t))(t - \llbracket t \rrbracket) \, dt \\
&+ \sup\Big\{ \big\|\psi'(\lambda\theta^\xi_T + (1-\lambda)\Xi)\big\|_{L(\mathbb{R}^d, \mathbb{R})} \in \mathbb{R} \colon \lambda \in [0,1] \Big\} \|\xi - \Xi\|_{\mathbb{R}^d} \exp(-LT).
\end{aligned} \tag{3.143}$$

Proof of Proposition 3.7.1. Throughout this proof let $T \in (0, \infty)$, let $E \subseteq [0, T]$ be the set given by

$$E = \{t \in [0, T] \colon \gamma(t) \neq \emptyset\} \cup \{T\}, \tag{3.144}$$

let $u = (u(t, \vartheta))_{(t,\vartheta)\in[0,T]\times\mathbb{R}^d} \in C([0, T] \times \mathbb{R}^d, \mathbb{R})$ satisfy for all $t \in [0, T]$, $\vartheta \in \mathbb{R}^d$ that

$$u(t, \vartheta) = \psi(\theta^{\vartheta}_{T-t}), \tag{3.145}$$

let $u_{1,0} = (u_{1,0}(t, \vartheta))_{(t,\vartheta)\in[0,T]\times\mathbb{R}^d} \in C([0, T] \times \mathbb{R}^d, L(\mathbb{R}, \mathbb{R}))$ satisfy for all $t \in [0, T]$, $\vartheta \in \mathbb{R}^d$ that

$$u_{1,0}(t, \vartheta) = (\tfrac{\partial}{\partial t} u)(t, \vartheta) \tag{3.146}$$

(cf. item (i) in Lemma 3.3.8), let $u_{0,1} = (u_{0,1}(t, \vartheta))_{(t,\vartheta)\in[0,T]\times\mathbb{R}^d} \in C([0, T] \times \mathbb{R}^d, L(\mathbb{R}^d, \mathbb{R}))$ satisfy for all $t \in [0, T]$, $\vartheta \in \mathbb{R}^d$ that

$$u_{0,1}(t, \vartheta) = (\tfrac{\partial}{\partial \vartheta} u)(t, \vartheta) \tag{3.147}$$

(cf. item (i) in Lemma 3.3.8), let $\delta \colon [0, T] \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $t \in [0, T]$ that

$$\delta_t = \Theta_t - \Theta_{[\![t]\!]}, \tag{3.148}$$

let $\theta^{1,\vartheta} \in C([0, \infty), L(\mathbb{R}^d, \mathbb{R}^d))$, $\vartheta \in \mathbb{R}^d$, satisfy for all $t \in [0, \infty)$, $\vartheta \in \mathbb{R}^d$ that

$$\theta^{1,\vartheta}_t = \tfrac{\partial}{\partial \vartheta} \theta^{\vartheta}_t \tag{3.149}$$

(cf. item (ii) in Lemma 3.4.4), let $\theta^{2,\vartheta} \in C([0, \infty), L^{(2)}(\mathbb{R}^d, \mathbb{R}^d))$, $\vartheta \in \mathbb{R}^d$, satisfy for all $t \in [0, \infty)$, $\vartheta \in \mathbb{R}^d$ that

$$\theta^{2,\vartheta}_t = \tfrac{\partial^2}{\partial \vartheta^2} \theta^{\vartheta}_t \tag{3.150}$$

(cf. item (ii) in Lemma 3.4.4), let $a^\lambda \colon [0, \infty) \times \Omega \to \mathbb{R}^d$, $\lambda \in [0, 1]$, be the stochastic processes which satisfy for all $\lambda \in [0, 1]$, $t \in [0, \infty)$ that

$$a^\lambda_t = \lambda \Theta_t + (1 - \lambda)\Theta_{[\![t]\!]}, \tag{3.151}$$

and let $\Delta \colon [0, \infty) \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $t \in [0, \infty)$ that

$$\Delta_t = \tfrac{1}{\#_{\gamma([\![t]\!])}} \Big[\textstyle\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}, Z_j) \Big] - g(\Theta_{[\![t]\!]}) = Q_t - g(\Theta_{[\![t]\!]}). \tag{3.152}$$

Observe that item (i) in Lemma 3.4.3 establishes item (i). Next note that Lemma 3.3.6 ensures that there exists a unique $\Xi \in \mathbb{R}^d$ which satisfies that $g(\Xi) = 0$ and

$$\limsup_{t\to\infty} \|\theta^{\xi}_t - \Xi\|_{\mathbb{R}^d} = 0. \tag{3.153}$$

This establishes item (ii). Next observe that the assumption that $\forall\, t \in [0,\infty)\colon 0 < \#_{\{s\in[0,t]\colon \gamma(s)\neq\emptyset\}} < \infty$ ensures that there exist $k \in \mathbb{N}$, $\mathfrak{t}_1, \mathfrak{t}_2, \dots, \mathfrak{t}_k \in [0,T]$ which satisfy that

$$0 = \mathfrak{t}_1 < \mathfrak{t}_2 < \cdots < \mathfrak{t}_k = [\![T]\!] \quad \text{and} \quad \{t \in [0,T]\colon \gamma(t) \neq \emptyset\} = \{\mathfrak{t}_1, \mathfrak{t}_2, \dots, \mathfrak{t}_k\}. \tag{3.154}$$

Note that (3.154) implies that there exists $\mathfrak{j}\colon \{1,2,\dots,k\} \to \mathbb{N}$ which satisfies for all $n \in \{1,2,\dots,k\}$ that

$$\mathfrak{j}_n \in \gamma(\mathfrak{t}_n). \tag{3.155}$$

Next note that (3.140) and Lemma 3.2.1 assure that there exists $c \in (0,\infty)$ which satisfies for all $x \in \mathbb{R}^d$ that

$$\max\left\{\mathbb{E}\big[\|G(x,Z_1)\|_{\mathbb{R}^d}^2\big]^{1/2}, \|g(x)\|_{\mathbb{R}^d}, \|g'(x)\|_{L(\mathbb{R}^d,\mathbb{R}^d)}\right\} \leq c(1+\|x\|_{\mathbb{R}^d}), \tag{3.156}$$

$$\mathbb{E}\big[\|G(x,Z_1)\|_{\mathbb{R}^d}^2\big] \leq c(1+\|x\|_{\mathbb{R}^d}^2), \tag{3.157}$$

and

$$\|\psi'(x)\|_{L(\mathbb{R}^d,\mathbb{R})} \leq c. \tag{3.158}$$

Item (ii) in Lemma 3.3.9 hence ensures that for all $t \in [0,T]$, $\vartheta \in \mathbb{R}^d$ we have that

$$\|u_{0,1}(t,\vartheta)\|_{L(\mathbb{R}^d,\mathbb{R})} = \|(\tfrac{\partial}{\partial\vartheta}u)(t,\vartheta)\|_{L(\mathbb{R}^d,\mathbb{R})} \leq \|\psi'(\theta_{T-t}^{\vartheta})\|_{L(\mathbb{R}^d,\mathbb{R})}e^{-L(T-t)} \leq c. \tag{3.159}$$

Next note Lemma 3.6.2 and (3.157) imply that for all $t \in [0,T]$ we have that

$$\mathbb{E}\big[\|Q_t\|_{\mathbb{R}^d}^2\big] \leq c\big(1+\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^2\big]\big). \tag{3.160}$$

Jensen's inequality therefore proves that for all $t \in [0,T]$ we have that

$$\mathbb{E}\big[\|Q_t\|_{\mathbb{R}^d}\big] \leq \big|\mathbb{E}\big[\|Q_t\|_{\mathbb{R}^d}^2\big]\big|^{1/2} \leq \big|c\big(1+\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|_{\mathbb{R}^d}^2\big]\big)\big|^{1/2}. \tag{3.161}$$

Moreover, observe that (3.156) and Jensen's inequality ensure that for all $t \in [0,T]$ we have that

$$\mathbb{E}\big[\|g(\Theta_t)\|_{\mathbb{R}^d}\big] \leq c\big(1+\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}\big]\big) \leq c\big(1+\big|\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^2\big]\big|^{1/2}\big). \tag{3.162}$$

This, (3.159), and (3.161) assure that

$$\begin{aligned}
&\int_0^T \mathbb{E}\big[(\|u_{0,1}(t,\Theta_t)\|_{L(\mathbb{R}^d,\mathbb{R})} + \|u_{0,1}(t,\Theta_{[\![t]\!]})\|_{L(\mathbb{R}^d,\mathbb{R})}) \\
&\quad \cdot (\|Q_t\|_{\mathbb{R}^d} + \|g(\Theta_{[\![t]\!]})\|_{\mathbb{R}^d} + \|g(\Theta_t)\|_{\mathbb{R}^d})\big]\,dt \\
&\le \int_0^T 2c\Big(\big|c\big(1+\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|^2_{\mathbb{R}^d}\big]\big)\big|^{1/2} + c\big(1+\big|\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|^2_{\mathbb{R}^d}\big]\big|^{1/2}\big) \\
&\quad + c\big(1+\big|\mathbb{E}\big[\|\Theta_t\|^2_{\mathbb{R}^d}\big]\big|^{1/2}\big)\Big)\,dt \\
&\le 2cT\Big(\big|c\big(1+\big(\sup_{t\in[0,T]}\mathbb{E}\big[\|\Theta_{[\![t]\!]}\|^2_{\mathbb{R}^d}\big]\big)\big)\big|^{1/2} + 2c\big(1+\big|\big(\sup_{t\in[0,T]}\mathbb{E}\big[\|\Theta_t\|^2_{\mathbb{R}^d}\big]\big)\big|^{1/2}\big)\Big).
\end{aligned} \tag{3.163}$$

In addition, observe that Corollary 3.6.4 and (3.157) prove that

$$\sup_{t\in[0,T]}\mathbb{E}\big[\|\Theta_t\|^2_{\mathbb{R}^d}\big] < \infty. \tag{3.164}$$

This and (3.163) demonstrate that

$$\begin{aligned}
&\int_0^T \mathbb{E}\big[(\|u_{0,1}(t,\Theta_t)\|_{L(\mathbb{R}^d,\mathbb{R})} + \|u_{0,1}(t,\Theta_{[\![t]\!]})\|_{L(\mathbb{R}^d,\mathbb{R})}) \\
&\quad \cdot (\|Q_t\|_{\mathbb{R}^d} + \|g(\Theta_{[\![t]\!]})\|_{\mathbb{R}^d} + \|g(\Theta_t)\|_{\mathbb{R}^d})\big]\,dt < \infty.
\end{aligned} \tag{3.165}$$

Furthermore, note that (3.159) and Jensen's inequality imply that for all $s,t\in[0,T]$, $j\in\mathbb{N}$ we have that

$$\begin{aligned}
&\mathbb{E}\big[|u_{0,1}(t,\Theta_s)G(\Theta_s,Z_j)| + |u_{0,1}(t,\Theta_s)g(\Theta_s)|\big] \\
&\le \mathbb{E}\big[\|u_{0,1}(t,\Theta_s)\|_{L(\mathbb{R}^d,\mathbb{R})}\|G(\Theta_s,Z_j)\|_{\mathbb{R}^d} + \|u_{0,1}(t,\Theta_s)\|_{L(\mathbb{R}^d,\mathbb{R})}\|g(\Theta_s)\|_{\mathbb{R}^d}\big] \\
&\le c\,\mathbb{E}\big[\|G(\Theta_s,Z_j)\|_{\mathbb{R}^d}\big] + c\,\mathbb{E}\big[\|g(\Theta_s)\|_{\mathbb{R}^d}\big] \\
&\le c\,\big|\mathbb{E}\big[\|G(\Theta_s,Z_j)\|^2_{\mathbb{R}^d}\big]\big|^{1/2} + c\,\mathbb{E}\big[\|g(\Theta_s)\|_{\mathbb{R}^d}\big].
\end{aligned} \tag{3.166}$$

Moreover, observe that the assumption that Z_j, $j\in\mathbb{N}$, are i.i.d. random variables, the assumption that $\forall\,(v,w)\in\{(a,b)\in[0,\infty)^2\colon a\neq b\}\colon \gamma(v)\cap\gamma(w)=\emptyset$, and (3.4) prove that for all $s\in[0,T]$, $j\in\gamma([\![s]\!])$ we have that Z_j and $\Theta_{[\![s]\!]}$ are independent. This and the assumption that Z_j, $j\in\mathbb{N}$, are i.i.d. random variables ensure that for all $s\in[0,T]$, $j\in\gamma([\![s]\!])$ we have that

$$\begin{aligned}
\mathbb{E}\big[\|G(\Theta_{[\![s]\!]},Z_j)\|^2_{\mathbb{R}^d}\big] &= \int_\Omega \|G(\Theta_{[\![s]\!]}(\omega),Z_j(\omega))\|^2_{\mathbb{R}^d}\,\mathbb{P}(d\omega) \\
&= \int_\Omega\int_\Omega \|G(\Theta_{[\![s]\!]}(\omega),Z_j(\tilde\omega))\|^2_{\mathbb{R}^d}\,\mathbb{P}(d\tilde\omega)\,\mathbb{P}(d\omega) \\
&= \int_\Omega\int_\Omega \|G(\Theta_{[\![s]\!]}(\omega),Z_1(\tilde\omega))\|^2_{\mathbb{R}^d}\,\mathbb{P}(d\tilde\omega)\,\mathbb{P}(d\omega).
\end{aligned} \tag{3.167}$$

Combining this with (3.157) implies that for all $s \in [0,T]$, $j \in \gamma(\llbracket s \rrbracket)$ we have that

$$\mathbb{E}\big[\|G(\Theta_{\llbracket s \rrbracket}, Z_j)\|_{\mathbb{R}^d}^2\big] \leq \int_\Omega c\big(1 + \|\Theta_{\llbracket s \rrbracket}(\omega)\|_{\mathbb{R}^d}^2\big)\, \mathbb{P}(d\omega) = c\big(1 + \mathbb{E}\big[\|\Theta_{\llbracket s \rrbracket}\|_{\mathbb{R}^d}^2\big]\big). \quad (3.168)$$

This and (3.166) demonstrate that for all $s, t \in [0,T]$, $j \in \gamma(\llbracket s \rrbracket)$ we have that

$$\begin{aligned} &\mathbb{E}\big[|u_{0,1}(t, \Theta_{\llbracket s \rrbracket}) G(\Theta_{\llbracket s \rrbracket}, Z_j)| + |u_{0,1}(t, \Theta_{\llbracket s \rrbracket}) g(\Theta_{\llbracket s \rrbracket})|\big] \\ &\leq c^{3/2} \big|1 + \mathbb{E}\big[\|\Theta_{\llbracket s \rrbracket}\|_{\mathbb{R}^d}^2\big]\big|^{1/2} + c\, \mathbb{E}\big[\|g(\Theta_{\llbracket s \rrbracket})\|_{\mathbb{R}^d}\big]. \end{aligned} \quad (3.169)$$

Combining this and (3.156) assures that for all $s, t \in [0,T]$, $j \in \gamma(\llbracket s \rrbracket)$ we have that

$$\begin{aligned} &\mathbb{E}\big[|u_{0,1}(t, \Theta_{\llbracket s \rrbracket}) G(\Theta_{\llbracket s \rrbracket}, Z_j)| + |u_{0,1}(t, \Theta_{\llbracket s \rrbracket}) g(\Theta_{\llbracket s \rrbracket})|\big] \\ &\leq c^{3/2} \big|1 + \mathbb{E}\big[\|\Theta_{\llbracket s \rrbracket}\|_{\mathbb{R}^d}^2\big]\big|^{1/2} + c^2 \big(1 + \mathbb{E}\big[\|\Theta_{\llbracket s \rrbracket}\|_{\mathbb{R}^d}\big]\big). \end{aligned} \quad (3.170)$$

Jensen's inequality and (3.164) hence prove that for all $s, t \in [0,T]$, $j \in \gamma(\llbracket s \rrbracket)$ we have that

$$\begin{aligned} &\mathbb{E}\big[|u_{0,1}(t, \Theta_{\llbracket s \rrbracket}) G(\Theta_{\llbracket s \rrbracket}, Z_j)| + |u_{0,1}(t, \Theta_{\llbracket s \rrbracket}) g(\Theta_{\llbracket s \rrbracket})|\big] \\ &\leq c^{3/2} \Big|1 + \sup_{u \in [0,T]} \mathbb{E}\big[\|\Theta_u\|_{\mathbb{R}^d}^2\big]\Big|^{1/2} + c^2 \Big(1 + \sup_{u \in [0,T]} \big|\mathbb{E}\big[\|\Theta_u\|_{\mathbb{R}^d}^2\big]\big|^{1/2}\Big) < \infty. \end{aligned} \quad (3.171)$$

Next note that item (ii) in Lemma 3.3.6 and Lemma 3.3.7 ensure that

$$\begin{aligned} &|\psi(\theta_T^\xi) - \psi(\Xi)| = |\psi(\theta_T^\xi) - \psi(\theta_T^\Xi)| \\ &\leq \sup\Big\{\|\psi'(\lambda \theta_T^\xi + (1-\lambda)\theta_T^\Xi)\|_{L(\mathbb{R}^d, \mathbb{R})} \in \mathbb{R} \colon \lambda \in [0,1]\Big\} \|\xi - \Xi\|_{\mathbb{R}^d} e^{-LT} \\ &= \sup\Big\{\|\psi'(\lambda \theta_T^\xi + (1-\lambda)\Xi)\|_{L(\mathbb{R}^d, \mathbb{R})} \in \mathbb{R} \colon \lambda \in [0,1]\Big\} \|\xi - \Xi\|_{\mathbb{R}^d} e^{-LT}. \end{aligned} \quad (3.172)$$

In the next step we combine (3.145) and (3.2) to obtain that for all $\vartheta \in \mathbb{R}^d$ we have that

$$\psi(\vartheta) = \psi(\theta_0^\vartheta) = u(T, \vartheta). \quad (3.173)$$

This and the assumption that $\forall\, \omega \in \Omega \colon \Theta_0(\omega) = \xi$ prove that for all $\omega \in \Omega$ we have that

$$\psi(\Theta_T(\omega)) = u(T, \Theta_T(\omega)) \quad (3.174)$$

and

$$\psi(\theta_T^\xi) = \psi(\theta_T^{\Theta_0(\omega)}) = u(0, \Theta_0(\omega)). \quad (3.175)$$

Next observe that Lemma 3.5.1 and (3.141) assure that for all $\omega \in \Omega$, $t \in [0,T]\backslash E$ we have that

$$[0,T]\backslash E \subseteq \big\{u \in [0,T]\colon [0,T] \ni s \mapsto \Theta_s(\omega) \in \mathbb{R}^d \text{ is differentiable at } u\big\} \tag{3.176}$$

and

$$\tfrac{\partial}{\partial t}\Theta_t(\omega) = Q_t(\omega). \tag{3.177}$$

This, item (i) in Lemma 3.3.8, and the fact that $\psi \in C^1(\mathbb{R}^d, \mathbb{R})$ imply that for all $\omega \in \Omega$ we have that

$$[0,T]\backslash E \subseteq \{t \in [0,T]\colon [0,T] \ni s \mapsto u(s, \Theta_s(\omega)) \in \mathbb{R} \text{ is differentiable at } t\}\,. \tag{3.178}$$

This reveals that for all $\omega \in \Omega$, $t \in [0,T]\backslash E$ it holds that

$$\tfrac{\partial}{\partial t}[u(t, \Theta_t(\omega))] = (\tfrac{\partial}{\partial t}u)(t, \Theta_t(\omega)) + (\tfrac{\partial}{\partial \vartheta}u)(t, \Theta_t(\omega))\tfrac{\partial}{\partial t}\Theta_t(\omega). \tag{3.179}$$

Next note that (3.177) and (3.147) demonstrate that

$$\begin{aligned} &\int_{[0,T]\backslash E} \mathbb{E}\big[|(\tfrac{\partial}{\partial \vartheta}u)(t, \Theta_t)\tfrac{\partial}{\partial t}\Theta_t|\big]\,dt \\ &\leq \int_{[0,T]\backslash E} \mathbb{E}\big[\|(\tfrac{\partial}{\partial \vartheta}u)(t, \Theta_t)\|_{L(\mathbb{R}^d,\mathbb{R})}\|\tfrac{\partial}{\partial t}\Theta_t\|_{\mathbb{R}^d}\big]\,dt \\ &= \int_{[0,T]\backslash E} \mathbb{E}\big[\|u_{0,1}(t, \Theta_t)\|_{L(\mathbb{R}^d,\mathbb{R})}\|Q_t\|_{\mathbb{R}^d}\big]\,dt. \end{aligned} \tag{3.180}$$

This and (3.165) assure that

$$\int_{[0,T]\backslash E} \mathbb{E}\big[|(\tfrac{\partial}{\partial \vartheta}u)(t, \Theta_t)\tfrac{\partial}{\partial t}\Theta_t|\big]\,dt < \infty. \tag{3.181}$$

Next observe that (3.145) ensures that for all $t \in [0,T], \vartheta \in \mathbb{R}^d$ we have that

$$\begin{aligned} |(\tfrac{\partial}{\partial t}u)(t, \vartheta)| &= |\psi'(\theta^\vartheta_{T-t})\tfrac{\partial}{\partial t}(\theta^\vartheta_{T-t})| = |\psi'(\theta^\vartheta_{T-t})g(\theta^\vartheta_{T-t})| \\ &\leq \|\psi'(\theta^\vartheta_{T-t})\|_{L(\mathbb{R}^d,\mathbb{R})}\|g(\theta^\vartheta_{T-t})\|_{\mathbb{R}^d}. \end{aligned} \tag{3.182}$$

This, (3.158), and (3.156) imply that for all $t \in [0,T], \vartheta \in \mathbb{R}^d$ we have that

$$|(\tfrac{\partial}{\partial t}u)(t, \vartheta)| \leq c\|g(\theta^\vartheta_{T-t})\|_{\mathbb{R}^d} \leq c^2(1 + \|\theta^\vartheta_{T-t}\|_{\mathbb{R}^d}). \tag{3.183}$$

The triangle inequality hence proves that for all $t \in [0,T]$ we have that

$$\begin{aligned} \mathbb{E}\big[|(\tfrac{\partial}{\partial t}u)(t,\Theta_t)|\big] &\leq c^2\big(1+\mathbb{E}\big[\|\theta^{\Theta_t}_{T-t}\|_{\mathbb{R}^d}\big]\big) \\ &= c^2\big(1+\mathbb{E}\big[\|\theta^{\Xi+(\Theta_t-\Xi)}_{T-t}\|_{\mathbb{R}^d}\big]\big) \\ &\leq c^2\big(1+\|\Xi\|_{\mathbb{R}^d}+\mathbb{E}\big[\|\theta^{\Xi+(\Theta_t-\Xi)}_{T-t}-\Xi\|_{\mathbb{R}^d}\big]\big). \end{aligned} \tag{3.184}$$

This, item (iii) in Lemma 3.3.6, and the triangle inequality ensure that for all $t \in [0,T]$ we have that

$$\begin{aligned} \mathbb{E}\big[|(\tfrac{\partial}{\partial t}u)(t,\Theta_t)|\big] &\leq c^2\big(1+\|\Xi\|_{\mathbb{R}^d}+\mathbb{E}\big[\|\Theta_t-\Xi\|_{\mathbb{R}^d}\big]e^{-L(T-t)}\big) \\ &\leq c^2\big(1+\|\Xi\|_{\mathbb{R}^d}+\mathbb{E}\big[\|\Theta_t-\Xi\|_{\mathbb{R}^d}\big]\big) \\ &\leq c^2\big(1+2\|\Xi\|_{\mathbb{R}^d}+\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}\big]\big). \end{aligned} \tag{3.185}$$

This reveals that

$$\begin{aligned} \int_0^T \mathbb{E}\big[|(\tfrac{\partial}{\partial t}u)(t,\Theta_t)|\big]\,dt &\leq \int_0^T c^2\big(1+2\|\Xi\|_{\mathbb{R}^d}+\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}\big]\big)\,dt \\ &= Tc^2(1+2\|\Xi\|_{\mathbb{R}^d})+c^2\int_0^T \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}\big]\,dt \\ &\leq Tc^2(1+2\|\Xi\|_{\mathbb{R}^d})+Tc^2\sup_{t\in[0,T]}\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}\big]. \end{aligned} \tag{3.186}$$

Jensen's inequality and (3.164) hence imply that

$$\begin{aligned} &\int_0^T \mathbb{E}\big[|(\tfrac{\partial}{\partial t}u)(t,\Theta_t)|\big]\,dt \\ &\leq Tc^2(1+2\|\Xi\|_{\mathbb{R}^d})+Tc^2\sup_{t\in[0,T]}\big(\mathbb{E}\big[\|\Theta_t\|^2_{\mathbb{R}^d}\big]\big)^{1/2}<\infty. \end{aligned} \tag{3.187}$$

Combining this, (3.181), (3.179), and the triangle inequality demonstrates that

$$\begin{aligned} &\int_{[0,T]\setminus E}\int_\Omega \big|\tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))]\big|\,\mathbb{P}(d\omega)\,dt \\ &\leq \int_{[0,T]\setminus E}\int_\Omega \big|(\tfrac{\partial}{\partial t}u)(t,\Theta_t(\omega))\big|+\big|(\tfrac{\partial}{\partial\vartheta}u)(t,\Theta_t(\omega))\tfrac{\partial}{\partial t}\Theta_t(\omega)\big|\,\mathbb{P}(d\omega)\,dt<\infty. \end{aligned} \tag{3.188}$$

This, (3.174), (3.175), the fundamental theorem of calculus, the fact that $\#_E < \infty$, and Tonelli's theorem prove that

$$\begin{aligned}\mathbb{E}\big[|\psi(\Theta_T)-\psi(\theta_T^\xi)|\big] &= \int_\Omega |u(T,\Theta_T(\omega))-u(0,\Theta_0(\omega))|\,\mathbb{P}(d\omega)\\ &\leq \int_\Omega\int_{[0,T]\backslash E} \big|\tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))]\big|\,dt\,\mathbb{P}(d\omega)\\ &= \int_{[0,T]\backslash E}\int_\Omega \big|\tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))]\big|\,\mathbb{P}(d\omega)\,dt<\infty.\end{aligned} \tag{3.189}$$

The fundamental theorem of calculus, (3.174), (3.175), the fact that $\#_E<\infty$, and Fubini's theorem therefore assure that

$$\begin{aligned}\mathbb{E}[\psi(\Theta_T)]-\psi(\theta_T^\xi) &= \int_\Omega u(T,\Theta_T(\omega))-u(0,\Theta_0(\omega))\,\mathbb{P}(d\omega)\\ &= \int_\Omega\int_{[0,T]\backslash E} \tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))]\,dt\,\mathbb{P}(d\omega)\\ &= \int_{[0,T]\backslash E}\int_\Omega \tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))]\,\mathbb{P}(d\omega)\,dt.\end{aligned} \tag{3.190}$$

Furthermore, note that item (ii) in Lemma 3.3.8 implies that for all $t\in[0,T]$, $\vartheta\in\mathbb{R}^d$ we have that

$$u_{1,0}(t,\vartheta)=-u_{0,1}(t,\vartheta)g(\vartheta). \tag{3.191}$$

This and (3.179) assure that for all $t\in[0,T]\backslash E$, $\omega\in\Omega$ we have that

$$\begin{aligned}\tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))] &= u_{1,0}(t,\Theta_t(\omega))+u_{0,1}(t,\Theta_t(\omega))\tfrac{\partial}{\partial t}\Theta_t(\omega)\\ &= u_{0,1}(t,\Theta_t(\omega))\left((\tfrac{\partial}{\partial t}\Theta_t(\omega))-g(\Theta_t(\omega))\right).\end{aligned} \tag{3.192}$$

Combining this and (3.177) demonstrates that for all $t\in[0,T]\backslash E$, $\omega\in\Omega$ we have that

$$\tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))]=u_{0,1}(t,\Theta_t(\omega))\left(Q_t(\omega)-g(\Theta_t(\omega))\right). \tag{3.193}$$

The fact that $\#_E<\infty$ and (3.190) hence imply that

$$\begin{aligned}\mathbb{E}[\psi(\Theta_T)]-\psi(\theta_T^\xi) &= \int_{[0,T]\backslash E}\int_\Omega \tfrac{\partial}{\partial t}[u(t,\Theta_t(\omega))]\,\mathbb{P}(d\omega)\,dt\\ &= \int_{[0,T]\backslash E}\int_\Omega u_{0,1}(t,\Theta_t(\omega))\left(Q_t(\omega)-g(\Theta_t(\omega))\right)\mathbb{P}(d\omega)\,dt\\ &= \int_{[0,T]}\int_\Omega u_{0,1}(t,\Theta_t(\omega))\left(Q_t(\omega)-g(\Theta_t(\omega))\right)\mathbb{P}(d\omega)\,dt.\end{aligned} \tag{3.194}$$

This reveals that

$$\mathbb{E}[\psi(\Theta_T)] - \psi(\theta^{\xi}_T) = \int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)\,(Q_t - g(\Theta_t))\big]\,dt. \tag{3.195}$$

This and (3.165) ensure that

$$\begin{aligned}
&\mathbb{E}[\psi(\Theta_T)] - \psi(\theta^{\xi}_T) \\
&= \int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)(Q_t - g(\Theta_{[\![t]\!]}))\big]\,dt + \int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)(g(\Theta_{[\![t]\!]}) - g(\Theta_t))\big]\,dt.
\end{aligned} \tag{3.196}$$

Combining this with (3.152) assures that

$$\begin{aligned}
&\mathbb{E}[\psi(\Theta_T)] - \psi(\theta^{\xi}_T) \\
&= \int_0^T \mathbb{E}[u_{0,1}(t,\Theta_t)\Delta_t]\,dt + \int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)\big(g(\Theta_{[\![t]\!]}) - g(\Theta_t)\big)\big]\,dt.
\end{aligned} \tag{3.197}$$

Next note that (3.154) and (3.3) demonstrate that for all $i \in \{1,2,\ldots,k-1\}$, $t \in [\mathfrak{t}_i, \mathfrak{t}_{i+1})$ we have that $[\![t]\!] = \mathfrak{t}_i$. This assures that

$$\begin{aligned}
&\int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_{[\![t]\!]})(Q_t - g(\Theta_{[\![t]\!]}))\big]\,dt \\
&= \left[\sum_{i=1}^{k-1} \int_{\mathfrak{t}_i}^{\mathfrak{t}_{i+1}} \mathbb{E}\big[u_{0,1}(t,\Theta_{[\![t]\!]})(Q_t - g(\Theta_{[\![t]\!]}))\big]\,dt\right] \\
&\quad + \int_{\mathfrak{t}_k}^T \mathbb{E}\big[u_{0,1}(t,\Theta_{[\![t]\!]})\,\big(Q_t - g(\Theta_{[\![t]\!]})\big)\big]\,dt \\
&= \left[\sum_{i=1}^{k-1} \int_{\mathfrak{t}_i}^{\mathfrak{t}_{i+1}} \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathfrak{t}_i})(Q_{\mathfrak{t}_i} - g(\Theta_{\mathfrak{t}_i}))\big]\,dt\right] \\
&\quad + \int_{\mathfrak{t}_k}^T \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathfrak{t}_k})(Q_{\mathfrak{t}_k} - g(\Theta_{\mathfrak{t}_k}))\big]\,dt.
\end{aligned} \tag{3.198}$$

Combining (3.141) and (3.171) hence proves that

$$\begin{aligned}
&\int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_{[\![t]\!]})(Q_t - g(\Theta_{[\![t]\!]}))\big]\,dt \\
&= \left[\sum_{i=1}^{k-1} \int_{\mathfrak{t}_i}^{\mathfrak{t}_{i+1}} \left(\left(\frac{1}{\#_{\gamma(\mathfrak{t}_i)}} \sum_{j\in\gamma(\mathfrak{t}_i)} \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathfrak{t}_i})G(\Theta_{\mathfrak{t}_i}, Z_j)\big]\right) - \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathfrak{t}_i})g(\Theta_{\mathfrak{t}_i})\big]\right) dt\right] \\
&\quad + \int_{t_k}^T \left(\left(\frac{1}{\#_{\gamma(\mathfrak{t}_k)}} \sum_{j\in\gamma(\mathfrak{t}_k)} \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathfrak{t}_k})G(\Theta_{\mathfrak{t}_k}, Z_j)\big]\right) - \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathfrak{t}_k})g(\Theta_{\mathfrak{t}_k})\big]\right) dt.
\end{aligned} \tag{3.199}$$

Furthermore, note that the assumption that Z_j, $j \in \mathbb{N}$, are i.i.d. random variables, the fact that for all $s \in [0,T]$, $j \in \gamma(\llbracket s \rrbracket)$ it holds that Z_j and $\Theta_{\llbracket s \rrbracket}$ are independent, and (3.155) ensure that for all $i \in \{1,2,\dots,k\}$, $t \in [0,T]$ we have that

$$\begin{aligned}
&\frac{1}{\#_{\gamma(\mathrm{t}_i)}} \sum_{j \in \gamma(\mathrm{t}_i)} \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathrm{t}_i}) G(\Theta_{\mathrm{t}_i}, Z_j)\big] = \tfrac{1}{\#_{\gamma(\mathrm{t}_i)}} \sum_{j \in \gamma(\mathrm{t}_i)} \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathrm{t}_i}) G(\Theta_{\mathrm{t}_i}, Z_{\mathrm{j}_i})\big] \\
&= \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathrm{t}_i}) G(\Theta_{\mathrm{t}_i}, Z_{\mathrm{j}_i})\big] = \int_\Omega u_{0,1}(t,\Theta_{\mathrm{t}_i}(\omega)) G(\Theta_{\mathrm{t}_i}(\omega), Z_{\mathrm{j}_i}(\omega))\, \mathbb{P}(d\omega) \\
&= \int_\Omega \int_\Omega u_{0,1}(t,\Theta_{\mathrm{t}_i}(\omega)) G(\Theta_{\mathrm{t}_i}(\omega), Z_{\mathrm{j}_i}(\omega'))\, \mathbb{P}(d\omega')\, \mathbb{P}(d\omega) \\
&= \int_\Omega u_{0,1}(t,\Theta_{\mathrm{t}_i}(\omega)) \bigg(\int_\Omega G(\Theta_{\mathrm{t}_i}(\omega), Z_{\mathrm{j}_i}(\omega'))\, \mathbb{P}(d\omega') \bigg) \mathbb{P}(d\omega) \\
&= \int_\Omega u_{0,1}(t,\Theta_{\mathrm{t}_i}(\omega)) \bigg(\int_\Omega G(\Theta_{\mathrm{t}_i}(\omega), Z_1(\omega'))\, \mathbb{P}(d\omega') \bigg) \mathbb{P}(d\omega) \\
&= \int_\Omega u_{0,1}(t,\Theta_{\mathrm{t}_i}(\omega)) \mathbb{E}[G(\Theta_{\mathrm{t}_i}(\omega), Z_1)]\, \mathbb{P}(d\omega) \\
&= \int_\Omega u_{0,1}(t,\Theta_{\mathrm{t}_i}(\omega)) g(\Theta_{\mathrm{t}_i}(\omega))\, \mathbb{P}(d\omega) = \mathbb{E}\big[u_{0,1}(t,\Theta_{\mathrm{t}_i}) g(\Theta_{\mathrm{t}_i})\big].
\end{aligned} \tag{3.200}$$

This and (3.199) imply that

$$\int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_{\llbracket t \rrbracket})(Q_t - g(\Theta_{\llbracket t \rrbracket}))\big]\, dt = 0. \tag{3.201}$$

Next observe that the fundamental theorem of calculus, Lemma 3.2.1, the fact that $\#_E < \infty$, and (3.177) ensure that

$$\begin{aligned}
&\int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)\big(g(\Theta_{\llbracket t \rrbracket}) - g(\Theta_t)\big)\big]\, dt \\
&= -\int_0^T \mathbb{E}\bigg[u_{0,1}(t,\Theta_t) \int_{[\llbracket t \rrbracket, t] \setminus E} g'(\Theta_s) Q_s\, ds\bigg]\, dt \\
&= -\int_0^T \mathbb{E}\bigg[u_{0,1}(t,\Theta_t) \int_{\llbracket t \rrbracket}^t g'(\Theta_s) Q_s\, ds\bigg]\, dt \\
&= -\int_0^T \mathbb{E}\bigg[\int_{\llbracket t \rrbracket}^t u_{0,1}(t,\Theta_t) g'(\Theta_s) Q_s\, ds\bigg]\, dt.
\end{aligned} \tag{3.202}$$

This and Tonelli's theorem assure that

$$\begin{aligned}
&\left|\int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)\big(g(\Theta_{[\![t]\!]})-g(\Theta_t)\big)\big]\,dt\right| \\
&= \left|\int_0^T \mathbb{E}\left[\int_{[\![t]\!]}^t u_{0,1}(t,\Theta_t)g'(\Theta_s)Q_s\,ds\right]dt\right| \\
&\le \int_0^T \mathbb{E}\left[\int_{[\![t]\!]}^t |u_{0,1}(t,\Theta_t)g'(\Theta_s)Q_s|\,ds\right]dt \\
&= \int_0^T \int_{[\![t]\!]}^t \mathbb{E}\big[|u_{0,1}(t,\Theta_t)g'(\Theta_s)Q_s|\big]\,ds\,dt \\
&\le \int_0^T \int_{[\![t]\!]}^t \mathbb{E}\big[\|u_{0,1}(t,\Theta_t)\|_{L(\mathbb{R}^d,\mathbb{R})}\|g'(\Theta_s)Q_s\|_{\mathbb{R}^d}\big]\,ds\,dt.
\end{aligned} \tag{3.203}$$

Item (ii) in Lemma 3.3.9 hence implies that

$$\begin{aligned}
&\left|\int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)\big(g(\Theta_{[\![t]\!]})-g(\Theta_t)\big)\big]\,dt\right| \\
&\le \int_0^T e^{-L(T-t)}\int_{[\![t]\!]}^t \mathbb{E}\big[\|\psi'(\theta_{T-t}^{\Theta_t})\|_{L(\mathbb{R}^d,\mathbb{R})}\|g'(\Theta_s)Q_s\|_{\mathbb{R}^d}\big]\,ds\,dt \\
&\le \int_0^T e^{-L(T-t)}(t-[\![t]\!])\sup_{s\in[0,T]}\mathbb{E}\big[\|\psi'(\theta_{T-t}^{\Theta_t})\|_{L(\mathbb{R}^d,\mathbb{R})}\|g'(\Theta_s)Q_s\|_{\mathbb{R}^d}\big]\,dt.
\end{aligned} \tag{3.204}$$

Next note that the fundamental theorem of calculus for the Bochner integral (see, e.g., [54, Lemma 2.1]) demonstrates that for all $t \in [0,T]$ we have that

$$\begin{aligned}
&u_{0,1}(t,\Theta_t) - u_{0,1}(t,\Theta_{[\![t]\!]}) \\
&= \int_0^1 \big((\tfrac{\partial}{\partial\vartheta}u_{0,1})(t,\lambda\Theta_t+(1-\lambda)\Theta_{[\![t]\!]})\big)(\Theta_t-\Theta_{[\![t]\!]})\,d\lambda.
\end{aligned} \tag{3.205}$$

This, (3.148), and (3.151) ensure that for all $t \in [0,T]$ we have that

$$u_{0,1}(t,\Theta_t) - u_{0,1}(t,\Theta_{[\![t]\!]}) = \int_0^1 \big((\tfrac{\partial}{\partial\vartheta}u_{0,1})(t,a_t^\lambda)\big)\delta_t\,d\lambda. \tag{3.206}$$

Moreover, observe that the chain rule, (3.149), and (3.150) assure that for all $t \in [0,T]$, $\vartheta, y, z \in \mathbb{R}^d$ we have that

$$\begin{aligned}
\big((\tfrac{\partial}{\partial\vartheta}u_{0,1})(t,\vartheta)\big)(y,z) &= \big((\tfrac{\partial^2}{\partial\vartheta^2}u)(t,\vartheta)\big)(y,z) = \big(\tfrac{\partial^2}{\partial\vartheta^2}\psi(\theta_{T-t}^\vartheta)\big)(y,z) \\
&= \big(\tfrac{\partial}{\partial\vartheta}(\psi'(\theta_{T-t}^\vartheta)\,\theta_{T-t}^{1,\vartheta}y)\big)(z) \\
&= \psi''(\theta_{T-t}^\vartheta)(\theta_{T-t}^{1,\vartheta}z,\theta_{T-t}^{1,\vartheta}y) + \psi'(\theta_{T-t}^\vartheta)(\theta_{T-t}^{2,\vartheta}(z,y)).
\end{aligned} \tag{3.207}$$

This and (3.206) imply that for all $t \in [0,T]$ we have that

$$
\begin{aligned}
&(u_{0,1}(t,\Theta_t) - u_{0,1}(t,\Theta_{[\![t]\!]}))\Delta_t \\
&= \int_0^1 \psi''(\theta^{a_t^\lambda}_{T-t})(\theta^{1,a_t^\lambda}_{T-t}\Delta_t, \theta^{1,a_t^\lambda}_{T-t}\delta_t) + \psi'(\theta^{a_t^\lambda}_{T-t})(\theta^{2,a_t^\lambda}_{T-t}(\Delta_t,\delta_t))\,d\lambda.
\end{aligned} \tag{3.208}
$$

Furthermore, note that (3.149) and Lemma 3.3.4 assure that for all $t \in [0,T]$, $\lambda \in [0,1]$ we have that

$$
\begin{aligned}
\big|\psi''(\theta^{a_t^\lambda}_{T-t})(\theta^{1,a_t^\lambda}_{T-t}\Delta_t, \theta^{1,a_t^\lambda}_{T-t}\delta_t)\big| &\le \big\|\psi''(\theta^{a_t^\lambda}_{T-t})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\big\|\theta^{1,a_t^\lambda}_{T-t}\Delta_t\big\|_{\mathbb{R}^d}\big\|\theta^{1,a_t^\lambda}_{T-t}\delta_t\big\|_{\mathbb{R}^d} \\
&\le \big\|\psi''(\theta^{a_t^\lambda}_{T-t})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\big\|\theta^{1,a_t^\lambda}_{T-t}\big\|^2_{L(\mathbb{R}^d,\mathbb{R}^d)}\|\Delta_t\|_{\mathbb{R}^d}\big\|\delta_t\big\|_{\mathbb{R}^d} \\
&\le e^{-2L(T-t)}\big\|\psi''(\theta^{a_t^\lambda}_{T-t})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\|\Delta_t\|_{\mathbb{R}^d}\|\delta_t\|_{\mathbb{R}^d}.
\end{aligned} \tag{3.209}
$$

Next observe that Lemma 3.4.6 and (3.150) ensure that for all $\lambda \in [0,1]$, $t \in [0,T]$ we have that

$$
\begin{aligned}
&\big|\psi'\big(\theta^{a_t^\lambda}_{T-t}\big)\big(\theta^{2,a_t^\lambda}_{T-t}\big(\Delta_t,\delta_t\big)\big)\big| \\
&\le \big\|\psi'(\theta^{a_t^\lambda}_{T-t})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\big\|\theta^{2,a_t^\lambda}_{T-t}(\Delta_t,\delta_t)\big\|_{\mathbb{R}^d} \\
&\le \big\|\psi'(\theta^{a_t^\lambda}_{T-t})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\big\|\theta^{2,a_t^\lambda}_{T-t}\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\|\Delta_t\|_{\mathbb{R}^d}\|\delta_t\|_{\mathbb{R}^d} \\
&\le e^{-L(T-t)}\big\|\psi'(\theta^{a_t^\lambda}_{T-t})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\|\Delta_t\|_{\mathbb{R}^d}\|\delta_t\|_{\mathbb{R}^d}\int_0^{T-t} e^{-Ls}\big\|g''(\theta^{a_t^\lambda}_s)\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,ds.
\end{aligned} \tag{3.210}
$$

Combining this, (3.201), Jensen's inequality, (3.208), and (3.209) demonstrates that

$$\begin{aligned}
&\left|\int_0^T \mathbb{E}[u_{0,1}(t,\Theta_t)\Delta_t]\,dt\right| = \left|\int_0^T \mathbb{E}\big[(u_{0,1}(t,\Theta_t) - u_{0,1}(t,\Theta_{[\![t]\!]}))\Delta_t\big]\,dt\right| \\
&\le \int_0^T \big|\mathbb{E}\big[(u_{0,1}(t,\Theta_t) - u_{0,1}(t,\Theta_{[\![t]\!]}))\Delta_t\big]\big|\,dt \\
&\le \int_0^T \mathbb{E}\big[\big|(u_{0,1}(t,\Theta_t) - u_{0,1}(t,\Theta_{[\![t]\!]}))\Delta_t\big|\big]\,dt \\
&= \int_0^T \mathbb{E}\bigg[\bigg|\int_0^1 \psi''(\theta^{a_t^\lambda}_{T-t})(\theta^{1,a_t^\lambda}_{T-t}\Delta_t, \theta^{1,a_t^\lambda}_{T-t}\delta_t) + \psi'(\theta^{a_t^\lambda}_{T-t})(\theta^{2,a_t^\lambda}_{T-t}(\Delta_t,\delta_t))\,d\lambda\bigg|\bigg]\,dt \qquad (3.211)\\
&\le \int_0^T \mathbb{E}\bigg[\int_0^1 \big|\psi''(\theta^{a_t^\lambda}_{T-t})(\theta^{1,a_t^\lambda}_{T-t}\Delta_t, \theta^{1,a_t^\lambda}_{T-t}\delta_t)\big| + \big|\psi'(\theta^{a_t^\lambda}_{T-t})(\theta^{2,a_t^\lambda}_{T-t}(\Delta_t,\delta_t))\big|\,d\lambda\bigg]\,dt \\
&\le \int_0^T \mathbb{E}\bigg[\int_0^1 e^{-2L(T-t)}\big\|\psi''(\theta^{a_t^\lambda}_{T-t})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\|\Delta_t\|_{\mathbb{R}^d}\|\delta_t\|_{\mathbb{R}^d} \\
&\quad + e^{-L(T-t)}\big\|\psi'(\theta^{a_t^\lambda}_{T-t})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\|\Delta_t\|_{\mathbb{R}^d}\|\delta_t\|_{\mathbb{R}^d}\int_0^{T-t} e^{-Ls}\big\|g''(\theta^{a_t^\lambda}_s)\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,ds\,d\lambda\bigg]\,dt.
\end{aligned}$$

Next note that (3.4) and (3.141) assure that for all $t \in [0,T]$, $\omega \in \Omega$ we have that

$$\begin{aligned}
\|\delta_t(\omega)\|_{\mathbb{R}^d} &= \big\|\Theta_t(\omega) - \Theta_{[\![t]\!]}(\omega)\big\|_{\mathbb{R}^d} \\
&= (t - [\![t]\!])\bigg\|\frac{1}{\#_{\gamma([\![t]\!])}}\sum_{j\in\gamma([\![t]\!])} G(\Theta_{[\![t]\!]}(\omega), Z_j(\omega))\bigg\|_{\mathbb{R}^d} \qquad (3.212)\\
&= (t - [\![t]\!])\,\|Q_t(\omega)\|_{\mathbb{R}^d}\,.
\end{aligned}$$

Combining this and (3.211) proves that

$$\begin{aligned}
&\left|\int_0^T \mathbb{E}[u_{0,1}(t,\Theta_t)\Delta_t]\,dt\right| \\
&\le \int_0^T e^{-L(T-t)}(t-[\![t]\!])\mathbb{E}\bigg[\|\Delta_t\|_{\mathbb{R}^d}\|Q_t\|_{\mathbb{R}^d}\bigg(\int_0^1 e^{-L(T-t)}\big\|\psi''(\theta^{a_t^\lambda}_{T-t})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} \qquad (3.213)\\
&\quad + \big\|\psi'(\theta^{a_t^\lambda}_{T-t})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\int_0^{T-t} e^{-Ls}\big\|g''(\theta^{a_t^\lambda}_s)\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,ds\,d\lambda\bigg)\bigg]\,dt.
\end{aligned}$$

This, (3.197), and (3.204) imply that

$$
\begin{aligned}
&\big|\mathbb{E}[\psi(\Theta_T)] - \psi(\theta_T^\xi)\big| \\
&\leq \left|\int_0^T \mathbb{E}[u_{0,1}(t,\Theta_t)\Delta_t]\,dt\right| + \left|\int_0^T \mathbb{E}\big[u_{0,1}(t,\Theta_t)\big(g(\Theta_{[\![t]\!]}) - g(\Theta_t)\big)\big]\,dt\right| \\
&\leq \int_0^T e^{-L(T-t)}(t-[\![t]\!])\mathbb{E}\bigg[\,\|\Delta_t\|_{\mathbb{R}^d}\,\|Q_t\|_{\mathbb{R}^d}\bigg(\int_0^1 e^{-L(T-t)}\big\|\psi''(\theta_{T-t}^{a_t^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} \\
&\quad + \big\|\psi'(\theta_{T-t}^{a_t^\lambda})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\int_0^{T-t} e^{-Lu}\big\|g''(\theta_u^{a_t^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,du\,d\lambda\bigg)\bigg]\,dt \qquad (3.214)\\
&\quad + \int_0^T e^{-L(T-t)}(t-[\![t]\!])\sup_{v\in[0,T]}\mathbb{E}\Big[\big\|\psi'(\theta_{T-t}^{\Theta_t})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\big\|g'(\Theta_v)Q_v\big\|_{\mathbb{R}^d}\Big]\,dt \\
&= \int_0^T e^{-L(T-t)}(t-[\![t]\!])\bigg(\mathbb{E}\bigg[\,\|\Delta_t\|_{\mathbb{R}^d}\,\|Q_t\|_{\mathbb{R}^d}\bigg(\int_0^1 e^{-L(T-t)}\big\|\psi''(\theta_{T-t}^{a_t^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} \\
&\quad + \big\|\psi'(\theta_{T-t}^{a_t^\lambda})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\int_0^{T-t} e^{-Lu}\big\|g''(\theta_u^{a_t^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,du\,d\lambda\bigg)\bigg] \\
&\quad + \sup_{v\in[0,T]}\mathbb{E}\Big[\big\|\psi'(\theta_{T-t}^{\Theta_t})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\big\|g'(\Theta_v)Q_v\big\|_{\mathbb{R}^d}\Big]\bigg)\,dt.
\end{aligned}
$$

This reveals that

$$
\begin{aligned}
&|\mathbb{E}[\psi(\Theta_T)] - \psi(\theta_T^\xi)| \leq \int_0^T e^{-L(T-t)}(t-[\![t]\!]) \\
&\quad \cdot \sup_{s,v\in[0,T]}\mathbb{E}\bigg[\,\|\Delta_s\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d}\bigg(\int_0^1 e^{-L(T-s)}\big\|\psi''(\theta_{T-s}^{a_s^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} \\
&\quad + \big\|\psi'(\theta_{T-s}^{a_s^\lambda})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\int_0^{T-s} e^{-Lu}\big\|g''(\theta_u^{a_s^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,du\,d\lambda\bigg) \\
&\quad + \big\|\psi'(\theta_{T-s}^{\Theta_s})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\big\|g'(\Theta_v)Q_v\big\|_{\mathbb{R}^d}\bigg]\,dt \qquad (3.215)\\
&= \sup_{s,v\in[0,T]}\mathbb{E}\bigg[\,\|\Delta_s\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d}\bigg(\int_0^1 e^{-L(T-s)}\big\|\psi''(\theta_{T-s}^{a_s^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} \\
&\quad + \big\|\psi'(\theta_{T-s}^{a_s^\lambda})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\int_0^{T-s} e^{-Lu}\big\|g''(\theta_u^{a_s^\lambda})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,du\,d\lambda\bigg) \\
&\quad + \big\|\psi'(\theta_{T-s}^{\Theta_s})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\big\|g'(\Theta_v)Q_v\big\|_{\mathbb{R}^d}\bigg]\int_0^T e^{-L(T-t)}(t-[\![t]\!])\,dt.
\end{aligned}
$$

Combining this, the triangle inequality, (3.151), and (3.172) proves that

$$\begin{aligned}
&|\mathbb{E}[\psi(\Theta_T)] - \psi(\Xi)| \le |\mathbb{E}[\psi(\Theta_T)] - \psi(\theta_T^\xi)| + |\psi(\theta_T^\xi) - \psi(\Xi)| \\
&\le \sup_{s,v\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d}\bigg(\int_0^1 e^{-L(T-s)}\big\|\psi''(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} \\
&\quad + \big\|\psi'(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\int_0^{T-s} e^{-Lu}\big\|g''(\theta_u^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\,du\,d\lambda\bigg) \\
&\quad + \big\|\psi'(\theta_{T-s}^{\Theta_s})\big\|_{L(\mathbb{R}^d,\mathbb{R})}\big\|g'(\Theta_v)Q_v\big\|_{\mathbb{R}^d}\bigg]\int_0^T e^{-L(T-t)}(t-[\![t]\!])\,dt \\
&\quad + \sup\Big\{\big\|\psi'(\lambda\theta_T^\xi + (1-\lambda)\Xi)\big\|_{L(\mathbb{R}^d,\mathbb{R})}\in\mathbb{R}\colon \lambda\in[0,1]\Big\}\big\|\xi-\Xi\big\|_{\mathbb{R}^d}e^{-LT}.
\end{aligned} \tag{3.216}$$

This establishes item (iii). This completes the proof of Proposition 3.7.1. □

3.8 Upper bounds for integrals of certain exponentially decaying functions

Lemma 3.8.1. *Assume Setting 3.1.1, let $L \in [0,\infty)$, let $\mathfrak{t}\colon \mathbb{N} \to [0,\infty)$ be non-decreasing, and assume that*

$$\{t \in [0,\infty)\colon \gamma(t) \neq \emptyset\} = \{\mathfrak{t}_n\colon n \in \mathbb{N}\}. \tag{3.217}$$

Then we have for all $k \in \{2,3,\ldots\}$ that

$$\int_0^{\mathfrak{t}_k} \exp(-L(\mathfrak{t}_k - t))(t - [\![t]\!])\,dt \le \frac{1}{2}\sum_{n=1}^{k-1}\exp(-L(\mathfrak{t}_k - \mathfrak{t}_{n+1}))(\mathfrak{t}_{n+1} - \mathfrak{t}_n)^2. \tag{3.218}$$

Proof of Lemma 3.8.1. First, observe that the assumption that $\forall\, t \in [0,\infty)\colon 0 < \#_{\{s\in[0,t]\colon \gamma(s)\neq\emptyset\}} < \infty$ ensures that

$$\gamma(0) \neq \emptyset. \tag{3.219}$$

This, (3.217), and the assumption that $\mathfrak{t}\colon \mathbb{N} \to [0,\infty)$ is non-decreasing imply that

$$\mathfrak{t}_1 = 0. \tag{3.220}$$

This reveals that for all $k \in \{2, 3, \ldots\}$ it holds that

$$\begin{aligned}
&\int_0^{\mathfrak{t}_k} e^{-L(\mathfrak{t}_k - t)}(t - [\![t]\!])\, dt = \sum_{n=1}^{k-1} \int_{\mathfrak{t}_n}^{\mathfrak{t}_{n+1}} e^{-L(\mathfrak{t}_k - t)}(t - [\![t]\!])\, dt \\
&\leq \sum_{n=1}^{k-1} e^{-L(\mathfrak{t}_k - \mathfrak{t}_{n+1})} \int_{\mathfrak{t}_n}^{\mathfrak{t}_{n+1}} (t - \mathfrak{t}_n)\, dt \\
&= \sum_{n=1}^{k-1} e^{-L(\mathfrak{t}_k - \mathfrak{t}_{n+1})} \left[\frac{1}{2}(t - \mathfrak{t}_n)^2\right]_{t=\mathfrak{t}_n}^{t=\mathfrak{t}_{n+1}} = \frac{1}{2} \sum_{n=1}^{k-1} e^{-L(\mathfrak{t}_k - \mathfrak{t}_{n+1})}(\mathfrak{t}_{n+1} - \mathfrak{t}_n)^2.
\end{aligned} \tag{3.221}$$

The proof of Lemma 3.8.1 is thus completed. □

Lemma 3.8.2. *Let $\nu \in [0, 1)$. Then*

(i) we have for all $l \in \mathbb{N}$ that

$$\sum_{n=1}^{l} \frac{1}{n^\nu} \geq \frac{1}{1-\nu} \left((l+1)^{1-\nu} - 1\right) \tag{3.222}$$

and

(ii) we have for all $l \in \mathbb{N}$ that

$$\sum_{n=1}^{l} \frac{1}{n^\nu} \leq 1 + \frac{1}{1-\nu} \left(l^{1-\nu} - 1\right) = \frac{1}{1-\nu} \left(l^{1-\nu} - \nu\right). \tag{3.223}$$

Proof of Lemma 3.8.2. First, observe that for all $l \in \mathbb{N}$ we have that

$$\begin{aligned}
\sum_{n=1}^{l} \frac{1}{n^\nu} &\geq \sum_{n=1}^{l} \int_n^{n+1} \frac{1}{x^\nu}\, dx = \sum_{n=1}^{l} \left[\frac{1}{1-\nu} x^{1-\nu}\right]_{x=n}^{x=n+1} \\
&= \frac{1}{1-\nu} \sum_{n=1}^{l} \left[(n+1)^{1-\nu} - n^{1-\nu}\right] = \frac{1}{1-\nu} \left((l+1)^{1-\nu} - 1\right).
\end{aligned} \tag{3.224}$$

This proves item (i). Moreover, note that for all $l \in \{2, 3, \ldots\}$ we have that

$$\begin{aligned}
\sum_{n=1}^{l} \frac{1}{n^\nu} &= 1 + \sum_{n=2}^{l} \frac{1}{n^\nu} \leq 1 + \sum_{n=2}^{l} \int_{n-1}^{n} \frac{1}{x^\nu}\, dx = 1 + \sum_{n=2}^{l} \left[\frac{1}{1-\nu} x^{1-\nu}\right]_{x=n-1}^{x=n} \\
&= 1 + \sum_{n=2}^{l} \left[\frac{1}{1-\nu} n^{1-\nu} - \frac{1}{1-\nu}(n-1)^{1-\nu}\right] \\
&= 1 + \frac{1}{1-\nu} \left(l^{1-\nu} - 1\right) = \frac{1}{1-\nu} \left(l^{1-\nu} - \nu\right).
\end{aligned} \tag{3.225}$$

Next observe that

$$\sum_{n=1}^{1} \frac{1}{n^\nu} = 1 = \frac{1}{1-\nu}\left(1^{1-\nu} - \nu\right). \tag{3.226}$$

This and (3.225) establish item (ii). The proof of Lemma 3.8.2 is thus completed. □

Lemma 3.8.3. *Assume Setting 3.1.1, let $L \in [0,\infty)$, $\eta \in (0,\infty)$, $\nu \in [0,1)$, let $\mathfrak{t}\colon \mathbb{N} \to [0,\infty)$ satisfy for all $m \in \{2,3,\ldots\}$ that $\mathfrak{t}_1 = 0$ and*

$$\mathfrak{t}_m = \sum_{n=1}^{m-1} \frac{\eta}{n^\nu}, \tag{3.227}$$

and assume that

$$\{s \in [0,\infty)\colon \gamma(s) \neq \emptyset\} = \{\mathfrak{t}_m \colon m \in \mathbb{N}\}. \tag{3.228}$$

Then we have for all $k \in \{2,3,\ldots\}$ that

$$\begin{aligned} &\int_0^{\mathfrak{t}_k} \exp(-L(\mathfrak{t}_k - t))(t - [\![t]\!])\,dt \\ &\leq \frac{\eta^2 \exp(L\eta)}{2}\left[\sum_{n=1}^{k-2} \frac{\exp\!\left(-\frac{L\eta}{1-\nu}(k^{1-\nu} - n^{1-\nu})\right)}{n^{2\nu}}\right] + \frac{\eta^2 (k-1)^{-2\nu}}{2}. \end{aligned} \tag{3.229}$$

Proof of Lemma 3.8.3. Throughout this proof let $k \in \{2,3,\ldots\}$. Observe that Lemma 3.8.2 implies that for all $n \in \{1,2,\ldots,k-1\}$ we have that

$$\mathfrak{t}_{n+1} \leq \eta + \frac{\eta}{1-\nu}(n^{1-\nu} - 1) \qquad \text{and} \qquad \mathfrak{t}_k \geq \frac{\eta}{1-\nu}\left(k^{1-\nu} - 1\right). \tag{3.230}$$

Lemma 3.8.1 hence ensures that

$$\begin{aligned} &\int_0^{\mathfrak{t}_k} e^{-L(\mathfrak{t}_k - t)}(t - [\![t]\!])\,dt \leq \frac{1}{2}\sum_{n=1}^{k-1} e^{-L(\mathfrak{t}_k - \mathfrak{t}_{n+1})}(\mathfrak{t}_{n+1} - \mathfrak{t}_n)^2 \\ &\leq \frac{1}{2}\left[\sum_{n=1}^{k-2} \exp\!\left(-L\left(\tfrac{\eta}{1-\nu}(k^{1-\nu} - 1) - \left(\eta + \tfrac{\eta}{1-\nu}(n^{1-\nu} - 1)\right)\right)\right)\frac{\eta^2}{n^{2\nu}}\right] + \frac{\eta^2}{2(k-1)^{2\nu}} \\ &= \frac{\eta^2 e^{L\eta}}{2}\left[\sum_{n=1}^{k-2} \frac{\exp\!\left(\frac{-L\eta}{1-\nu}(k^{1-\nu} - n^{1-\nu})\right)}{n^{2\nu}}\right] + \frac{\eta^2 (k-1)^{-2\nu}}{2}. \end{aligned} \tag{3.231}$$

This establishes (3.229). The proof of Lemma 3.8.3 is thus completed. □

Lemma 3.8.4. *Let $c \in (0,\infty)$, $\varepsilon \in (0, 1/2)$, $\alpha = \left(\frac{2-2\varepsilon}{c\varepsilon}\right)^{1/\varepsilon}$ and let $v\colon (0,\infty) \to \mathbb{R}$ satisfy for all $x \in (0,\infty)$ that*

$$v(x) = x^{2\varepsilon - 2}\exp(cx^\varepsilon). \tag{3.232}$$

Then v is non-increasing on $(0,\alpha]$ and non-decreasing on $[\alpha,\infty)$.

Proof of Lemma 3.8.4. First, observe that for all $x \in (0,\infty)$ we have that

$$\begin{aligned} v'(x) &= c\varepsilon x^{\varepsilon-1}\exp(cx^\varepsilon)x^{2\varepsilon-2} + \exp(cx^\varepsilon)(-2+2\varepsilon)x^{2\varepsilon-3} \\ &= c\varepsilon\exp(cx^\varepsilon)x^{3\varepsilon-3} - (2-2\varepsilon)\exp(cx^\varepsilon)x^{2\varepsilon-3} \\ &= \exp(cx^\varepsilon)x^{2\varepsilon-3}\big[c\varepsilon x^\varepsilon - (2-2\varepsilon)\big]. \end{aligned} \tag{3.233}$$

This reveals that

$$\{x \in (0,\infty)\colon v'(x) = 0\} = \{\alpha\}. \tag{3.234}$$

Next note that for all $x \in (0,\infty)$ we have that

$$\begin{aligned} v''(x) &= (c\varepsilon)^2 x^{\varepsilon-1}\exp(cx^\varepsilon)x^{3\varepsilon-3} - c\varepsilon\exp(cx^\varepsilon)(3-3\varepsilon)x^{3\varepsilon-4} \\ &\quad - (2-2\varepsilon)c\varepsilon x^{\varepsilon-1}\exp(cx^\varepsilon)x^{2\varepsilon-3} + (2-2\varepsilon)\exp(cx^\varepsilon)(3-2\varepsilon)x^{2\varepsilon-4} \\ &= (c\varepsilon)^2\exp(cx^\varepsilon)x^{4\varepsilon-4} - c\varepsilon\exp(cx^\varepsilon)(3-3\varepsilon)x^{3\varepsilon-4} \\ &\quad - (2-2\varepsilon)c\varepsilon\exp(cx^\varepsilon)x^{3\varepsilon-4} + (2-2\varepsilon)(3-2\varepsilon)\exp(cx^\varepsilon)x^{2\varepsilon-4} \\ &= (c\varepsilon)^2\exp(cx^\varepsilon)x^{4\varepsilon-4} - (5-5\varepsilon)c\varepsilon\exp(cx^\varepsilon)x^{3\varepsilon-4} \\ &\quad + (2-2\varepsilon)(3-2\varepsilon)\exp(cx^\varepsilon)x^{2\varepsilon-4} \\ &= c\varepsilon\exp(cx^\varepsilon)x^{2\varepsilon-4}\left(c\varepsilon x^{2\varepsilon} - (5-5\varepsilon)x^\varepsilon + \tfrac{(2-2\varepsilon)(3-2\varepsilon)}{c\varepsilon}\right). \end{aligned} \tag{3.235}$$

This implies that

$$\begin{aligned} v''(\alpha) &= c\varepsilon e^{c\alpha^\varepsilon}\alpha^{2\varepsilon-4}\left(c\varepsilon\alpha^{2\varepsilon} - (5-5\varepsilon)\alpha^\varepsilon + \tfrac{(2-2\varepsilon)(3-2\varepsilon)}{c\varepsilon}\right) \\ &= c\varepsilon e^{c\alpha^\varepsilon}\alpha^{2\varepsilon-4}\left(c\varepsilon\left(\tfrac{2-2\varepsilon}{c\varepsilon}\right)^2 - (5-5\varepsilon)\left(\tfrac{2-2\varepsilon}{c\varepsilon}\right) + \tfrac{(2-2\varepsilon)(3-2\varepsilon)}{c\varepsilon}\right) \\ &= e^{c\alpha^\varepsilon}\alpha^{2\varepsilon-4}\big((4-8\varepsilon+4\varepsilon^2) - (10-20\varepsilon+10\varepsilon^2) + (6-10\varepsilon+4\varepsilon^2)\big) \\ &= e^{c\alpha^\varepsilon}\alpha^{2\varepsilon-4}(2\varepsilon-2\varepsilon^2) > 0. \end{aligned} \tag{3.236}$$

Combining this with (3.234) verifies that v is non-increasing on $(0,\alpha]$ and non-decreasing on $[\alpha,\infty)$. The proof of Lemma 3.8.4 is thus completed. □

Lemma 3.8.5. *Let $a \in (0,\infty)$, $\varepsilon \in (0, 1/2)$. Then we have for all $l \in \mathbb{N}$ that*

$$\sum_{n=1}^{l} n^{2\varepsilon-2}\exp(an^\varepsilon) \le e^a + \int_1^{l+1} x^{2\varepsilon-2}\exp(ax^\varepsilon)\,dx. \tag{3.237}$$

Proof of Lemma 3.8.5. Throughout this proof let $v\colon (0,\infty) \to \mathbb{R}$ satisfy for all $x \in (0,\infty)$ that

$$v(x) = x^{2\varepsilon-2}\exp(ax^\varepsilon) \tag{3.238}$$

and let

$$\mathfrak{N} = \max\Big\{ \Big(-\infty, \big|\tfrac{2-2\varepsilon}{a\varepsilon}\big|^{1/\varepsilon} \Big] \cap \mathbb{Z} \Big\}. \tag{3.239}$$

Note that for all $l \in \mathbb{N}$ we have that

$$\sum_{n=1}^{l} n^{2\varepsilon-2} \exp(an^{\varepsilon}) = \sum_{n=1}^{l} v(n) \leq v(1) + \sum_{n=2}^{\min\{\mathfrak{N},l\}} v(n) + \sum_{n=\min\{\mathfrak{N},l\}+1}^{l} v(n). \tag{3.240}$$

Combining this and Lemma 3.8.4 assures that for all $l \in \mathbb{N}$ we have that

$$\begin{aligned}
&\sum_{n=1}^{l} n^{2\varepsilon-2} \exp(an^{\varepsilon}) \\
&\leq v(1) + \sum_{n=2}^{\min\{\mathfrak{N},l\}} \int_{n-1}^{n} v(n)\, dx + \sum_{n=\min\{\mathfrak{N},l\}+1}^{l} \int_{n}^{n+1} v(n)\, dx \\
&\leq v(1) + \sum_{n=2}^{\min\{\mathfrak{N},l\}} \int_{n-1}^{n} v(x)\, dx + \sum_{n=\min\{\mathfrak{N},l\}+1}^{l} \int_{n}^{n+1} v(x)\, dx \\
&= v(1) + \sum_{n=1}^{\min\{\mathfrak{N},l\}-1} \int_{n}^{n+1} v(x)\, dx + \sum_{n=\min\{\mathfrak{N},l\}+1}^{l} \int_{n}^{n+1} v(x)\, dx \\
&\leq v(1) + \sum_{n=1}^{l} \int_{n}^{n+1} v(x)\, dx = v(1) + \int_{1}^{l+1} v(x)\, dx.
\end{aligned} \tag{3.241}$$

This demonstrates (3.237). The proof of Lemma 3.8.5 is thus completed. □

Lemma 3.8.6. *Let $n \in \mathbb{N}$, $a \in [0,\infty)$. Then we have for all $\varepsilon \in (0, 1/2)$, $\lambda \in (0,1]$ that*

$$\int_{1}^{n} x^{2\varepsilon-2} \exp(a(x^{\varepsilon} - n^{\varepsilon}))\, dx \leq \frac{1}{1-2\varepsilon}\big(\exp(a(\lambda^{\varepsilon} n^{\varepsilon} - n^{\varepsilon})) + (\lambda n)^{2\varepsilon-1} \big). \tag{3.242}$$

Proof of Lemma 3.8.6. Observe that for all $\varepsilon \in (0, 1/2)$, $\lambda \in (0, 1]$ we have that

$$\begin{aligned}
&\int_1^n x^{2\varepsilon-2} \exp(a(x^\varepsilon - n^\varepsilon))\, dx \\
&= \int_1^{\lambda n} x^{2\varepsilon-2} \exp(a(x^\varepsilon - n^\varepsilon))\, dx + \int_{\lambda n}^n x^{2\varepsilon-2} \exp(a(x^\varepsilon - n^\varepsilon))\, dx \\
&\leq \exp(a(\lambda^\varepsilon n^\varepsilon - n^\varepsilon)) \int_1^{\lambda n} x^{2\varepsilon-2}\, dx + \int_{\lambda n}^n x^{2\varepsilon-2}\, dx \\
&\leq \exp(a(\lambda^\varepsilon n^\varepsilon - n^\varepsilon)) \int_1^{\infty} x^{2\varepsilon-2}\, dx + \left[\frac{1}{2\varepsilon - 1} x^{2\varepsilon-1}\right]_{x=\lambda n}^{x=\infty} \\
&= \exp(a(\lambda^\varepsilon n^\varepsilon - n^\varepsilon)) \frac{1}{1-2\varepsilon} + \frac{1}{1-2\varepsilon} (\lambda n)^{2\varepsilon-1} \\
&= \frac{1}{1-2\varepsilon} \big(\exp(a(\lambda^\varepsilon n^\varepsilon - n^\varepsilon)) + (\lambda n)^{2\varepsilon-1} \big).
\end{aligned} \tag{3.243}$$

The proof of Lemma 3.8.6 is thus completed. $\square$

Lemma 3.8.7. *Assume Setting 3.1.1, let $L, \eta \in (0, \infty)$, $\varepsilon \in (0, 1/2)$, let $K \colon (0,1) \to (0, \infty]$ satisfy for all $\lambda \in (0,1)$ that*

$$\begin{aligned}
&K(\lambda) \\
&= \sup_{n \in \mathbb{N} \cap [2, \infty)} \left[\frac{\eta^2 \exp(L\eta + \frac{L\eta}{\varepsilon})}{2(1-2\varepsilon)} \Big(n^{1-2\varepsilon} \Big[2 \exp\big(-\tfrac{L\eta}{\varepsilon}(1-\lambda^\varepsilon) n^\varepsilon\big) + (n-1)^{2\varepsilon-2} \Big] + \lambda^{2\varepsilon-1} \Big) \right],
\end{aligned} \tag{3.244}$$

and assume that

$$\{s \in [0, \infty) \colon \gamma(s) \neq \emptyset\} = \{0\} \cup \left\{ \sum_{n=1}^m \frac{\eta}{n^{1-\varepsilon}} \colon m \in \mathbb{N} \right\}. \tag{3.245}$$

Then we have for all $\lambda \in (0,1)$, $k \in \mathbb{N}$ that

$$\int_0^{\sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}}} (t - [\![t]\!]) \exp\big(-L\big(\textstyle\sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}}\big) + Lt\big)\, dt \leq K(\lambda) k^{2\varepsilon-1} < \infty. \tag{3.246}$$

Proof of Lemma 3.8.7. Throughout this proof let $\lambda \in (0,1)$, $k \in \mathbb{N} \cap [2, \infty)$, let $a = \frac{L\eta}{\varepsilon} \in (0, \infty)$, let $v \colon (0, \infty) \to \mathbb{R}$ satisfy for all $x \in (0, \infty)$ that

$$v(x) = x^{2\varepsilon-2} \exp(a x^\varepsilon), \tag{3.247}$$

let $\kappa\colon \mathbb{N}\cap[2,\infty)\to(0,\infty)$ satisfy for all $n\in\mathbb{N}\cap[2,\infty)$ that

$$\kappa(n)=\frac{\eta^2 e^{L\eta+a}}{2(1-2\varepsilon)}\big(n^{1-2\varepsilon}\big[2e^{-a(1-\lambda^\varepsilon)n^\varepsilon}+(n-1)^{2\varepsilon-2}\big]+\lambda^{2\varepsilon-1}\big), \tag{3.248}$$

and let $\mathfrak{t}\colon \mathbb{N}\to[0,\infty)$ satisfy for all $n\in\{2,3,\ldots\}$ that $\mathfrak{t}_1=0$ and

$$\mathfrak{t}_n=\sum_{m=1}^{n-1}\frac{\eta}{m^{1-\varepsilon}}. \tag{3.249}$$

Note that Lemma 3.8.3 implies that

$$\begin{aligned}&\int_0^{\mathfrak{t}_k} e^{-L(\mathfrak{t}_k-t)}(t-[\![t]\!])\,dt\\&\leq\left[\frac{\eta^2e^{L\eta}}{2}\sum_{n=1}^{k-2}\frac{\exp\!\big(-\frac{L\eta}{\varepsilon}(k^\varepsilon-n^\varepsilon)\big)}{n^{2-2\varepsilon}}\right]+\frac{\eta^2(k-1)^{2\varepsilon-2}}{2}\\&=\left[\frac{\eta^2e^{L\eta}}{2}\sum_{n=1}^{k-2}n^{2\varepsilon-2}\exp(a(n^\varepsilon-k^\varepsilon))\right]+\frac{\eta^2(k-1)^{2\varepsilon-2}}{2}.\end{aligned} \tag{3.250}$$

Next observe that Lemma 3.8.5 ensures that

$$\sum_{n=1}^{k-2}n^{2\varepsilon-2}\exp(a(n^\varepsilon-k^\varepsilon))\leq e^{-ak^\varepsilon}\left(e^a+\int_1^{k-1}v(x)\,dx\right). \tag{3.251}$$

Combining this and (3.250) demonstrates that

$$\begin{aligned}\int_0^{\mathfrak{t}_k} e^{-L(\mathfrak{t}_k-t)}(t-[\![t]\!])\,dt&\leq\frac{\eta^2e^{L\eta}}{2}e^{-ak^\varepsilon}\left(e^a+\int_1^{k-1}v(x)\,dx\right)+\frac{\eta^2(k-1)^{2\varepsilon-2}}{2}\\&\leq\frac{\eta^2e^{L\eta-ak^\varepsilon}}{2}\left(e^a+\int_1^{k}v(x)\,dx\right)+\frac{\eta^2(k-1)^{2\varepsilon-2}}{2}.\end{aligned} \tag{3.252}$$

Lemma 3.8.6 hence assures that

$$\begin{aligned}&\int_0^{\mathfrak{t}_k} e^{-L(\mathfrak{t}_k-t)}(t-[\![t]\!])\,dt\\&\leq\frac{\eta^2e^{L\eta-ak^\varepsilon+a}}{2}+\frac{\eta^2e^{L\eta}}{2}\int_1^k x^{2\varepsilon-2}\exp(a(x^\varepsilon-k^\varepsilon))\,dx+\frac{\eta^2(k-1)^{2\varepsilon-2}}{2}\\&\leq\frac{\eta^2e^{L\eta-ak^\varepsilon+a}}{2}+\frac{\eta^2e^{L\eta}}{2(1-2\varepsilon)}\big(\exp(a(\lambda^\varepsilon k^\varepsilon-k^\varepsilon))+(\lambda k)^{2\varepsilon-1}\big)+\frac{\eta^2(k-1)^{2\varepsilon-2}}{2}\\&\leq\frac{\eta^2e^{L\eta+a}}{2(1-2\varepsilon)}\Big(\exp(-ak^\varepsilon)+\exp(-a(1-\lambda^\varepsilon)k^\varepsilon)+(\lambda k)^{2\varepsilon-1}+(k-1)^{2\varepsilon-2}\Big).\end{aligned} \tag{3.253}$$

This reveals that

$$\begin{aligned}
&\int_0^{\mathrm{t}_k} e^{-L(\mathrm{t}_k-t)}(t-[\![t]\!])\,dt \\
&\leq \frac{\eta^2 e^{L\eta+a}}{2(1-2\varepsilon)}\Big(2\exp(-a(1-\lambda^\varepsilon)k^\varepsilon)+(\lambda k)^{2\varepsilon-1}+(k-1)^{2\varepsilon-2}\Big) \\
&= \frac{\eta^2 e^{L\eta+a}k^{2\varepsilon-1}}{2(1-2\varepsilon)}\Big(k^{1-2\varepsilon}\Big[2\exp(-a(1-\lambda^\varepsilon)k^\varepsilon)+(k-1)^{2\varepsilon-2}\Big]+\lambda^{2\varepsilon-1}\Big) \\
&\leq K(\lambda)k^{2\varepsilon-1}.
\end{aligned} \tag{3.254}$$

In addition, note that the fact that

$$\limsup_{n\to\infty}\Big(n^{1-2\varepsilon}\exp(-a(1-\lambda^\varepsilon)n^\varepsilon)+n^{1-2\varepsilon}(n-1)^{2\varepsilon-2}\Big)=0 \tag{3.255}$$

ensures that

$$\limsup_{n\to\infty}\kappa(n)=\frac{\eta^2 e^{L\eta+a}\lambda^{2\varepsilon-1}}{2(1-2\varepsilon)}<\infty. \tag{3.256}$$

This reveals that

$$\sup_{n\in\mathbb{N}\cap[2,\infty)}\kappa(n)<\infty. \tag{3.257}$$

Combining this and (3.254) establishes (3.246). The proof of Lemma 3.8.7 is thus completed. □

3.9 Weak error estimates for SAAs in the case of polynomially decaying learning rates with mini-batches

Corollary 3.9.1. *Assume Setting 3.1.1, assume for all $v,w\in[0,\infty)$ with $v\neq w$ that $\gamma(v)\cap\gamma(w)=\emptyset$, let $\psi\in C^2(\mathbb{R}^d,\mathbb{R})$, $\varepsilon\in(0,1/2)$, $L,\eta\in(0,\infty)$, assume for all $y,z\in\mathbb{R}^d$ that*

$$\langle g(y)-g(z),y-z\rangle_{\mathbb{R}^d}\leq -L\|y-z\|_{\mathbb{R}^d}^2, \tag{3.258}$$

$$\sup_{x\in\mathbb{R}^d}\left(\frac{\mathbb{E}\big[\|G(x,Z_1)\|_{\mathbb{R}^d}^2\big]}{\big[1+\|x\|_{\mathbb{R}^d}\big]^2}+\frac{\big\|\mathbb{E}\big[\big(\frac{\partial}{\partial x}G\big)(x,Z_1)\big]\big\|_{L(\mathbb{R}^d,\mathbb{R}^d)}}{\big[1+\|x\|_{\mathbb{R}^d}\big]}+\|\psi'(x)\|_{L(\mathbb{R}^d,\mathbb{R})}\right)<\infty, \tag{3.259}$$

and

$$\{s \in [0,\infty) \colon \gamma(s) \neq \emptyset\} = \{0\} \cup \left\{ \sum_{n=1}^{m} \frac{\eta}{n^{1-\varepsilon}} \colon m \in \mathbb{N} \right\}, \tag{3.260}$$

let $Q\colon [0,\infty) \times \Omega \to \mathbb{R}^d$ *be the stochastic process which satisfies for all* $t \in [0,\infty)$ *that*

$$Q_t = \tfrac{1}{\#_{\gamma(\llbracket t \rrbracket)}} \textstyle\sum_{j \in \gamma(\llbracket t \rrbracket)} G(\Theta_{\llbracket t \rrbracket}, Z_j), \tag{3.261}$$

let $K\colon (0,1) \to (0,\infty]$ *satisfy for all* $\lambda \in (0,1)$ *that*

$$\begin{aligned} &K(\lambda) \\ &= \sup_{n \in \mathbb{N} \cap [2,\infty)} \left[\frac{\eta^2 \exp(L\eta + \frac{L\eta}{\varepsilon})}{2(1-2\varepsilon)} \Big(n^{1-2\varepsilon} \Big[2 \exp\big(-\tfrac{L\eta}{\varepsilon}(1-\lambda^{\varepsilon}) n^{\varepsilon} \big) + (n-1)^{2\varepsilon-2} \Big] + \lambda^{2\varepsilon-1} \Big) \right], \end{aligned} \tag{3.262}$$

and let $C\colon [0,\infty) \to [0,\infty]$ *satisfy for all* $T \in [0,\infty)$ *that*

$$\begin{aligned} C(T) = \sup_{s,v \in [0,T]} \mathbb{E}\bigg[&\|Q_s - g(\Theta_{\llbracket s \rrbracket})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} \\ &\cdot \bigg(\int_0^1 \exp(-L(T-s)) \|\psi''(\theta_{T-s}^{\lambda\Theta_s + (1-\lambda)\Theta_{\llbracket s \rrbracket}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} \\ &+ \|\psi'(\theta_{T-s}^{\lambda\Theta_s + (1-\lambda)\Theta_{\llbracket s \rrbracket}})\|_{L(\mathbb{R}^d,\mathbb{R})} \int_0^{T-s} \exp(-Lu) \big\| g''(\theta_u^{\lambda\Theta_s + (1-\lambda)\Theta_{\llbracket s \rrbracket}}) \big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} \, du \, d\lambda \bigg) \\ &+ \|\psi'(\theta_{T-s}^{\Theta_s})\|_{L(\mathbb{R}^d,\mathbb{R})} \big\| g'(\Theta_v) Q_v \big\|_{\mathbb{R}^d} \bigg] \end{aligned} \tag{3.263}$$

(cf. item (i) in Lemma 3.4.3). Then

(i) we have that there exists a unique $\Xi \in \mathbb{R}^d$ *which satisfies that*

$$\limsup_{t \to \infty} \|\theta_t^{\xi} - \Xi\|_{\mathbb{R}^d} = 0 \tag{3.264}$$

and

(ii) we have for all $\lambda \in (0,1)$*,* $k \in \mathbb{N}$ *that*

$$\begin{aligned} &\Big| \mathbb{E}[\psi(\Theta_{\sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}}})] - \psi(\Xi) \Big| \\ &\leq k^{2\varepsilon-1} \bigg[K(\lambda) C\big(\textstyle\sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}} \big) + k^{1-2\varepsilon} \exp\big(-L\big(\textstyle\sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}} \big) \big) \\ &\quad \cdot \sup_{\alpha \in [0,1]} \Big(\|\psi'(\alpha \theta^{\xi}_{\sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}}} + (1-\alpha)\Xi)\|_{L(\mathbb{R}^d,\mathbb{R})} \Big) \|\xi - \Xi\|_{\mathbb{R}^d} \bigg]. \end{aligned} \tag{3.265}$$

Proof of Corollary 3.9.1. First, observe that item (ii) in Proposition 3.7.1 implies that there exists a unique $\Xi \in \mathbb{R}^d$ which satisfies that

$$\limsup_{t\to\infty} \|\theta_t^\xi - \Xi\|_{\mathbb{R}^d} = 0. \tag{3.266}$$

This establishes item (i). Next note that item (iii) in Proposition 3.7.1 and Lemma 3.3.7 demonstrate that for all $T \in [0,\infty)$ we have that

$$\begin{aligned}
\big|\mathbb{E}\big[\psi(\Theta_T)\big] - \psi(\Xi)\big| &\leq \sup_{s,v\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}\, \|Q_s\|_{\mathbb{R}^d} \\
&\cdot \bigg(\int_0^1 e^{-L(T-s)} \big\|\psi''(\theta_{T-s}^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} + \big\|\psi'(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L(\mathbb{R}^d,\mathbb{R})} \\
&\cdot \int_0^{T-s} e^{-Lu} \big\|g''(\theta_u^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\, du\, d\lambda\bigg) \\
&+ \big\|\psi'(\theta_{T-s}^{\Theta_s})\big\|_{L(\mathbb{R}^d,\mathbb{R})} \big\|g'(\Theta_v) Q_v\big\|_{\mathbb{R}^d}\bigg] \int_0^T e^{-L(T-t)}(t - [\![t]\!])\, dt \\
&+ \sup\Big\{ \big\|\psi'(\alpha\theta_T^\xi + (1-\alpha)\Xi)\big\|_{L(\mathbb{R}^d,\mathbb{R})} \in \mathbb{R} \colon \alpha \in [0,1]\Big\} \|\xi - \Xi\|_{\mathbb{R}^d} e^{-LT}.
\end{aligned} \tag{3.267}$$

This and (3.263) imply that for all $T \in [0,\infty)$ we have that

$$\begin{aligned}
\big|\mathbb{E}\big[\psi(\Theta_T)\big] - \psi(\Xi)\big| &\leq C(T) \int_0^T e^{-L(T-t)} (t - [\![t]\!])\, dt \\
&+ \sup_{\alpha\in[0,1]} \Big(\big\|\psi'(\alpha\theta_T^\xi + (1-\alpha)\Xi)\big\|_{L(\mathbb{R}^d,\mathbb{R})}\Big) \|\xi - \Xi\|_{\mathbb{R}^d} e^{-LT}.
\end{aligned} \tag{3.268}$$

Lemma 3.8.7 therefore ensures that for all $\lambda \in (0,1)$, $k \in \mathbb{N}$ we have that

$$\begin{aligned}
&\Big|\mathbb{E}[\psi(\Theta_{\sum_{n=1}^{k-1}\frac{\eta}{n^{1-\varepsilon}}})] - \psi(\Xi)\Big| \leq C\big(\textstyle\sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}}\big) K(\lambda) k^{2\varepsilon - 1} \\
&\quad + \sup_{\alpha\in[0,1]} \Big(\|\psi'(\alpha\theta^\xi_{\sum_{n=1}^{k-1}\frac{\eta}{n^{1-\varepsilon}}} + (1-\alpha)\Xi)\|_{L(\mathbb{R}^d,\mathbb{R})}\Big) \|\xi - \Xi\|_{\mathbb{R}^d} e^{-L(\sum_{n=1}^{k-1}\frac{\eta}{n^{1-\varepsilon}})} \\
&= k^{2\varepsilon-1} \bigg[K(\lambda) C\big(\textstyle\sum_{n=1}^{k-1}\frac{\eta}{n^{1-\varepsilon}}\big) + k^{1-2\varepsilon} e^{-L(\sum_{n=1}^{k-1}\frac{\eta}{n^{1-\varepsilon}})} \\
&\quad \cdot \sup_{\alpha\in[0,1]} \Big(\|\psi'(\alpha\theta^\xi_{\sum_{n=1}^{k-1}\frac{\eta}{n^{1-\varepsilon}}} + (1-\alpha)\Xi)\|_{L(\mathbb{R}^d,\mathbb{R})}\Big) \|\xi - \Xi\|_{\mathbb{R}^d}\bigg].
\end{aligned} \tag{3.269}$$

This establishes item (ii). The proof of Corollary 3.9.1 is thus completed. □

Chapter 4

Weak error estimates for SAAs in the case of polynomially decaying learning rates

In this chapter we specialize the weak error analysis for SAAs in the case of general learning rates from Chapter 3 to accomplish weak error estimates for SAAs in the case of polynomially decaying learning rates. In particular, we present and prove in this chapter the main result of this paper, Theorem 4.6.2 in Section 4.6 below, which establishes weak convergence rates for SAAs in the case of polynomially decaying learning rates with mini-batches. In Section 4.7 we apply Theorem 4.6.2 to establish in Corollary 4.7.1 in Section 4.7 below weak convergence rates for SAAs in the case of polynomially decaying learning rates without mini-batches. In Section 4.8 below we illustrate Corollary 4.7.1 by means of an elementary example. Our proof of Theorem 4.6.2 employs the weak error analysis result in Corollary 3.9.1 in Section 3.9 above, the elementary results on suitable sequences of uniformly bounded functions in Section 4.2 below, the elementary result on differentiable functions with bounded derivatives in Lemma 4.5.4 in Section 4.5 below, and the a priori estimates for suitable approximation error constants associated to SAAs in Lemma 4.5.5 in Section 4.5 below. Our proof of Lemma 4.5.5, in turn, uses the result on the possibility of interchanging derivatives and expectations in Lemma 3.2.2 in Section 3.2 above, the a priori estimates for SAAs in the case of general learning rates in Lemma 3.6.3 in Section 3.6 above, the a priori estimates for SAAs in the case of polynomially decaying learning rates in Section 4.3 below, and the a posteriori estimates for conditional variances associated to SAAs in Lemma 4.4.2 in Section 4.4 below. In the scientific literature a posteriori estimates similar to the ones as in Lemma 4.4.2 can, e.g., be found in [53, (3) in Theorem 1.1, (169) in Corollary 3.5, and (217) in Theorem 3.7]). Our proof of Lemma 4.4.2, in turn, employs the elementary growth bound estimates in Lemma 4.4.1 in Section 4.4 below. In the scientific literature similar results to Lemma 4.4.1 can, e.g., be found in Dereich & Müller-Gronbach [34, Remark 2.1]. In Setting 4.1.1 in Section 4.1 below we present a mathematical framework for describing SAAs in the case of polynomially decaying learning rates. In the results of this chapter we frequently employ Setting 4.1.1.

4.1 Mathematical description for SAAs in the case of polynomially decaying learning rates

Setting 4.1.1. *Let $d \in \mathbb{N}$, $\xi, \Xi \in \mathbb{R}^d$, $\varepsilon \in (0, 1/2)$, $\eta, L \in (0,\infty)$, $(\mathfrak{M}_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{N}$, let $(S,\mathcal{S})$ be a measurable space, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $Z_{m,n} \colon \Omega \to S$, $(m,n) \in \mathbb{N}^2$, be i.i.d. random variables, let $\mathfrak{t} \colon \mathbb{N}_0 \to [0,\infty)$ satisfy for all $m \in \mathbb{N}_0$ that $\mathfrak{t}_m = \eta[\sum_{n=1}^m n^{\varepsilon-1}]$, let $G = (G(x,s))_{(x,s)\in\mathbb{R}^d\times S} \colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be $(\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{S})/\mathcal{B}(\mathbb{R}^d)$-measurable, let $g \colon \mathbb{R}^d \to \mathbb{R}^d$ be a function, assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto G(x,s) \in \mathbb{R}^d) \in C^2(\mathbb{R}^d,\mathbb{R}^d)$, assume for all $x, y \in \mathbb{R}^d$ that*

$$\max_{i\in\{1,2\}} \inf_{\delta\in(0,\infty)} \sup_{u\in[-\delta,\delta]^d} \mathbb{E}\Big[\|G(x,Z_{1,1})\|_{\mathbb{R}^d} + \|(\tfrac{\partial^i}{\partial x^i}G)(x+u,Z_{1,1})\|^{1+\delta}_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)}\Big] < \infty, \quad (4.1)$$

$$\langle x-\Xi, g(x)\rangle_{\mathbb{R}^d} \leq -L\|g(x)\|^2_{\mathbb{R}^d}, \qquad \langle x-y, g(x)-g(y)\rangle_{\mathbb{R}^d} \leq -L\|x-y\|^2_{\mathbb{R}^d}, \quad (4.2)$$

and $g(x) = \mathbb{E}[G(x,Z_{1,1})]$ (cf. Corollary 2.2.5), let $\theta^\vartheta \in C([0,\infty),\mathbb{R}^d)$, $\vartheta \in \mathbb{R}^d$, satisfy for all $t \in [0,\infty)$, $\vartheta \in \mathbb{R}^d$ that

$$\theta^\vartheta_t = \vartheta + \int_0^t g(\theta^\vartheta_s)\, ds \quad (4.3)$$

(cf. item (i) in Lemma 2.2.6), and let $\Theta \colon [0,\infty)\times\Omega \to \mathbb{R}^d$ be the stochastic process w.c.s.p. which satisfies for all $m \in \mathbb{N}_0$, $t \in [\mathfrak{t}_m, \mathfrak{t}_{m+1})$ that $\Theta_0 = \xi$ and

$$\Theta_t = \Theta_{\mathfrak{t}_m} + \frac{(t-\mathfrak{t}_m)}{\mathfrak{M}_m}\left[\sum_{n=1}^{\mathfrak{M}_m} G(\Theta_{\mathfrak{t}_m}, Z_{m+1,n})\right]. \quad (4.4)$$

4.2 On a sequence of uniformly bounded functions

Lemma 4.2.1. *Let $d_1, d_2, d_3 \in \mathbb{N}$, $f \in C(\mathbb{R}^{d_1}\times\mathbb{R}^{d_2},\mathbb{R}^{d_3})$ and let $K \subseteq \mathbb{R}^{d_2}$ be a non-empty compact set. Then* *(i) we have for all* $x \in \mathbb{R}^{d_1}$ *that* $\sup_{y\in K} \|f(x,y)\|_{\mathbb{R}^{d_3}} < \infty$ *and*

(ii) we have that $\mathbb{R}^{d_1} \ni x \mapsto \sup_{y\in K} \|f(x,y)\|_{\mathbb{R}^{d_3}} \in \mathbb{R}$ *is continuous.*

Proof of Lemma 4.2.1. Throughout this proof let $g\colon \mathbb{R}^{d_1} \to \mathbb{R} \cup \{\infty\}$ satisfy for all $x \in \mathbb{R}^{d_1}$ that

$$g(x) = \sup_{y\in K} \|f(x,y)\|_{\mathbb{R}^{d_3}}, \tag{4.5}$$

let $x = (x_n)_{n\in\mathbb{N}_0}\colon \mathbb{N}_0 \to \mathbb{R}^{d_1}$ satisfy that

$$\limsup_{n\to\infty} \|x_n - x_0\|_{\mathbb{R}^{d_1}} = 0, \tag{4.6}$$

and let $k\colon \mathbb{N} \to \mathbb{N}$ be strictly increasing. Note that the assumption that f is continuous ensures that for all $z \in \mathbb{R}^{d_1}$ we have that

$$\big(\mathbb{R}^{d_2} \ni y \mapsto f(z,y) \in \mathbb{R}^{d_3}\big) \in C(\mathbb{R}^{d_2}, \mathbb{R}^{d_3}). \tag{4.7}$$

Lemma 2.2.2 and the assumption that K is a non-empty compact set hence establish item (i). Next observe that (4.7) and the assumption that K is a non-empty compact set assure that there exists $y = (y_n)_{n\in\mathbb{N}_0}\colon \mathbb{N}_0 \to K$ which satisfies for all $n \in \mathbb{N}$ that

$$g(x_0) = \|f(x_0, y_0)\|_{\mathbb{R}^{d_3}} \tag{4.8}$$

and

$$g(x_{k(n)}) = \|f(x_{k(n)}, y_n)\|_{\mathbb{R}^{d_3}} \tag{4.9}$$

(see, e.g., Coleman [25, Theorem 1.3]). Furthermore, observe that the assumption that K is a non-empty compact set and the Bolzano-Weierstrass theorem demonstrate that there exist $\mathfrak{y} \in K$ and strictly increasing $l\colon \mathbb{N} \to \mathbb{N}$ which satisfy that

$$\limsup_{n\to\infty} \|y_{l(n)} - \mathfrak{y}\|_{\mathbb{R}^{d_2}} = 0. \tag{4.10}$$

This and (4.6) imply that $(x_{k(l(n))}, y_{l(n)}) \in \mathbb{R}^{d_1} \times \mathbb{R}^{d_2}, n \in \mathbb{N}$, is a convergent sequence in $\mathbb{R}^{d_1} \times \mathbb{R}^{d_2}$. The assumption that f is continuous, (4.6), and (4.9) hence assure that $g(x_{k(l(n))}) \in \mathbb{R}, n \in \mathbb{N}$, is a convergent sequence in $\mathbb{R}$ and

$$\lim_{n\to\infty} g(x_{k(l(n))}) = \lim_{n\to\infty} \|f(x_{k(l(n))}, y_{l(n)})\|_{\mathbb{R}^{d_3}} = \|f(x_0, \mathfrak{y})\|_{\mathbb{R}^{d_3}}. \tag{4.11}$$

This and (4.5) prove that

$$\lim_{n\to\infty} g(x_{k(l(n))}) = \|f(x_0, \mathfrak{y})\|_{\mathbb{R}^{d_3}} \le g(x_0). \tag{4.12}$$

Moreover, note that (4.8), (4.6), the assumption that f is continuous, and (4.5) imply that

$$\begin{aligned} g(x_0) &= \|f(x_0, y_0)\|_{\mathbb{R}^{d_3}} \\ &= \big\|f\big(\lim_{n\to\infty} x_{k(l(n))}, y_0\big)\big\|_{\mathbb{R}^{d_3}} \\ &= \big\|\lim_{n\to\infty} f(x_{k(l(n))}, y_0)\big\|_{\mathbb{R}^{d_3}} \\ &= \lim_{n\to\infty} \|f(x_{k(l(n))}, y_0)\|_{\mathbb{R}^{d_3}} \\ &\le \lim_{n\to\infty} g(x_{k(l(n))}). \end{aligned} \tag{4.13}$$

Combining this and (4.12) assures that

$$\limsup_{n\to\infty} |g(x_{k(l(n))}) - g(x_0)| = 0. \tag{4.14}$$

This and, e.g., [54, Lemma 3.2] prove that the sequence $g(x_n) \in \mathbb{R}$, $n \in \mathbb{N}$, converges to $g(x_0)$. This reveals that g is continuous at x_0. This establishes item (ii). The proof of Lemma 4.2.1 is thus completed. □

Lemma 4.2.2. *Assume Setting 4.1.1 and let $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$. Then*

(i) we have that $\limsup_{t\to\infty} \|\theta_t^\xi - \Xi\|_{\mathbb{R}^d} = 0$ and

(ii) we have that

$$\sup_{n\in\mathbb{N}} \sup_{\lambda\in[0,1]} \Big(n^{1-2\varepsilon} \exp(-L\mathfrak{t}_{n-1}) \|\psi'(\lambda\theta^\xi_{\mathfrak{t}_{n-1}} + (1-\lambda)\Xi)\|_{L(\mathbb{R}^d,\mathbb{R})}\Big) < \infty. \tag{4.15}$$

Proof of Lemma 4.2.2. First, note that (4.2) assures that

$$\{x \in \mathbb{R}^d \colon g(x) = 0\} = \{\Xi\}. \tag{4.16}$$

This and item (iv) in Lemma 3.3.6 establish item (i). In the next step observe that item (i) in Lemma 3.8.2 assures that for all $k \in \{2, 3, \ldots\}$ we have that

$$\mathfrak{t}_{k-1} = \sum_{n=1}^{k-1} \frac{\eta}{n^{1-\varepsilon}} \ge \frac{\eta}{\varepsilon}\big(k^\varepsilon - 1\big). \tag{4.17}$$

This implies that for all $k \in \{2, 3, \ldots\}$ we have that

$$k^{1-2\varepsilon} e^{-L\mathfrak{t}_{k-1}} \le k^{1-2\varepsilon} e^{-\frac{L\eta}{\varepsilon}(k^\varepsilon-1)} = k^{1-2\varepsilon} e^{-\frac{L\eta}{\varepsilon}k^\varepsilon} e^{\frac{L\eta}{\varepsilon}} = (k^\varepsilon)^{\frac{1-2\varepsilon}{\varepsilon}} e^{-\frac{L\eta}{\varepsilon}k^\varepsilon} e^{\frac{L\eta}{\varepsilon}}. \tag{4.18}$$

This reveals that

$$\limsup_{k\to\infty} k^{1-2\varepsilon} e^{-L\mathfrak{t}_{k-1}} = 0. \tag{4.19}$$

Moreover, note that for all $t \in [0,\infty)$ we have that

$$\sup_{\lambda\in[0,1]} \|\psi'(\lambda\theta_t^\xi + (1-\lambda)\Xi)\|_{L(\mathbb{R}^d,\mathbb{R})} = \sup_{\lambda\in[0,1]} \|\nabla\psi(\lambda\theta_t^\xi + (1-\lambda)\Xi)\|_{\mathbb{R}^d}. \tag{4.20}$$

In the next step we combine item (i) and the fact that $\liminf_{k\to\infty} \mathfrak{t}_k = \infty$ to obtain that

$$\limsup_{k\to\infty} \|\theta_{\mathfrak{t}_k}^\xi - \Xi\|_{\mathbb{R}^d} = 0. \tag{4.21}$$

Furthermore, observe that Lemma 4.2.1 and the fact that $\nabla\psi$ is continuous assure that for all $x \in \mathbb{R}^d$ we have that

$$\sup_{\lambda\in[0,1]} \|\nabla\psi(\lambda x + (1-\lambda)\Xi)\|_{\mathbb{R}^d} < \infty \tag{4.22}$$

and

$$\Big(\mathbb{R}^d \ni y \mapsto \sup_{\lambda\in[0,1]} \|\nabla\psi(\lambda y + (1-\lambda)\Xi)\|_{\mathbb{R}^d} \in \mathbb{R}\Big) \in C(\mathbb{R}^d,\mathbb{R}). \tag{4.23}$$

This and (4.21) prove that the sequence

$$\Big(\sup\nolimits_{\lambda\in[0,1]} \|\nabla\psi(\lambda\theta_{\mathfrak{t}_{k-1}}^\xi + (1-\lambda)\Xi)\|_{\mathbb{R}^d}\Big)_{k\in\mathbb{N}} \tag{4.24}$$

is convergent in $\mathbb{R}$ and

$$\begin{aligned}\limsup_{k\to\infty} \sup_{\lambda\in[0,1]} \|\nabla\psi(\lambda\theta_{\mathfrak{t}_{k-1}}^\xi + (1-\lambda)\Xi)\|_{\mathbb{R}^d} &= \sup_{\lambda\in[0,1]} \|\nabla\psi(\lambda\Xi + (1-\lambda)\Xi)\|_{\mathbb{R}^d} \\ &= \|\nabla\psi(\Xi)\|_{\mathbb{R}^d}.\end{aligned} \tag{4.25}$$

This, (4.20), and (4.19) establish (4.15). The proof of Lemma 4.2.2 is thus completed. □

4.3 A priori estimates for SAAs in the case of polynomially decaying learning rates

Lemma 4.3.1. *Assume Setting 4.1.1 and let $\mathbb{G}_t \subseteq \mathcal{F}$, $t \in [0,\infty)$, satisfy for all $t \in (0,\infty)$ that*

$$\mathbb{G}_0 = \{\emptyset, \Omega\} \qquad \text{and} \qquad \mathbb{G}_t = \sigma_\Omega(Z_{m+1,n} \colon (m,n) \in \mathbb{N}_0 \times \mathbb{N},\, n \leq \mathfrak{M}_m,\, \mathfrak{t}_m < t). \tag{4.26}$$

Then we have for all $t \in [0,\infty)$ that Θ_t is $\mathbb{G}_t/\mathcal{B}(\mathbb{R}^d)$-measurable.

Proof of Lemma 4.3.1. First, observe that (4.4) ensures that for all $m \in \mathbb{N}_0$, $t \in (\mathfrak{t}_m, \mathfrak{t}_{m+1})$ we have that

$$\begin{aligned}\Theta_t = \xi + \left[\sum_{n=0}^{m-1} \frac{(\mathfrak{t}_{n+1} - \mathfrak{t}_n)}{\mathfrak{M}_n} \sum_{j=1}^{\mathfrak{M}_n} G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\right] \\ + \frac{(t - \mathfrak{t}_m)}{\mathfrak{M}_m} \sum_{n=1}^{\mathfrak{M}_m} G(\Theta_{\mathfrak{t}_m}, Z_{m+1,n}).\end{aligned} \tag{4.27}$$

Next we claim that for all $m \in \mathbb{N}_0$, $t \in (\mathfrak{t}_m, \mathfrak{t}_{m+1}]$ we have that Θ_t is $\mathbb{G}_t/\mathcal{B}(\mathbb{R}^d)$-measurable. We prove this by induction on $m \in \mathbb{N}_0$. For the base case $m = 0$ note that (4.26) assures that for all $t \in (\mathfrak{t}_0, \mathfrak{t}_1]$ we have that

$$\mathbb{G}_t = \sigma_\Omega(Z_{1,n} \colon n \in \{1, 2, \ldots, \mathfrak{M}_0\}). \tag{4.28}$$

Furthermore, observe that (4.27) and the assumption that $\Theta \colon [0, \infty) \times \Omega \to \mathbb{R}^d$ is a stochastic process w.c.s.p. prove that for all $t \in (\mathfrak{t}_0, \mathfrak{t}_1]$ we have that

$$\Theta_t = \xi + \frac{t}{\mathfrak{M}_0} \sum_{n=1}^{\mathfrak{M}_0} G(\xi, Z_{1,n}). \tag{4.29}$$

This and (4.28) prove that for all $t \in (\mathfrak{t}_0, \mathfrak{t}_1]$ we have that Θ_t is $\mathbb{G}_t/\mathcal{B}(\mathbb{R}^d)$-measurable. For the induction step $\mathbb{N}_0 \ni m \to m+1 \in \mathbb{N}_0$ observe that (4.26) assures that for all $t \in (\mathfrak{t}_{m+1}, \mathfrak{t}_{m+2}]$ we have that

$$\mathbb{G}_t = \sigma_\Omega(Z_{k+1,n} \colon (k, n) \in \mathbb{N}_0 \times \mathbb{N},\, n \le \mathfrak{M}_k,\, k \le m+1). \tag{4.30}$$

Moreover, note that (4.27) and the assumption that $\Theta \colon [0, \infty) \times \Omega \to \mathbb{R}^d$ is a stochastic process w.c.s.p. demonstrate that for all $t \in (\mathfrak{t}_{m+1}, \mathfrak{t}_{m+2}]$ we have that

$$\begin{aligned}\Theta_t = \xi + \left[\sum_{n=0}^{m} \frac{(\mathfrak{t}_{n+1} - \mathfrak{t}_n)}{\mathfrak{M}_n} \sum_{j=1}^{\mathfrak{M}_n} G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\right] \\ + \frac{(t - \mathfrak{t}_{m+1})}{\mathfrak{M}_{m+1}} \sum_{n=1}^{\mathfrak{M}_{m+1}} G(\Theta_{\mathfrak{t}_{m+1}}, Z_{m+2,n}).\end{aligned} \tag{4.31}$$

The induction hypothesis and (4.30) hence imply that for all $t \in (\mathfrak{t}_{m+1}, \mathfrak{t}_{m+2}]$ we have that Θ_t is $\mathbb{G}_t/\mathcal{B}(\mathbb{R}^d)$-measurable. This finishes the proof of the induction step. Induction therefore establishes that for all $m \in \mathbb{N}_0$, $t \in (\mathfrak{t}_m, \mathfrak{t}_{m+1}]$ we have that Θ_t is $\mathbb{G}_t/\mathcal{B}(\mathbb{R}^d)$-measurable. Combining this with the fact that Θ_0 is $\mathbb{G}_0/\mathcal{B}(\mathbb{R}^d)$-measurable ensures that for all $t \in [0, \infty)$ we have that Θ_t is $\mathbb{G}_t/\mathcal{B}(\mathbb{R}^d)$-measurable. The proof of Lemma 4.3.1 is thus completed. □

Lemma 4.3.2. *Assume Setting 4.1.1, let $\mathbb{F}_n \subseteq \mathcal{F}$, $n \in \mathbb{N}_0$, be the sigma-algebras which satisfy for all $n \in \mathbb{N}$ that*

$$\mathbb{F}_0 = \{\emptyset, \Omega\} \qquad \text{and} \qquad \mathbb{F}_n = \sigma_\Omega(Z_{m+1,j} \colon (m,j) \in \mathbb{N}_0 \times \mathbb{N},\, j \leq \mathfrak{M}_m,\, m \leq n-1), \tag{4.32}$$

and assume for all $n \in \mathbb{N}_0$, $j \in \{1, 2, \dots, \mathfrak{M}_n\}$ that

$$\mathbb{E}\big[\|G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\|_{\mathbb{R}^d}\big] < \infty. \tag{4.33}$$

Then we have for all $n \in \mathbb{N}_0$, $A \in \mathbb{F}_n$ that

$$\mathbb{E}\Big[\tfrac{1}{\mathfrak{M}_n} \textstyle\sum_{j=1}^{\mathfrak{M}_n} (G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j}) - g(\Theta_{\mathfrak{t}_n}))\mathbb{1}_A\Big] = 0. \tag{4.34}$$

Proof of Lemma 4.3.2. Throughout this proof let $D \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $n \in \mathbb{N}_0$ that

$$D_n = \frac{1}{\mathfrak{M}_n} \sum_{j=1}^{\mathfrak{M}_n} (G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j}) - g(\Theta_{\mathfrak{t}_n})) \tag{4.35}$$

and let $Y \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $n \in \mathbb{N}_0$ that

$$Y_n = \Theta_{\mathfrak{t}_n}. \tag{4.36}$$

Observe that the assumption that $Z_{m,j}$, $(m,j) \in \mathbb{N}^2$, are i.i.d. random variables and (4.32) ensure that for all $n \in \mathbb{N}_0$, $j \in \{1, 2, \dots, \mathfrak{M}_n\}$ we have that

$$Z_{n+1,j} \quad \text{is independent of} \quad \mathbb{F}_n. \tag{4.37}$$

Moreover, note that Lemma 4.3.1 proves that Y is an $(\mathbb{F}_n)_{n \in \mathbb{N}_0}/\mathcal{B}(\mathbb{R}^d)$-adapted stochastic process. This, (4.37), the fact that $\forall\, x \in \mathbb{R}^d, (n,j) \in \mathbb{N}^2 \colon \mathbb{E}[\|G(x, Z_{n,j})\|_{\mathbb{R}^d}] < \infty$, the fact that $\forall\, x \in \mathbb{R}^d, (n,j) \in \mathbb{N}^2 \colon g(x) = \mathbb{E}[G(x, Z_{n,j})]$, (4.33), and, e.g., [53, Corollary 2.9] establish that for all $n \in \mathbb{N}_0$, $j \in \{1, 2, \dots, \mathfrak{M}_n\}$, $A \in \mathbb{F}_n$ we have that

$$\mathbb{E}[G(Y_n, Z_{n+1,j})\mathbb{1}_A] = \mathbb{E}[g(Y_n)\mathbb{1}_A]. \tag{4.38}$$

Hence, we obtain that for all $n \in \mathbb{N}_0$, $A \in \mathbb{F}_n$ it holds that

$$\begin{aligned}
\mathbb{E}[D_n \mathbb{1}_A] &= \frac{1}{\mathfrak{M}_n} \sum_{j=1}^{\mathfrak{M}_n} \mathbb{E}\big[\big(G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j}) - g(\Theta_{\mathfrak{t}_n})\big)\mathbb{1}_A\big] \\
&= \frac{1}{\mathfrak{M}_n} \sum_{j=1}^{\mathfrak{M}_n} \big(\mathbb{E}[G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\mathbb{1}_A] - \mathbb{E}[g(\Theta_{\mathfrak{t}_n})\mathbb{1}_A]\big) \\
&= \frac{1}{\mathfrak{M}_n} \sum_{j=1}^{\mathfrak{M}_n} \big(\mathbb{E}[G(Y_n, Z_{n+1,j})\mathbb{1}_A] - \mathbb{E}[g(Y_n)\mathbb{1}_A]\big) = 0.
\end{aligned} \tag{4.39}$$

The proof of Lemma 4.3.2 is thus completed. □

Proposition 4.3.3. *Let $d \in \mathbb{N}$, $c, \kappa \in (0,\infty)$, $\Xi \in \mathbb{R}^d$, let $\gamma = (\gamma_n)_{n\in\mathbb{N}_0} \colon \mathbb{N}_0 \to (0,\infty)$ be a function, let $g \colon \mathbb{R}^d \to \mathbb{R}^d$ be $\mathcal{B}(\mathbb{R}^d)/\mathcal{B}(\mathbb{R}^d)$-measurable, let $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_n)_{n\in\mathbb{N}_0})$ be a filtered probability space, let $\Theta \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ be an $(\mathbb{F}_n)_{n\in\mathbb{N}_0}/\mathcal{B}(\mathbb{R}^d)$-adapted stochastic process, and assume for all $x \in \mathbb{R}^d$, $n \in \mathbb{N}$, $A \in \mathbb{F}_{n-1}$ that*

$$\mathbb{E}\big[\|\Theta_0\|_{\mathbb{R}^d}^2 + \|\Theta_n - (\Theta_{n-1} + \gamma_n g(\Theta_{n-1}))\|_{\mathbb{R}^d}^2\big] < \infty, \tag{4.40}$$

$$\mathbb{E}\big[\big(\Theta_n - (\Theta_{n-1} + \gamma_n g(\Theta_{n-1}))\big)\mathbb{1}_A\big] = 0, \tag{4.41}$$

$$\langle x - \Xi, g(x)\rangle_{\mathbb{R}^d} \leq -c \max\big\{\|x - \Xi\|_{\mathbb{R}^d}^2, \|g(x)\|_{\mathbb{R}^d}^2\big\}, \tag{4.42}$$

$$\mathbb{E}\big[\|\Theta_n - (\Theta_{n-1} + \gamma_n g(\Theta_{n-1}))\|_{\mathbb{R}^d}^2\big] \leq (\gamma_n)^2 \kappa\big(1 + \mathbb{E}\big[\|\Theta_{n-1} - \Xi\|_{\mathbb{R}^d}^2\big]\big), \tag{4.43}$$

and

$$\limsup_{k\to\infty} \gamma_k = 0 < \liminf_{k\to\infty} \left[\tfrac{\gamma_k - \gamma_{k-1}}{(\gamma_k)^2} + \tfrac{c\gamma_{k-1}}{2\gamma_k}\right]. \tag{4.44}$$

Then there exists $C \in (0,\infty)$ such that for all $n \in \mathbb{N}_0$ we have that

$$\mathbb{E}\big[\|\Theta_n - \Xi\|_{\mathbb{R}^d}^2\big] \leq C\gamma_n \tag{4.45}$$

and

$$\sup_{m\in\mathbb{N}_0} \mathbb{E}\big[\|\Theta_m - \Xi\|_{\mathbb{R}^d}^2\big] < \infty. \tag{4.46}$$

Proof of Proposition 4.3.3. First, note that [53, Corollary 3.5] (with $d = d$, $(\gamma_n)_{n\in\mathbb{N}_0} = (\gamma_n)_{n\in\mathbb{N}_0}$, $g = g$, $(\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_n)_{n\in\mathbb{N}_0}) = (\Omega, \mathcal{F}, \mathbb{P}, (\mathbb{F}_n)_{n\in\mathbb{N}_0})$, $D_k = {}^{[\Theta_k - (\Theta_{k-1} + \gamma_k g(\Theta_{k-1}))]}/{}_{\gamma_k}$, $\Theta = \Theta$, $\langle\cdot,\cdot\rangle = \langle\cdot,\cdot\rangle_{\mathbb{R}^d}$, $c = c$, $\kappa = \kappa$, $\vartheta = \Xi$ for $k \in \mathbb{N}$ in the notation of [53, Corollary 3.5]) implies that there exists $C \in (0,\infty)$ such that for all $n \in \mathbb{N}_0$ we have that

$$\mathbb{E}\big[\|\Theta_n - \Xi\|_{\mathbb{R}^d}^2\big] \leq C\gamma_n. \tag{4.47}$$

The assumption that $\limsup_{n\to\infty} \gamma_n = 0$ hence ensures that

$$\sup_{m\in\mathbb{N}_0} \mathbb{E}\big[\|\Theta_m - \Xi\|_{\mathbb{R}^d}^2\big] < \infty. \tag{4.48}$$

Combining this with (4.47) completes the proof of Proposition 4.3.3. □

Lemma 4.3.4. *It holds for all $\varepsilon \in (-\infty, 1)$ that*

$$\limsup_{n\to\infty} |n^\varepsilon - (n-1)^\varepsilon| = 0. \tag{4.49}$$

Proof of Lemma 4.3.4. First, observe that for all $\varepsilon \in (-\infty, 0)$ we have that

$$\limsup_{n\to\infty} n^{\varepsilon} = 0. \tag{4.50}$$

This reveals that for all $\varepsilon \in (-\infty, 0)$ it holds that

$$\limsup_{n\to\infty} |n^{\varepsilon} - (n-1)^{\varepsilon}| = 0. \tag{4.51}$$

Next note that for all $\varepsilon \in [0, 1)$, $n \in \mathbb{N} \cap [2, \infty)$ we have that

$$\begin{aligned} 0 \leq |n^{\varepsilon} - (n-1)^{\varepsilon}| &= n^{\varepsilon} - (n-1)^{\varepsilon} = \frac{n^{1-\varepsilon}(n^{\varepsilon} - (n-1)^{\varepsilon})}{n^{1-\varepsilon}} \\ &= \frac{n - n^{1-\varepsilon}(n-1)^{\varepsilon}}{n^{1-\varepsilon}} \leq \frac{n-(n-1)}{n^{1-\varepsilon}} = \frac{1}{n^{1-\varepsilon}} = n^{\varepsilon-1}. \end{aligned} \tag{4.52}$$

This and (4.50) imply that for all $\varepsilon \in [0, 1)$ we have that

$$\limsup_{n\to\infty} |n^{\varepsilon} - (n-1)^{\varepsilon}| = 0. \tag{4.53}$$

Combining this and (4.51) establishes (4.49). The proof of Lemma 4.3.4 is thus completed. ☐

Corollary 4.3.5. *Assume Setting 4.1.1, let $\kappa \in (0, \infty)$, and assume for all $n \in \mathbb{N}$, $m \in \mathbb{N}_0$, $j \in \{1, 2, \ldots, \mathfrak{M}_m\}$ that*

$$\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - (\Theta_{\mathfrak{t}_{n-1}} + (\mathfrak{t}_n - \mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|^2_{\mathbb{R}^d} + \|G(\Theta_{\mathfrak{t}_m}, Z_{m+1,j})\|_{\mathbb{R}^d}\big] < \infty \tag{4.54}$$

and

$$\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - (\Theta_{\mathfrak{t}_{n-1}} + (\mathfrak{t}_n - \mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|^2_{\mathbb{R}^d}\big] \leq (\mathfrak{t}_n - \mathfrak{t}_{n-1})^2 \kappa \big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}} - \Xi\|^2_{\mathbb{R}^d}\big]\big). \tag{4.55}$$

Then we have that

$$\sup_{n\in\mathbb{N}_0} \mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - \Xi\|^2_{\mathbb{R}^d}\big] < \infty. \tag{4.56}$$

Proof of Corollary 4.3.5. Throughout this proof let $\alpha = (\alpha_n)_{n\in\mathbb{N}_0} \colon \mathbb{N}_0 \to (0, \infty)$ satisfy for all $n \in \mathbb{N}$ that

$$\alpha_n = \mathfrak{t}_n - \mathfrak{t}_{n-1} = \frac{\eta}{n^{1-\varepsilon}}, \tag{4.57}$$

let $\mathbb{F}_n \subseteq \mathcal{F}$, $n \in \mathbb{N}_0$, be the sigma-algebras which satisfy for all $n \in \mathbb{N}$ that

$$\mathbb{F}_0 = \{\emptyset, \Omega\} \qquad \text{and} \qquad \mathbb{F}_n = \sigma_{\Omega}(Z_{m+1,j} \colon (m, j) \in \mathbb{N}_0 \times \mathbb{N},\, j \leq \mathfrak{M}_m,\, m \leq n-1), \tag{4.58}$$

and let $Y\colon \mathbb{N}_0\times\Omega\to\mathbb{R}^d$ be the stochastic process which satisfies for all $n\in\mathbb{N}_0$ that

$$Y_n = \Theta_{\mathfrak{t}_n}. \tag{4.59}$$

Note that Lemma 4.3.1 ensures that Y is an $(\mathbb{F}_n)_{n\in\mathbb{N}_0}/\mathcal{B}(\mathbb{R}^d)$-adapted stochastic process. Next observe that (4.54), (4.57), and (4.59) imply that for all $n\in\mathbb{N}$ we have that

$$\begin{aligned}&\mathbb{E}\big[\|Y_n-(Y_{n-1}+\alpha_n g(Y_{n-1}))\|^2_{\mathbb{R}^d}\big]\\&=\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}-(\Theta_{\mathfrak{t}_{n-1}}+(\mathfrak{t}_n-\mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|^2_{\mathbb{R}^d}\big]<\infty.\end{aligned}\tag{4.60}$$

Moreover, note that combining Lemma 4.3.2 and (4.54) assures that for all $n\in\mathbb{N}$, $A\in\mathbb{F}_{n-1}$ we have that

$$\begin{aligned}&\mathbb{E}\big[\big(Y_n-(Y_{n-1}+\alpha_n g(Y_{n-1}))\big)\mathbb{1}_A\big]\\&=\alpha_n\,\mathbb{E}\Big[\tfrac{1}{\mathfrak{M}_{n-1}}\textstyle\sum_{j=1}^{\mathfrak{M}_{n-1}}\big(G(\Theta_{\mathfrak{t}_{n-1}},Z_{n,j})-g(\Theta_{\mathfrak{t}_{n-1}})\big)\mathbb{1}_A\Big]=0.\end{aligned}\tag{4.61}$$

Next observe that (4.2) ensures that $g(\Xi)=0$. This and (4.2) establish that for all $x\in\mathbb{R}^d$ we have that

$$\langle x-\Xi,g(x)\rangle_{\mathbb{R}^d}=\langle x-\Xi,g(x)-g(\Xi)\rangle_{\mathbb{R}^d}\le -L\|x-\Xi\|^2_{\mathbb{R}^d}. \tag{4.62}$$

Combining this with (4.2) implies that for all $x\in\mathbb{R}^d$ we have that

$$\langle x-\Xi,g(x)\rangle_{\mathbb{R}^d}\le -L\max\big\{\|x-\Xi\|^2_{\mathbb{R}^d},\|g(x)\|^2_{\mathbb{R}^d}\big\}. \tag{4.63}$$

Next note that (4.57) assures that for all $n\in\mathbb{N}\cap[2,\infty)$ we have that

$$\begin{aligned}\frac{\alpha_n-\alpha_{n-1}}{(\alpha_n)^2}+\frac{L\alpha_{n-1}}{2\alpha_n}&=\frac{\frac{\eta}{n^{1-\varepsilon}}-\frac{\eta}{(n-1)^{1-\varepsilon}}}{\big(\frac{\eta}{n^{1-\varepsilon}}\big)^2}+\frac{\frac{L\eta}{2(n-1)^{1-\varepsilon}}}{\frac{\eta}{n^{1-\varepsilon}}}\\&=\frac{n^{1-\varepsilon}\big[(n-1)^{1-\varepsilon}-n^{1-\varepsilon}\big]}{\eta(n-1)^{1-\varepsilon}}+\frac{Ln^{1-\varepsilon}}{2(n-1)^{1-\varepsilon}}\\&=\tfrac{1}{\eta}\big[1+\tfrac{1}{n-1}\big]^{1-\varepsilon}\big[(n-1)^{1-\varepsilon}-n^{1-\varepsilon}\big]+\tfrac{L}{2}\big[1+\tfrac{1}{n-1}\big]^{1-\varepsilon}.\end{aligned}\tag{4.64}$$

Lemma 4.3.4 (with $\varepsilon=(1-\varepsilon)\in(-\infty,1)$ in the notation of Lemma 4.3.4) hence ensures that

$$\liminf_{n\to\infty}\left[\frac{\alpha_n-\alpha_{n-1}}{(\alpha_n)^2}+\frac{L\alpha_{n-1}}{2\alpha_n}\right]=\frac{L}{2}>0=\limsup_{n\to\infty}\alpha_n. \tag{4.65}$$

Combining this, (4.60), (4.61), (4.63), and Proposition 4.3.3 (with $d=d$, $\gamma_n=\alpha_n$, $c=L$, $\kappa=\kappa$, $\Xi=\Xi$, $g=g$, $(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_k)_{k\in\mathbb{N}_0})=(\Omega,\mathcal{F},\mathbb{P},(\mathbb{F}_k)_{k\in\mathbb{N}_0})$, $\Theta_n=Y_n$ for $n\in\mathbb{N}_0$ in the notation of Proposition 4.3.3) establishes (4.56). The proof of Corollary 4.3.5 is thus completed. □

4.4 A posteriori estimates for conditional variances associated to SAAs

Lemma 4.4.1. *Let $d \in \mathbb{N}$, $\Xi \in \mathbb{R}^d$, $M, L \in (0, \infty)$ and let $f : \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $x \in \mathbb{R}^d$ that*

$$\langle x - \Xi, f(x) \rangle_{\mathbb{R}^d} \le - \max\{L\|x - \Xi\|^2_{\mathbb{R}^d}, M\|f(x)\|^2_{\mathbb{R}^d}\}. \tag{4.66}$$

Then we have for all $x \in \mathbb{R}^d$ that

$$\begin{aligned} L\|x\|_{\mathbb{R}^d} - L\|\Xi\|_{\mathbb{R}^d} \le L\|x - \Xi\|_{\mathbb{R}^d} \le \|f(x)\|_{\mathbb{R}^d} \\ \le \tfrac{1}{M}\|x - \Xi\|_{\mathbb{R}^d} \le \tfrac{\max\{1, \|\Xi\|_{\mathbb{R}^d}\}}{M}(1 + \|x\|_{\mathbb{R}^d}). \end{aligned} \tag{4.67}$$

Proof of Lemma 4.4.1. First, note that (4.66) assures that for all $x \in \mathbb{R}^d \backslash \{\Xi\}$ we have that

$$f(x) \neq 0 \tag{4.68}$$

and

$$f(\Xi) = 0. \tag{4.69}$$

Furthermore, observe that (4.66) and the Cauchy-Schwartz inequality imply that for all $x \in \mathbb{R}^d$ we have that

$$\|f(x)\|^2_{\mathbb{R}^d} \le -\tfrac{1}{M}\langle x - \Xi, f(x) \rangle_{\mathbb{R}^d} \le \tfrac{1}{M}\|x - \Xi\|_{\mathbb{R}^d}\|f(x)\|_{\mathbb{R}^d}. \tag{4.70}$$

This, (4.68), (4.69), and the triangle inequality ensure that for all $x \in \mathbb{R}^d$ we have that

$$\|f(x)\|_{\mathbb{R}^d} \le \tfrac{1}{M}\|x - \Xi\|_{\mathbb{R}^d} \le \tfrac{1}{M}(\|x\|_{\mathbb{R}^d} + \|\Xi\|_{\mathbb{R}^d}) \le \tfrac{\max\{1, \|\Xi\|_{\mathbb{R}^d}\}}{M}(1 + \|x\|_{\mathbb{R}^d}). \tag{4.71}$$

Moreover, note that (4.66) and the Cauchy-Schwartz inequality demonstrate that for all $x \in \mathbb{R}^d$ we have that

$$\|x - \Xi\|^2_{\mathbb{R}^d} \le -\tfrac{1}{L}\langle x - \Xi, f(x) \rangle_{\mathbb{R}^d} \le \tfrac{1}{L}\|x - \Xi\|_{\mathbb{R}^d}\|f(x)\|_{\mathbb{R}^d}. \tag{4.72}$$

This reveals that for all $x \in \mathbb{R}^d$ it holds that

$$L\|x - \Xi\|_{\mathbb{R}^d} \le \|f(x)\|_{\mathbb{R}^d}. \tag{4.73}$$

This, (4.71), and the triangle inequality establish (4.67). The proof of Lemma 4.4.1 is thus completed. ☐

Lemma 4.4.2. *Assume Setting 4.1.1 and assume that*

$$\sup_{x\in\mathbb{R}^d}\left(\frac{\mathbb{E}\big[\|G(x,Z_{1,1})\|_{\mathbb{R}^d}^2\big]}{\big[1+\|x\|_{\mathbb{R}^d}\big]^2}\right)<\infty. \tag{4.74}$$

Then there exists $\kappa\in(0,\infty)$ such that for all $n\in\mathbb{N}$ we have that

$$\begin{aligned}&\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}-(\Theta_{\mathfrak{t}_{n-1}}+(\mathfrak{t}_n-\mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big]\\&\leq(\mathfrak{t}_n-\mathfrak{t}_{n-1})^2\kappa\big(1+\mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}}-\Xi\|_{\mathbb{R}^d}^2\big]\big)<\infty.\end{aligned} \tag{4.75}$$

Proof of Lemma 4.4.2. First, note that (4.2) ensures that $g(\Xi)=0$. This and (4.2) establish that for all $x\in\mathbb{R}^d$ we have that

$$\langle x-\Xi,g(x)\rangle_{\mathbb{R}^d}\leq-\max\{L\|x-\Xi\|_{\mathbb{R}^d}^2,L\|g(x)\|_{\mathbb{R}^d}^2\}. \tag{4.76}$$

Lemma 4.4.1 hence assures that for all $x\in\mathbb{R}^d$ we have that

$$\|g(x)\|_{\mathbb{R}^d}\leq\tfrac{\max\{1,\|\Xi\|_{\mathbb{R}^d}\}}{L}(1+\|x\|_{\mathbb{R}^d}). \tag{4.77}$$

Next observe that (4.74) and the fact that $\forall\, a,b\in\mathbb{R}\colon(a+b)^2\leq 2|a|^2+2|b|^2$ imply that

$$\sup_{x\in\mathbb{R}^d}\left(\frac{\mathbb{E}\big[\|G(x,Z_{1,1})\|_{\mathbb{R}^d}^2\big]}{1+\|x\|_{\mathbb{R}^d}^2}\right)<\infty. \tag{4.78}$$

This and (4.77) demonstrate that there exists $c\in(0,\infty)$ which satisfies for all $x\in\mathbb{R}^d$ that

$$\|g(x)\|_{\mathbb{R}^d}\leq c(1+\|x\|_{\mathbb{R}^d})\quad\text{and}\quad\mathbb{E}\big[\|G(x,Z_{1,1})\|_{\mathbb{R}^d}^2\big]\leq c(1+\|x\|_{\mathbb{R}^d}^2). \tag{4.79}$$

Moreover, note that (4.4) and the fact that $\forall\, x,y\in\mathbb{R}^d\colon\|x+y\|_{\mathbb{R}^d}^2\leq 2\|x\|_{\mathbb{R}^d}^2+2\|y\|_{\mathbb{R}^d}^2$ ensure that for all $n\in\mathbb{N}$ we have that

$$\begin{aligned}&\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}-(\Theta_{\mathfrak{t}_{n-1}}+(\mathfrak{t}_n-\mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big]\\&=(\mathfrak{t}_n-\mathfrak{t}_{n-1})^2\,\mathbb{E}\Big[\Big\|\tfrac{1}{\mathfrak{M}_{n-1}}\big[\textstyle\sum_{j=1}^{\mathfrak{M}_{n-1}}G(\Theta_{\mathfrak{t}_{n-1}},Z_{n,j})\big]-g(\Theta_{\mathfrak{t}_{n-1}})\Big\|_{\mathbb{R}^d}^2\Big]\\&\leq 2(\mathfrak{t}_n-\mathfrak{t}_{n-1})^2\,\mathbb{E}\Big[\Big\|\tfrac{1}{\mathfrak{M}_{n-1}}\textstyle\sum_{j=1}^{\mathfrak{M}_{n-1}}G(\Theta_{\mathfrak{t}_{n-1}},Z_{n,j})\Big\|_{\mathbb{R}^d}^2\Big]\\&\quad+2(\mathfrak{t}_n-\mathfrak{t}_{n-1})^2\,\mathbb{E}\big[\|g(\Theta_{\mathfrak{t}_{n-1}})\|_{\mathbb{R}^d}^2\big].\end{aligned} \tag{4.80}$$

This, (4.79), and Lemma 3.6.2 assure that for all $n\in\mathbb{N}$ we have that

$$\begin{aligned}&\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}-(\Theta_{\mathfrak{t}_{n-1}}+(\mathfrak{t}_n-\mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big]\\&\leq 2(\mathfrak{t}_n-\mathfrak{t}_{n-1})^2\Big(c\big(1+\mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}^2\big]\big)+\mathbb{E}\big[(c(1+\|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}))^2\big]\Big).\end{aligned} \tag{4.81}$$

Lemma 3.6.1 hence implies that for all $n \in \mathbb{N}$ we have that

$$\begin{aligned}
&\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - (\Theta_{\mathfrak{t}_{n-1}} + (\mathfrak{t}_n - \mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big] \\
&\leq 2(\mathfrak{t}_n - \mathfrak{t}_{n-1})^2\Big(c\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}^2\big]\big) + \mathbb{E}\big[2c^2\big(1 + \|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}^2\big)\big]\Big).
\end{aligned} \tag{4.82}$$

Hence, we obtain that for all $n \in \mathbb{N}$ it holds that

$$\begin{aligned}
&\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - (\Theta_{\mathfrak{t}_{n-1}} + (\mathfrak{t}_n - \mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big] \\
&\leq 2(\mathfrak{t}_n - \mathfrak{t}_{n-1})^2\Big(c\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}^2\big]\big) + 2c^2\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}^2\big]\big)\Big) \\
&= 2(\mathfrak{t}_n - \mathfrak{t}_{n-1})^2(c + 2c^2)\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}^2\big]\big).
\end{aligned} \tag{4.83}$$

The fact that $\forall\, x, y \in \mathbb{R}^d \colon \|x + y\|_{\mathbb{R}^d}^2 \leq 2\|x\|_{\mathbb{R}^d}^2 + 2\|y\|_{\mathbb{R}^d}^2$ therefore demonstrates that for all $n \in \mathbb{N}$ we have that

$$\begin{aligned}
&\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - (\Theta_{\mathfrak{t}_{n-1}} + (\mathfrak{t}_n - \mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big] \\
&\leq 2(\mathfrak{t}_n - \mathfrak{t}_{n-1})^2(c + 2c^2)\big(1 + 2\|\Xi\|_{\mathbb{R}^d}^2 + 2\mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}} - \Xi\|_{\mathbb{R}^d}^2\big]\big) \\
&\leq (\mathfrak{t}_n - \mathfrak{t}_{n-1})^2\big[2(c + 2c^2)(2 + 2\|\Xi\|_{\mathbb{R}^d}^2)\big]\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}} - \Xi\|_{\mathbb{R}^d}^2\big]\big).
\end{aligned} \tag{4.84}$$

Next note that (4.74) and Lemma 3.6.3 imply that for all $n \in \mathbb{N}_0$ we have that

$$\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^2\big] < \infty. \tag{4.85}$$

The fact that $\forall\, x, y \in \mathbb{R}^d \colon \|x + y\|_{\mathbb{R}^d}^2 \leq 2\|x\|_{\mathbb{R}^d}^2 + 2\|y\|_{\mathbb{R}^d}^2$ hence ensures that for all $n \in \mathbb{N}$ we have that

$$\mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}} - \Xi\|_{\mathbb{R}^d}^2\big] \leq 2\big(\mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}}\|_{\mathbb{R}^d}^2\big] + \|\Xi\|_{\mathbb{R}^d}^2\big) < \infty. \tag{4.86}$$

Combining this and (4.84) establishes (4.75). The proof of Lemma 4.4.2 is thus completed. □

4.5 A priori estimates for suitable approximation error constants associated to SAAs

Lemma 4.5.1. *Assume Setting 4.1.1 and let $p, \mathfrak{m} \in \{0\} \cup [1, \infty)$ satisfy that*

$$\sup_{x \in \mathbb{R}^d}\left(\frac{\mathbb{E}\big[\|G(x, Z_{1,1})\|_{\mathbb{R}^d}^p\big]}{\big[1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{m}}\big]^p}\right) + \sup_{n \in \mathbb{N}_0} \mathbb{E}\big[\,\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^{\mathfrak{m}p}\,\big] < \infty. \tag{4.87}$$

Then we have that

$$\sup_{n \in \mathbb{N}_0} \mathbb{E}\Big[\big\|\tfrac{1}{\mathfrak{M}_n}\textstyle\sum_{j=1}^{\mathfrak{M}_n} G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\big\|_{\mathbb{R}^d}^p\Big] < \infty. \tag{4.88}$$

Proof of Lemma 4.5.1. First, observe that (4.87) and Lemma 3.6.1 imply that there exists $c \in [0, \infty)$ which satisfies for all $x \in \mathbb{R}^d$ that

$$\mathbb{E}\big[\|G(x, Z_{1,1})\|_{\mathbb{R}^d}^p\big] \le c(1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{m}p}). \tag{4.89}$$

Lemma 3.6.2 therefore assures that for all $n \in \mathbb{N}_0$ we have that

$$\mathbb{E}\Big[\big\|\tfrac{1}{\mathfrak{M}_n}\textstyle\sum_{j=1}^{\mathfrak{M}_n} G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\big\|_{\mathbb{R}^d}^p\Big] \le c\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^{\mathfrak{m}p}\big]\big). \tag{4.90}$$

Combining this and (4.87) establishes (4.88). The proof of Lemma 4.5.1 is thus completed. □

Lemma 4.5.2. *Assume Setting 4.1.1 and let $p, \mathfrak{m} \in \{0\} \cup [1, \infty)$ satisfy that*

$$\sup_{x \in \mathbb{R}^d} \left(\frac{\|g(x)\|_{\mathbb{R}^d}}{1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{m}}} \right) + \sup_{n \in \mathbb{N}_0} \mathbb{E}\big[\, \|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^{\mathfrak{m}p} \,\big] < \infty. \tag{4.91}$$

Then we have that

$$\sup_{t \in [0,\infty)} \mathbb{E}\big[\|g(\Theta_t)\|_{\mathbb{R}^d}^p\big] < \infty. \tag{4.92}$$

Proof of Lemma 4.5.2. First, observe that (4.91) implies that there exists $c \in [0, \infty)$ which satisfies for all $x \in \mathbb{R}^d$ that

$$\|g(x)\|_{\mathbb{R}^d} \le c(1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{m}}). \tag{4.93}$$

This and Lemma 3.6.1 ensure that for all $t \in [0, \infty)$ we have that

$$\begin{aligned}\mathbb{E}\big[\|g(\Theta_t)\|_{\mathbb{R}^d}^p\big] &\le \mathbb{E}\big[(c(1 + \|\Theta_t\|_{\mathbb{R}^d}^{\mathfrak{m}}))^p\big] \le \mathbb{E}\big[2^{p-1}c^p(1 + \|\Theta_t\|_{\mathbb{R}^d}^{\mathfrak{m}p})\big] \\ &= 2^{p-1}c^p + 2^{p-1}c^p\,\mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^{\mathfrak{m}p}\big].\end{aligned} \tag{4.94}$$

Lemma 3.6.5 and (4.91) hence demonstrate that

$$\sup_{t \in [0,\infty)} \mathbb{E}\big[\|g(\Theta_t)\|_{\mathbb{R}^d}^p\big] \le 2^{p-1}c^p + 2^{p-1}c^p \sup_{t \in [0,\infty)} \mathbb{E}\big[\|\Theta_t\|_{\mathbb{R}^d}^{\mathfrak{m}p}\big] < \infty. \tag{4.95}$$

This establishes (4.92). The proof of Lemma 4.5.2 is thus completed. □

Lemma 4.5.3. *Assume Setting 4.1.1 and let $p, \mathfrak{m} \in \{0\} \cup [1, \infty)$ satisfy that*

$$\sup_{x \in \mathbb{R}^d} \left(\frac{\big\|\mathbb{E}\big[(\tfrac{\partial}{\partial x} G)(x, Z_{1,1})\big]\big\|_{L(\mathbb{R}^d, \mathbb{R}^d)}}{1 + \|x\|_{\mathbb{R}^d}^{\mathfrak{m}}} \right) + \sup_{n \in \mathbb{N}_0} \mathbb{E}\big[\, \|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^{\mathfrak{m}p} \,\big] < \infty. \tag{4.96}$$

Then

(i) we have that $g \in C^1(\mathbb{R}^d, \mathbb{R}^d)$ and

(ii) we have that

$$\sup_{t \in [0,\infty)} \mathbb{E}\big[\|g'(\Theta_t)\|^p_{L(\mathbb{R}^d,\mathbb{R}^d)}\big] < \infty. \tag{4.97}$$

Proof of Lemma 4.5.3. First, note that item (i) in Lemma 3.2.1 proves item (i). Furthermore, observe that (4.96) and item (ii) in Lemma 3.2.1 demonstrate that there exists $c \in [0, \infty)$ which satisfies for all $x \in \mathbb{R}^d$ that

$$\|g'(x)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} = \big\|\mathbb{E}\big[(\tfrac{\partial}{\partial x} G)(x, Z_{1,1})\big]\big\|_{L(\mathbb{R}^d,\mathbb{R}^d)} \leq c(1 + \|x\|^{\mathfrak{m}}_{\mathbb{R}^d}). \tag{4.98}$$

This and Lemma 3.6.1 imply that for all $t \in [0, \infty)$ we have that

$$\begin{aligned}
\mathbb{E}\big[\|g'(\Theta_t)\|^p_{L(\mathbb{R}^d,\mathbb{R}^d)}\big] &\leq \mathbb{E}\big[(c(1 + \|\Theta_t\|^{\mathfrak{m}}_{\mathbb{R}^d}))^p\big] \\
&\leq \mathbb{E}\big[2^{p-1} c^p (1 + \|\Theta_t\|^{\mathfrak{m}p}_{\mathbb{R}^d})\big] \\
&= 2^{p-1} c^p + 2^{p-1} c^p \, \mathbb{E}\big[\|\Theta_t\|^{\mathfrak{m}p}_{\mathbb{R}^d}\big].
\end{aligned} \tag{4.99}$$

Lemma 3.6.5 and (4.96) hence demonstrate that

$$\sup_{t \in [0,\infty)} \mathbb{E}\big[\|g'(\Theta_t)\|^p_{L(\mathbb{R}^d,\mathbb{R}^d)}\big] \leq 2^{p-1} c^p + 2^{p-1} c^p \sup_{t \in [0,\infty)} \mathbb{E}\big[\|\Theta_t\|^{\mathfrak{m}p}_{\mathbb{R}^d}\big] < \infty. \tag{4.100}$$

The proof of Lemma 4.5.3 is thus completed. □

Lemma 4.5.4. *Let $(V, \|\cdot\|_V)$ be a non-trivial $\mathbb{R}$-Banach space, let $(W, \|\cdot\|_W)$ be an $\mathbb{R}$-Banach space, and let $f \in C^1(V, W)$, $c \in \mathbb{R}$ satisfy for all $x \in V$ that*

$$\|f'(x)\|_{L(V,W)} \leq c. \tag{4.101}$$

Then we have for all $x \in V$ that

$$\|f(x)\|_W \leq (c + \|f(0)\|_W)(1 + \|x\|_V). \tag{4.102}$$

Proof of Lemma 4.5.4. First, note that the fundamental theorem of calculus for the Bochner integral (see, e.g., [54, Lemma 2.1]) proves that for all $x \in V$ we have that

$$\|f(x) - f(0)\|_W = \left\| \int_0^1 f'(\lambda x) x \, d\lambda \right\|_W. \tag{4.103}$$

This and the triangle inequality for the Bochner integral demonstrate that for all $x \in V$ we have that

$$\begin{aligned} \|f(x) - f(0)\|_W \leq \int_0^1 \|f'(\lambda x)x\|_W \, d\lambda &\leq \int_0^1 \|f'(\lambda x)\|_{L(V,W)} \|x\|_V \, d\lambda \\ &\leq \int_0^1 c\|x\|_V \, d\lambda = c\|x\|_V. \end{aligned} \tag{4.104}$$

This reveals that for all $x \in V$ it holds that

$$\begin{aligned} \|f(x)\|_W &\leq \|f(x) - f(0)\|_W + \|f(0)\|_W \\ &\leq c\|x\|_V + \|f(0)\|_W \\ &\leq (c + \|f(0)\|_W)(1 + \|x\|_V). \end{aligned} \tag{4.105}$$

This establishes (4.102). The proof of Lemma 4.5.4 is thus completed. □

Lemma 4.5.5. *Assume Setting 4.1.1, let* $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$ *satisfy that*

$$\sup_{x \in \mathbb{R}^d} \left(\frac{\mathbb{E}\big[\|G(x, Z_{1,1})\|_{\mathbb{R}^d}^2\big]}{\big[1 + \|x\|_{\mathbb{R}^d}\big]^2} + \big\|\mathbb{E}\big[(\tfrac{\partial^2}{\partial x^2} G)(x, Z_{1,1})\big]\big\|_{L^{(2)}(\mathbb{R}^d, \mathbb{R}^d)} \right) < \infty \tag{4.106}$$

and $\sup_{x \in \mathbb{R}^d} \max_{i \in \{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^d, \mathbb{R})} < \infty$, *let* $Q \colon [0, \infty) \times \Omega \to \mathbb{R}^d$ *be the stochastic process which satisfies for all* $m \in \mathbb{N}_0$, $t \in [\mathfrak{t}_m, \mathfrak{t}_{m+1})$ *that*

$$Q_t = \frac{1}{\mathfrak{M}_m} \sum_{n=1}^{\mathfrak{M}_m} G(\Theta_{\mathfrak{t}_m}, Z_{m+1,n}), \tag{4.107}$$

and let $C \colon [0, \infty) \to [0, \infty]$ *satisfy for all* $T \in [0, \infty)$ *that*

$$\begin{aligned} C(T) = \sup_{s,v \in [0,T]} \mathbb{E}\bigg[&\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} \\ &\cdot \bigg(\int_0^1 e^{-L(T-s)} \|\psi''(\theta_{T-s}^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d, \mathbb{R})} + \|\psi'(\theta_{T-s}^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\|_{L(\mathbb{R}^d, \mathbb{R})} \\ &\cdot \int_0^{T-s} e^{-Lu} \big\|g''(\theta_u^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L^{(2)}(\mathbb{R}^d, \mathbb{R}^d)} \, du \, d\lambda \bigg) + \|\psi'(\theta_{T-s}^{\Theta_s})\|_{L(\mathbb{R}^d, \mathbb{R})} \|g'(\Theta_v) Q_v\|_{\mathbb{R}^d} \bigg]. \end{aligned} \tag{4.108}$$

Then we have that

$$\sup_{T \in [0, \infty)} C(T) < \infty. \tag{4.109}$$

Proof of Lemma 4.5.5. Throughout this proof let $\kappa \in (0,\infty)$ be a real number which satisfies for all $n \in \mathbb{N}$ that

$$\begin{aligned} &\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - (\Theta_{\mathfrak{t}_{n-1}} + (\mathfrak{t}_n - \mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big] \\ &\leq (\mathfrak{t}_n - \mathfrak{t}_{n-1})^2\kappa\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_{n-1}} - \Xi\|_{\mathbb{R}^d}^2\big]\big) \end{aligned} \tag{4.110}$$

(cf. Lemma 4.4.2). Note that Lemma 3.2.2 assures that for all $x \in \mathbb{R}^d$ we have that

$$g \in C^2(\mathbb{R}^d, \mathbb{R}^d) \qquad \text{and} \qquad g''(x) = \mathbb{E}\big[\big(\tfrac{\partial^2}{\partial x^2}G\big)(x, Z_{1,1})\big]. \tag{4.111}$$

This, (4.106), and Lemma 4.4.1 prove that there exists $c \in [0,\infty)$ which satisfies for all $x \in \mathbb{R}^d$ that

$$\mathbb{E}\big[\|G(x, Z_{1,1})\|_{\mathbb{R}^d}^2\big] \leq c(1 + \|x\|_{\mathbb{R}^d})^2, \qquad \|g(x)\|_{\mathbb{R}^d} \leq c(1 + \|x\|_{\mathbb{R}^d}), \tag{4.112}$$

and

$$\max\big\{\|g''(x)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}, \|\psi'(x)\|_{L(\mathbb{R}^d,\mathbb{R})}, \|\psi''(x)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\big\} \leq c. \tag{4.113}$$

Lemma 4.5.4 hence implies that for all $x \in \mathbb{R}^d$ we have that

$$\|g'(x)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} \leq (c + \|g'(0)\|_{L(\mathbb{R}^d,\mathbb{R}^d)})(1 + \|x\|_{\mathbb{R}^d}). \tag{4.114}$$

Moreover, note that Lemma 4.4.2 proves that for all $n \in \mathbb{N}$ we have that

$$\mathbb{E}\big[\|\Theta_{\mathfrak{t}_n} - (\Theta_{\mathfrak{t}_{n-1}} + (\mathfrak{t}_n - \mathfrak{t}_{n-1})g(\Theta_{\mathfrak{t}_{n-1}}))\|_{\mathbb{R}^d}^2\big] < \infty. \tag{4.115}$$

Next observe that the assumption that $Z_{m,n}$, $(m,n) \in \mathbb{N}^2$, are i.i.d. random variables and (4.4) ensure that for all $n \in \mathbb{N}_0$, $j \in \{1, 2, \ldots, \mathfrak{M}_n\}$ we have that $Z_{n+1,j}$ and $\Theta_{\mathfrak{t}_n}$ are independent. This and the assumption that $Z_{m,n}$, $(m,n) \in \mathbb{N}^2$, are i.i.d. random variables ensure that for all $n \in \mathbb{N}_0$, $j \in \{1, 2, \ldots, \mathfrak{M}_n\}$ we have that

$$\begin{aligned} \mathbb{E}\big[\|G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\|_{\mathbb{R}^d}^2\big] &= \int_\Omega \|G(\Theta_{\mathfrak{t}_n}(\omega), Z_{n+1,j}(\omega))\|_{\mathbb{R}^d}^2 \,\mathbb{P}(d\omega) \\ &= \int_\Omega \int_\Omega \|G(\Theta_{\mathfrak{t}_n}(\omega), Z_{n+1,j}(\tilde{\omega}))\|_{\mathbb{R}^d}^2 \,\mathbb{P}(d\tilde{\omega})\,\mathbb{P}(d\omega) \\ &= \int_\Omega \int_\Omega \|G(\Theta_{\mathfrak{t}_n}(\omega), Z_{1,1}(\tilde{\omega}))\|_{\mathbb{R}^d}^2 \,\mathbb{P}(d\tilde{\omega})\,\mathbb{P}(d\omega). \end{aligned} \tag{4.116}$$

Combining this with (4.112) and Lemma 3.6.3 demonstrates that for all $n \in \mathbb{N}_0$, $j \in \{1, 2, \ldots, \mathfrak{M}_n\}$ we have that

$$\begin{aligned} \mathbb{E}\big[\|G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\|_{\mathbb{R}^d}^2\big] &\leq \int_\Omega c\big(1 + \|\Theta_{\mathfrak{t}_n}(\omega)\|_{\mathbb{R}^d}\big)^2 \,\mathbb{P}(d\omega) \\ &\leq 2c\big(1 + \mathbb{E}\big[\|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^2\big]\big) < \infty. \end{aligned} \tag{4.117}$$

This reveals that for all $n \in \mathbb{N}_0$, $j \in \{1, 2, \ldots, \mathfrak{M}_n\}$ it holds that

$$\mathbb{E}\big[\|G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\|_{\mathbb{R}^d}\big] \leq \big|\mathbb{E}\big[\|G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\|_{\mathbb{R}^d}^2\big]\big|^{1/2} < \infty. \tag{4.118}$$

This and Corollary 4.3.5 imply that

$$\sup_{n\in\mathbb{N}_0} \big|\mathbb{E}\big[\, \|\Theta_{\mathfrak{t}_n} - \Xi\|_{\mathbb{R}^d}^2 \,\big]\big|^{1/2} < \infty. \tag{4.119}$$

The Minkowski inequality hence assures that

$$\sup_{n\in\mathbb{N}_0} \big|\mathbb{E}\big[\, \|\Theta_{\mathfrak{t}_n}\|_{\mathbb{R}^d}^2 \,\big]\big|^{1/2} \leq \|\Xi\|_{\mathbb{R}^d} + \sup_{n\in\mathbb{N}_0} \big|\mathbb{E}\big[\, \|\Theta_{\mathfrak{t}_n} - \Xi\|_{\mathbb{R}^d}^2 \,\big]\big|^{1/2} < \infty. \tag{4.120}$$

Next observe that (4.113) demonstrates that for all $T \in [0, \infty)$ we have that

$$\begin{aligned}
&\sup_{s\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d} \int_0^1 e^{-L(T-s)}\|\psi''(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\,d\lambda\bigg]\\
&\leq \sup_{s\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d} \int_0^1 ce^{-L(T-s)}\,d\lambda\bigg]\\
&\leq c \sup_{s\in[0,T]} \mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d}\,\big]\\
&\leq c \sup_{s\in[0,\infty)} \mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d}\,\big].
\end{aligned} \tag{4.121}$$

Hölder's inequality therefore assures that for all $T \in [0, \infty)$ we have that

$$\begin{aligned}
&\sup_{s\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}\,\|Q_s\|_{\mathbb{R}^d} \int_0^1 e^{-L(T-s)}\|\psi''(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\,d\lambda\bigg]\\
&\leq c \sup_{s\in[0,\infty)} \Big(\big|\mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}^2\big]\big|^{1/2}\big|\mathbb{E}\big[\|Q_s\|_{\mathbb{R}^d}^2\big]\big|^{1/2}\Big)\\
&\leq c \sup_{s\in[0,\infty)} \big|\mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}^2\big]\big|^{1/2} \sup_{s\in[0,\infty)} \big|\mathbb{E}\big[\|Q_s\|_{\mathbb{R}^d}^2\big]\big|^{1/2}.
\end{aligned} \tag{4.122}$$

Moreover, note that the Minkowski inequality implies that

$$\begin{aligned}
&\sup_{s\in[0,\infty)} \big|\mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}^2\big]\big|^{1/2}\\
&\leq \sup_{s\in[0,\infty)} \big|\mathbb{E}\big[\|Q_s\|_{\mathbb{R}^d}^2\big]\big|^{1/2} + \sup_{s\in[0,\infty)} \big|\mathbb{E}\big[\|g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}^2\big]\big|^{1/2}.
\end{aligned} \tag{4.123}$$

In the next step observe that (4.112), (4.120), and Lemma 4.5.1 assure that

$$\sup_{s\in[0,\infty)} \mathbb{E}\big[\|Q_s\|_{\mathbb{R}^d}^2\big] = \sup_{n\in\mathbb{N}_0} \mathbb{E}\big[\|\tfrac{1}{\mathfrak{M}_n}\textstyle\sum_{j=1}^{\mathfrak{M}_n} G(\Theta_{\mathfrak{t}_n}, Z_{n+1,j})\|_{\mathbb{R}^d}^2\big] < \infty. \tag{4.124}$$

Next note that (4.112), (4.120), and Lemma 4.5.2 demonstrate that

$$\sup_{s\in[0,\infty)} \mathbb{E}\big[\|g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}^2\big] \leq \sup_{t\in[0,\infty)} \mathbb{E}\big[\|g(\Theta_t)\|_{\mathbb{R}^d}^2\big] < \infty. \tag{4.125}$$

Combining this, (4.123), and (4.124) ensures that

$$\sup_{s\in[0,\infty)} \big|\mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}^2\big]\big|^{1/2} < \infty. \tag{4.126}$$

This, (4.122), and (4.124) prove that

$$\begin{aligned} &\sup_{T\in[0,\infty)} \sup_{s\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} \\ &\quad\cdot \int_0^1 e^{-L(T-s)} \|\psi''(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})}\, d\lambda\bigg] < \infty. \end{aligned} \tag{4.127}$$

Next observe that (4.113) implies that for all $T\in(0,\infty)$ we have that

$$\begin{aligned} &\sup_{s\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} \int_0^1 \|\psi'(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\|_{L(\mathbb{R}^d,\mathbb{R})} \\ &\quad\cdot \int_0^{T-s} e^{-Lu} \|g''(\theta_u^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\, du\, d\lambda\bigg] \\ &\leq c^2 \sup_{s\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} \int_0^1 \int_0^{T-s} e^{-Lu}\, du\, d\lambda\bigg] \\ &= c^2 \sup_{s\in[0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} \int_0^{T-s} e^{-Lu}\, du\bigg] \\ &= \tfrac{c^2}{L} \sup_{s\in[0,T]} \mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} (1 - e^{-L(T-s)})\big] \\ &\leq \tfrac{c^2}{L} \sup_{s\in[0,\infty)} \mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d}\big]. \end{aligned} \tag{4.128}$$

This and Hölder's inequality ensure that for all $T \in (0, \infty)$ we have that

$$\begin{aligned}
&\sup_{s \in [0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \, \|Q_s\|_{\mathbb{R}^d} \int_0^1 \|\psi'(\theta_{T-s}^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\|_{L(\mathbb{R}^d,\mathbb{R})} \\
&\quad \cdot \int_0^{T-s} e^{-Lu} \|g''(\theta_u^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} \, du \, d\lambda \bigg] \\
&\leq \tfrac{c^2}{L} \sup_{s \in [0,\infty)} \Big(\big| \mathbb{E}\big[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d}^2 \big] \big|^{1/2} \big| \mathbb{E}\big[\, \|Q_s\|_{\mathbb{R}^d}^2 \, \big] \big|^{1/2} \Big).
\end{aligned} \tag{4.129}$$

Combining this, (4.124), and (4.126) demonstrates that

$$\begin{aligned}
&\sup_{T \in [0,\infty)} \sup_{s \in [0,T]} \mathbb{E}\bigg[\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \, \|Q_s\|_{\mathbb{R}^d} \int_0^1 \|\psi'(\theta_{T-s}^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\|_{L(\mathbb{R}^d,\mathbb{R})} \\
&\cdot \int_0^{T-s} e^{-Lu} \|g''(\theta_u^{\lambda\Theta_s + (1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} \, du \, d\lambda \bigg] < \infty.
\end{aligned} \tag{4.130}$$

Next observe that (4.113) and Hölder's inequality imply that for all $T \in [0, \infty)$ we have that

$$\begin{aligned}
&\sup_{s,v \in [0,T]} \mathbb{E}\big[\|\psi'(\theta_{T-s}^{\Theta_s})\|_{L(\mathbb{R}^d,\mathbb{R})} \|g'(\Theta_v) Q_v\|_{\mathbb{R}^d} \big] \\
&\leq \sup_{v \in [0,\infty)} \mathbb{E}\big[c \|g'(\Theta_v) Q_v\|_{\mathbb{R}^d} \big] \\
&\leq c \sup_{v \in [0,\infty)} \mathbb{E}\big[\|g'(\Theta_v)\|_{L(\mathbb{R}^d,\mathbb{R}^d)} \|Q_v\|_{\mathbb{R}^d} \big] \\
&\leq c \sup_{v \in [0,\infty)} \Big(\big| \mathbb{E}\big[\|g'(\Theta_v)\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^2 \big] \big|^{1/2} \big| \mathbb{E}\big[\|Q_v\|_{\mathbb{R}^d}^2 \big] \big|^{1/2} \Big) \\
&\leq c \sup_{v \in [0,\infty)} \big| \mathbb{E}\big[\|g'(\Theta_v)\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^2 \big] \big|^{1/2} \sup_{v \in [0,\infty)} \big| \mathbb{E}\big[\|Q_v\|_{\mathbb{R}^d}^2 \big] \big|^{1/2}.
\end{aligned} \tag{4.131}$$

Furthermore, note that Lemma 4.5.3, (4.114), and (4.120) ensure that

$$\sup_{v \in [0,\infty)} \mathbb{E}\big[\|g'(\Theta_v)\|_{L(\mathbb{R}^d,\mathbb{R}^d)}^2 \big] < \infty. \tag{4.132}$$

Combining this, (4.124), and (4.131) proves that

$$\sup_{T \in [0,\infty)} \sup_{s,v \in [0,T]} \mathbb{E}\big[\|\psi'(\theta_{T-s}^{\Theta_s})\|_{L(\mathbb{R}^d,\mathbb{R})} \|g'(\Theta_v) Q_v\|_{\mathbb{R}^d} \big] < \infty. \tag{4.133}$$

This, (4.127), and (4.130) establish (4.109). The proof of Lemma 4.5.5 is thus completed. □

4.6 Weak convergence rates for SAAs in the case of polynomially decaying learning rates with mini-batches

Proposition 4.6.1. *Assume Setting 4.1.1 and let $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$ satisfy that*

$$\sup_{x\in\mathbb{R}^d} \left(\frac{\mathbb{E}\big[\|G(x, Z_{1,1})\|_{\mathbb{R}^d}^2\big]}{\big[1 + \|x\|_{\mathbb{R}^d}\big]^2} + \big\|\mathbb{E}\big[(\tfrac{\partial^2}{\partial x^2} G)(x, Z_{1,1})\big]\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} \right) < \infty \tag{4.134}$$

and $\sup_{x\in\mathbb{R}^d} \max_{i\in\{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})} < \infty$*. Then*

(i) we have that $\{x \in \mathbb{R}^d \colon g(x) = 0\} = \{\Xi\}$ and

(ii) there exists $C \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have that

$$|\mathbb{E}[\psi(\Theta_{\mathfrak{t}_n})] - \psi(\Xi)| \leq C n^{2\varepsilon-1}. \tag{4.135}$$

Proof of Proposition 4.6.1. Throughout this proof let $\lambda \in (0, 1)$, let $K(\lambda) \in (0, \infty)$ be the real number given by

$$K(\lambda) = \sup_{n\in\mathbb{N}\cap[2,\infty)} \left[\frac{\eta^2 e^{L\eta + \frac{L\eta}{\varepsilon}}}{2(1 - 2\varepsilon)} \Big(n^{1-2\varepsilon} \Big[2e^{-\frac{L\eta}{\varepsilon}(1-\lambda^\varepsilon)n^\varepsilon} + (n - 1)^{2\varepsilon-2} \Big] + \lambda^{2\varepsilon-1} \Big) \right] \tag{4.136}$$

(cf. Lemma 3.8.7), let $Q \colon [0, \infty) \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $m \in \mathbb{N}_0$, $t \in [\mathfrak{t}_m, \mathfrak{t}_{m+1})$ that

$$Q_t = \frac{1}{\mathfrak{M}_m} \sum_{n=1}^{\mathfrak{M}_m} G(\Theta_{\mathfrak{t}_m}, Z_{m+1,n}), \tag{4.137}$$

and let $R \colon [0, \infty) \to [0, \infty]$ satisfy for all $T \in [0, \infty)$ that

$$\begin{aligned} R(T) = \sup_{s,v\in[0,T]} \mathbb{E}\Bigg[&\|Q_s - g(\Theta_{[\![s]\!]})\|_{\mathbb{R}^d} \|Q_s\|_{\mathbb{R}^d} \\ &\cdot \bigg(\int_0^1 e^{-L(T-s)} \|\psi''(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R})} + \|\psi'(\theta_{T-s}^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\|_{L(\mathbb{R}^d,\mathbb{R})} \\ &\cdot \int_0^{T-s} e^{-Lu} \big\|g''(\theta_u^{\lambda\Theta_s+(1-\lambda)\Theta_{[\![s]\!]}})\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} \, du \, d\lambda \bigg) + \|\psi'(\theta_{T-s}^{\Theta_s})\|_{L(\mathbb{R}^d,\mathbb{R})} \|g'(\Theta_v) Q_v\|_{\mathbb{R}^d} \Bigg]. \end{aligned} \tag{4.138}$$

Note that (4.134) and Lemma 3.2.2 prove that

$$\sup_{x\in\mathbb{R}^d} \|g''(x)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} < \infty. \tag{4.139}$$

This, Lemma 3.2.1, and Lemma 4.5.4 demonstrate that

$$\sup_{x\in\mathbb{R}^d} \left(\frac{\big\|\mathbb{E}\big[(\tfrac{\partial}{\partial x}G)(x, Z_{1,1})\big]\big\|_{L(\mathbb{R}^d,\mathbb{R}^d)}}{\big[1+\|x\|_{\mathbb{R}^d}\big]} \right) = \sup_{x\in\mathbb{R}^d} \left(\frac{\|g'(x)\|_{L(\mathbb{R}^d,\mathbb{R}^d)}}{\big[1+\|x\|_{\mathbb{R}^d}\big]} \right) < \infty. \tag{4.140}$$

Next observe that (4.2) assures that Ξ is the unique zero of g. This proves item (i). Item (iv) in Lemma 3.3.6 therefore ensures that

$$\limsup_{s\to\infty} \|\theta_s^\xi - \Xi\|_{\mathbb{R}^d} = 0. \tag{4.141}$$

Corollary 3.9.1, (4.134), (4.140), and (4.2) hence assure that for all $k \in \mathbb{N}$ we have that

$$\begin{aligned}
&|\mathbb{E}[\psi(\Theta_{\mathfrak{t}_k})] - \psi(\Xi)| \\
&\leq (k+1)^{2\varepsilon-1}\Bigg[K(\lambda)R(\mathfrak{t}_k) \\
&+ (k+1)^{1-2\varepsilon}e^{-L\mathfrak{t}_k} \sup_{\alpha\in[0,1]} \Big(\big\|\psi'(\alpha\theta^\xi_{\mathfrak{t}_k} + (1-\alpha)\Xi)\big\|_{L(\mathbb{R}^d,\mathbb{R})}\Big)\|\xi-\Xi\|_{\mathbb{R}^d}\Bigg] \\
&\leq (k+1)^{2\varepsilon-1}\Bigg[K(\lambda) \sup_{T\in[0,\infty)} R(T) \\
&+ (k+1)^{1-2\varepsilon}e^{-L\mathfrak{t}_k} \sup_{\alpha\in[0,1]} \Big(\big\|\psi'(\alpha\theta^\xi_{\mathfrak{t}_k} + (1-\alpha)\Xi)\big\|_{L(\mathbb{R}^d,\mathbb{R})}\Big)\|\xi-\Xi\|_{\mathbb{R}^d}\Bigg] \\
&\leq (k+1)^{2\varepsilon-1}\Bigg[K(\lambda) \sup_{T\in[0,\infty)} R(T) \\
&+ \sup_{l\in\mathbb{N}_0}\bigg((l+1)^{1-2\varepsilon}e^{-L\mathfrak{t}_l} \sup_{\alpha\in[0,1]} \Big(\big\|\psi'(\alpha\theta^\xi_{\mathfrak{t}_l} + (1-\alpha)\Xi)\big\|_{L(\mathbb{R}^d,\mathbb{R})}\Big)\bigg)\|\xi-\Xi\|_{\mathbb{R}^d}\Bigg].
\end{aligned} \tag{4.142}$$

Next note that Lemma 4.2.2 and Lemma 4.5.5 imply that

$$\begin{aligned}
&\sup_{l\in\mathbb{N}_0}\bigg((l+1)^{1-2\varepsilon}e^{-L\mathfrak{t}_l} \sup_{\alpha\in[0,1]} \Big(\big\|\psi'(\alpha\theta^\xi_{\mathfrak{t}_l} + (1-\alpha)\Xi)\big\|_{L(\mathbb{R}^d,\mathbb{R})}\Big)\bigg)\|\xi-\Xi\|_{\mathbb{R}^d} \\
&\qquad + K(\lambda) \sup_{T\in[0,\infty)} R(T) < \infty.
\end{aligned} \tag{4.143}$$

Furthermore, observe that for all $k \in \mathbb{N}$ we have that

$$(k+1)^{2\varepsilon-1} \leq k^{2\varepsilon-1}. \tag{4.144}$$

This, (4.143), and (4.142) establish item (ii). The proof of Proposition 4.6.1 is thus completed. □

Theorem 4.6.2. *Let $d \in \mathbb{N}$, $\xi, \Xi \in \mathbb{R}^d$, $\varepsilon \in (0, 1/2)$, $\eta, L, c \in (0,\infty)$, $(\mathfrak{M}_n)_{n\in\mathbb{N}_0} \subseteq \mathbb{N}$, $\psi \in C^2(\mathbb{R}^d,\mathbb{R})$, let $(S,\mathcal{S})$ be a measurable space, let $(\Omega,\mathcal{F},\mathbb{P})$ be a probability space, let $Z_{m,n}\colon \Omega \to S$, $(m,n) \in \mathbb{N}^2$, be i.i.d. random variables, let $G = (G(x,s))_{(x,s)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be $(\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{S})/\mathcal{B}(\mathbb{R}^d)$-measurable, let $g\colon \mathbb{R}^d \to \mathbb{R}^d$ be a function, assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto G(x,s) \in \mathbb{R}^d) \in C^2(\mathbb{R}^d,\mathbb{R}^d)$, assume for all $x, y \in \mathbb{R}^d$ that*

$$\mathbb{E}\big[\|G(x,Z_{1,1})\|_{\mathbb{R}^d}^2\big] \leq c\big[1+\|x\|_{\mathbb{R}^d}\big]^2, \quad \langle x-\Xi, g(x)\rangle_{\mathbb{R}^d} \leq -L\|g(x)\|_{\mathbb{R}^d}^2, \tag{4.145}$$

$$g(x) = \mathbb{E}\big[G(x,Z_{1,1})\big], \quad \langle x-y, g(x)-g(y)\rangle_{\mathbb{R}^d} \leq -L\|x-y\|_{\mathbb{R}^d}^2, \tag{4.146}$$

$$\max_{i\in\{1,2\}} \inf_{\delta\in(0,\infty)} \sup_{u\in[-\delta,\delta]^d} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i}G)(x+u,Z_{1,1})\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)}^{1+\delta}\Big] < \infty, \tag{4.147}$$

and

$$\big\|\mathbb{E}\big[(\tfrac{\partial^2}{\partial x^2}G)(x,Z_{1,1})\big]\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} + \max_{i\in\{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})} < c \tag{4.148}$$

(cf. Corollary 2.2.5), and let $\Theta\colon \mathbb{N}_0\times\Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $n \in \mathbb{N}$ that $\Theta_0 = \xi$ and

$$\Theta_n = \Theta_{n-1} + \frac{\eta}{n^{1-\varepsilon}\mathfrak{M}_{n-1}} \sum_{j=1}^{\mathfrak{M}_{n-1}} G(\Theta_{n-1}, Z_{n,j}). \tag{4.149}$$

Then

(i) we have that $\{x \in \mathbb{R}^d\colon g(x) = 0\} = \{\Xi\}$ and

(ii) there exists $C \in [0,\infty)$ such that for all $n \in \mathbb{N}$ we have that

$$|\mathbb{E}[\psi(\Theta_n)] - \psi(\Xi)| \leq Cn^{2\varepsilon-1}. \tag{4.150}$$

Proof of Theorem 4.6.2. This is a direct consequence of Proposition 4.6.1. The proof of Theorem 4.6.2 is thus completed. □

4.7 Weak convergence rates for SAAs in the case of polynomially decaying learning rates without mini-batches

Corollary 4.7.1. *Let $d \in \mathbb{N}$, $\xi, \Xi \in \mathbb{R}^d$, $\varepsilon \in (0, 1/2)$, $\eta, L, c \in (0, \infty)$, $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$, let $(S, \mathcal{S})$ be a measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $G = (G(x,s))_{(x,s)\in\mathbb{R}^d\times S} \colon \mathbb{R}^d \times S \to \mathbb{R}^d$ be $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{S})/\mathcal{B}(\mathbb{R}^d)$-measurable, let $Z_n \colon \Omega \to S$, $n \in \mathbb{N}$, be i.i.d. random variables, let $g \colon \mathbb{R}^d \to \mathbb{R}^d$ be a function, assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto G(x,s) \in \mathbb{R}^d) \in C^2(\mathbb{R}^d, \mathbb{R}^d)$, assume for all $x, y \in \mathbb{R}^d$ that*

$$\mathbb{E}\big[\|G(x, Z_1)\|_{\mathbb{R}^d}^2\big] \leq c\big[1 + \|x\|_{\mathbb{R}^d}\big]^2, \quad \langle x - \Xi, g(x)\rangle_{\mathbb{R}^d} \leq -L\|g(x)\|_{\mathbb{R}^d}^2, \tag{4.151}$$

$$g(x) = \mathbb{E}\big[G(x, Z_1)\big], \quad \langle x - y, g(x) - g(y)\rangle_{\mathbb{R}^d} \leq -L\|x - y\|_{\mathbb{R}^d}^2, \tag{4.152}$$

$$\max_{i\in\{1,2\}} \inf_{\delta\in(0,\infty)} \sup_{u\in[-\delta,\delta]^d} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i}G)(x+u, Z_1)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)}^{1+\delta}\Big] < \infty, \tag{4.153}$$

and

$$\big\|\mathbb{E}\big[(\tfrac{\partial^2}{\partial x^2}G)(x, Z_1)\big]\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} + \max_{i\in\{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})} < c \tag{4.154}$$

(cf. Corollary 2.2.5), and let $\Theta \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $n \in \mathbb{N}$ that $\Theta_0 = \xi$ and

$$\Theta_n = \Theta_{n-1} + \tfrac{\eta}{n^{1-\varepsilon}} G(\Theta_{n-1}, Z_n). \tag{4.155}$$

Then

(i) we have that $\{x \in \mathbb{R}^d \colon g(x) = 0\} = \{\Xi\}$ and

(ii) there exists $C \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have that

$$|\mathbb{E}[\psi(\Theta_n)] - \psi(\Xi)| \leq C n^{2\varepsilon - 1}. \tag{4.156}$$

Proof of Corollary 4.7.1. This is a direct consequence of Theorem 4.6.2. The proof of Corollary 4.7.1 is thus completed. □

4.8 SAAs for random rotation problems

Lemma 4.8.1. *Let* $(\Omega, \mathcal{F}, \mathbb{P})$ *be a probability space, let* $Z_n \colon \Omega \to [\pi/4, 5\pi/4]$*,* $n \in \mathbb{N}$*, be i.i.d. random variables, assume that* Z_1 *is continuous uniformly distributed on* $(\pi/4, 5\pi/4)$*, let* $A \colon [\pi/4, 5\pi/4] \to \mathbb{R}^{2\times 2}$ *satisfy for all* $s \in [\pi/4, 5\pi/4]$ *that*

$$A(s) = \begin{pmatrix} \cos(s) & -\sin(s) \\ \sin(s) & \cos(s) \end{pmatrix}, \tag{4.157}$$

let $G = (G(x,s))_{(x,s)\in\mathbb{R}^2\times[\pi/4,5\pi/4]} \colon \mathbb{R}^2 \times [\pi/4, 5\pi/4] \to \mathbb{R}^2$ *satisfy for all* $x \in \mathbb{R}^2$*,* $s \in [\pi/4, 5\pi/4]$ *that*

$$G(x,s) = A(s)x, \tag{4.158}$$

and let $g \colon \mathbb{R}^2 \to \mathbb{R}^2$ *satisfy for all* $x \in \mathbb{R}^2$ *that*

$$g(x) = \mathbb{E}\big[G(x, Z_1)\big]. \tag{4.159}$$

Then

(i) we have for all $x \in \mathbb{R}^2$ *that*

$$\mathbb{E}\big[A(Z_1)\big] = \tfrac{\sqrt{2}}{\pi} \begin{pmatrix} -1 & -1 \\ 1 & -1 \end{pmatrix} = \tfrac{2}{\pi} A\big(\tfrac{3\pi}{4}\big) \tag{4.160}$$

and

$$g(x) = \tfrac{2}{\pi} A\big(\tfrac{3\pi}{4}\big)x, \tag{4.161}$$

(ii) we have for all $s \in [\pi/4, 5\pi/4]$ *that*

$$\big(\mathbb{R}^2 \ni x \mapsto A(s)x \in \mathbb{R}^2\big) \in C^2(\mathbb{R}^2, \mathbb{R}^2), \tag{4.162}$$

(iii) we have for all $x \in \mathbb{R}^2$ *that*

$$\max_{i\in\{1,2\}} \sup_{u\in[-1,1]^2} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i} G)(x+u, Z_1)\|^2_{L^{(i)}(\mathbb{R}^2,\mathbb{R}^2)}\Big] < \infty, \tag{4.163}$$

(iv) we have for all $s \in [\pi/4, 5\pi/4]$*,* $x \in \mathbb{R}^2$ *that*

$$\|A(s)x\|_{\mathbb{R}^2} = \|x\|_{\mathbb{R}^2}, \qquad \|g(x)\|_{\mathbb{R}^2} = \tfrac{2}{\pi}\|x\|_{\mathbb{R}^2}, \tag{4.164}$$

and

$$\mathbb{E}\big[\|G(x, Z_1)\|^2_{\mathbb{R}^2}\big] = \|x\|^2_{\mathbb{R}^2}, \tag{4.165}$$

(v) we have for all $s \in [\pi/4, 5\pi/4]$, $x \in \mathbb{R}^2$ that

$$\langle A(s)x, x\rangle_{\mathbb{R}^2} = \cos(s)\|x\|^2_{\mathbb{R}^2}, \tag{4.166}$$

and

(vi) we have for all $x, y \in \mathbb{R}^2$ that

$$\langle x - y, g(x) - g(y)\rangle_{\mathbb{R}^2} = -\tfrac{\sqrt{2}}{\pi}\|x - y\|^2_{\mathbb{R}^2}. \tag{4.167}$$

Proof of Lemma 4.8.1. First, observe that

$$\mathbb{E}\big[\cos(Z_1)\big] = \int_{\pi/4}^{5\pi/4} \cos(s)\tfrac{1}{\pi}\, ds = \tfrac{1}{\pi}\big[\sin(s)\big]_{s=\pi/4}^{s=5\pi/4} = \tfrac{1}{\pi}(-\tfrac{\sqrt{2}}{2} - \tfrac{\sqrt{2}}{2}) = -\tfrac{\sqrt{2}}{\pi} \tag{4.168}$$

and

$$\mathbb{E}\big[\sin(Z_1)\big] = \int_{\pi/4}^{5\pi/4} \sin(s)\tfrac{1}{\pi}\, ds = -\tfrac{1}{\pi}\big[\cos(s)\big]_{s=\pi/4}^{s=5\pi/4} = -\tfrac{1}{\pi}(-\tfrac{\sqrt{2}}{2} - \tfrac{\sqrt{2}}{2}) = \tfrac{\sqrt{2}}{\pi}. \tag{4.169}$$

This and (4.157) prove (4.160). Combining this, (4.158), and (4.159) demonstrates (4.161). This establishes item (i). Moreover, note that item (ii) is obvious. Next observe that for all $x \in \mathbb{R}^2$, $s \in [\pi/4, 5\pi/4]$ we have that

$$(\tfrac{\partial}{\partial x}G)(x, s) = A(s) \qquad \text{and} \qquad (\tfrac{\partial^2}{\partial x^2}G)(x, s) = 0. \tag{4.170}$$

Furthermore, note that for all $s \in [\pi/4, 5\pi/4]$, $x = (x_1, x_2) \in \mathbb{R}^2$ we have that

$$\begin{aligned}
\|A(s)x\|^2_{\mathbb{R}^2} &= (\cos(s)x_1 - \sin(s)x_2)^2 + (\sin(s)x_1 + \cos(s)x_2)^2 \\
&= \cos(s)^2x_1^2 - 2\cos(s)\sin(s)x_1x_2 + \sin(s)^2x_2^2 \\
&\quad + \sin(s)^2x_1^2 + 2\cos(s)\sin(s)x_1x_2 + \cos(s)^2x_2^2 \\
&= x_1^2 + x_2^2 = \|x\|^2_{\mathbb{R}^2}.
\end{aligned} \tag{4.171}$$

This and (4.170) establish item (iii). Next observe that (4.171) and (4.161) prove that for all $x \in \mathbb{R}^2$ we have that

$$\|g(x)\|_{\mathbb{R}^2} = \tfrac{2}{\pi}\|x\|_{\mathbb{R}^2}. \tag{4.172}$$

In addition, observe that (4.171) and (4.158) ensure that for all $x \in \mathbb{R}^2$ we have that

$$\mathbb{E}\big[\|G(x, Z_1)\|^2_{\mathbb{R}^2}\big] = \mathbb{E}\big[\|A(Z_1)x\|^2_{\mathbb{R}^2}\big] = \mathbb{E}\big[\|x\|^2_{\mathbb{R}^2}\big] = \|x\|^2_{\mathbb{R}^2}. \tag{4.173}$$

Combining this, (4.171), and (4.172) establishes item (iv). Next note that for all $x = (x_1, x_2) \in \mathbb{R}^2$, $s \in [\pi/4, 5\pi/4]$ we have that

$$\begin{aligned} \langle A(s)x, x\rangle_{\mathbb{R}^2} &= \langle (x_1\cos(s) - x_2\sin(s), x_1\sin(s) + x_2\cos(s)), (x_1, x_2)\rangle_{\mathbb{R}^2} \\ &= x_1^2\cos(s) + x_2^2\cos(s) \\ &= \cos(s)\|x\|_{\mathbb{R}^2}^2. \end{aligned} \tag{4.174}$$

This proves item (v). Combining this with (4.161) assures that for all $x, y \in \mathbb{R}^2$ we have that

$$\begin{aligned} \langle x - y, g(x) - g(y)\rangle_{\mathbb{R}^2} &= \tfrac{2}{\pi}\langle x - y, A\big(\tfrac{3\pi}{4}\big)(x - y)\rangle_{\mathbb{R}^2} \\ &= \tfrac{2}{\pi}\langle A\big(\tfrac{3\pi}{4}\big)(x - y), (x - y)\rangle_{\mathbb{R}^2} \\ &= \tfrac{2}{\pi}\cos\big(\tfrac{3\pi}{4}\big)\|x - y\|_{\mathbb{R}^2}^2 = -\tfrac{\sqrt{2}}{\pi}\|x - y\|_{\mathbb{R}^2}^2. \end{aligned} \tag{4.175}$$

This establishes item (vi). The proof of Lemma 4.8.1 is thus completed. □

Corollary 4.8.2. *Let $\xi \in \mathbb{R}^2$, $\varepsilon \in (0, 1/2)$, $\eta \in (0, \infty)$, $\psi \in C^2(\mathbb{R}^2, \mathbb{R})$ satisfy that $\sup_{x\in\mathbb{R}^2} \max_{i\in\{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^2,\mathbb{R})} < \infty$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $Z_n \colon \Omega \to [\pi/4, 5\pi/4]$, $n \in \mathbb{N}$, be i.i.d. random variables, assume that Z_1 is continuous uniformly distributed on $(\pi/4, 5\pi/4)$, let $A \colon [\pi/4, 5\pi/4] \to \mathbb{R}^{2\times 2}$ satisfy for all $s \in [\pi/4, 5\pi/4]$ that*

$$A(s) = \begin{pmatrix} \cos(s) & -\sin(s) \\ \sin(s) & \cos(s) \end{pmatrix}, \tag{4.176}$$

let $G = (G(x, s))_{(x,s)\in\mathbb{R}^2\times[\pi/4,5\pi/4]} \colon \mathbb{R}^2 \times [\pi/4, 5\pi/4] \to \mathbb{R}^2$ satisfy for all $x \in \mathbb{R}^2$, $s \in [\pi/4, 5\pi/4]$ that

$$G(x, s) = A(s)x, \tag{4.177}$$

let $g \colon \mathbb{R}^2 \to \mathbb{R}^2$ satisfy for all $x \in \mathbb{R}^2$ that

$$g(x) = \mathbb{E}\big[G(x, Z_1)\big], \tag{4.178}$$

and let $\Theta \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^2$ be the stochastic process which satisfies for all $n \in \mathbb{N}$ that $\Theta_0 = \xi$ and

$$\Theta_n = \Theta_{n-1} + \tfrac{\eta}{n^{1-\varepsilon}} G(\Theta_{n-1}, Z_n). \tag{4.179}$$

Then

(i) we have that $\{x \in \mathbb{R}^2 \colon g(x) = 0\} = \{0\}$ and

(ii) there exists $C \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have that

$$|\mathbb{E}[\psi(\Theta_n)] - \psi(0)| \le C n^{2\varepsilon - 1}. \tag{4.180}$$

Proof of Corollary 4.8.2. First, note that item (iv) in Lemma 4.8.1 proves that for all $x \in \mathbb{R}^2$ we have that

$$\mathbb{E}\big[\|G(x, Z_1)\|_{\mathbb{R}^2}^2\big] = \|x\|_{\mathbb{R}^2}^2 \leq \big[1 + \|x\|_{\mathbb{R}^2}\big]^2 \tag{4.181}$$

and

$$\|g(x)\|_{\mathbb{R}^2} = \tfrac{2}{\pi}\|x\|_{\mathbb{R}^2}. \tag{4.182}$$

Next observe that item (iii) in Lemma 4.8.1 establishes for all $x \in \mathbb{R}^2$ that

$$\max_{i \in \{1,2\}} \sup_{u \in [-1,1]^2} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i} G)(x + u, Z_1)\|_{L^{(i)}(\mathbb{R}^2, \mathbb{R}^2)}^2\Big] < \infty. \tag{4.183}$$

Moreover, note that item (vi) in Lemma 4.8.1 ensures that for all $x, y \in \mathbb{R}^2$ we have that

$$\langle x - y, g(x) - g(y)\rangle_{\mathbb{R}^2} = -\tfrac{\sqrt{2}}{\pi}\|x - y\|_{\mathbb{R}^2}^2. \tag{4.184}$$

This and (4.182) ensure that for all $x \in \mathbb{R}^2$ we have that

$$\langle x, g(x)\rangle_{\mathbb{R}^2} = \langle x - 0, g(x) - g(0)\rangle_{\mathbb{R}^2} = -\tfrac{\sqrt{2}}{\pi}\|x\|_{\mathbb{R}^2}^2 = -\tfrac{\pi\sqrt{2}}{4}\|g(x)\|_{\mathbb{R}^2}^2. \tag{4.185}$$

In addition, observe that for all $x \in \mathbb{R}^2$, $s \in [\pi/4, 5\pi/4]$ we have that

$$(\tfrac{\partial^2}{\partial x^2} G)(x, s) = 0. \tag{4.186}$$

This reveals that for all $s \in [\pi/4, 5\pi/4]$ it holds that

$$\sup_{x \in \mathbb{R}^2} \|(\tfrac{\partial^2}{\partial x^2} G)(x, s)\|_{L^{(2)}(\mathbb{R}^2, \mathbb{R}^2)} < \infty. \tag{4.187}$$

Combining this with Corollary 4.7.1, (4.181), (4.183), (4.184), (4.185), and the assumption that $\sup_{x \in \mathbb{R}^2} \max_{i \in \{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^2, \mathbb{R})} < \infty$ establishes item (i) and item (ii). The proof of Corollary 4.8.2 is thus completed. □

Chapter 5

Weak error estimates for stochastic gradient descent (SGD) optimization methods

In this chapter we apply the weak error analysis results for SAAs from Chapter 4 above to establish weak error estimates for SGD optimization methods. In particular, in Corollary 5.2.1 in Section 5.2 below we establish weak error estimates for SGD optimization methods in the case of objective functions with linearly growing derivatives. In our proof of Corollary 5.2.1 we employ the weak error estimates for SGD optimization methods in the case of coercive objective functions in Corollary 5.1.2 in Section 5.1 below. Our proof of Corollary 5.1.2, in turn, uses the elementary result on derivatives of gradients of smooth functions in Lemma 5.1.1 in Section 5.1 below and the weak convergence result for SAAs in Corollary 4.7.1 in Section 4.7 above.

5.1 Weak error estimates for SGD optimization methods in the case of coercive objective functions

Lemma 5.1.1. *Let $d, n \in \mathbb{N}$, $f \in C^n(\mathbb{R}^d, \mathbb{R})$ and let $g\colon \mathbb{R}^d \to \mathbb{R}^d$ satisfy for all $x \in \mathbb{R}^d$ that $g(x) = (\nabla f)(x)$. Then*

(i) we have that $g \in C^{(n-1)}(\mathbb{R}^d, \mathbb{R}^d)$,

(ii) we have for all $k \in \{1, 2, \dots, n\}$, $x, y_1, y_2, \dots, y_k \in \mathbb{R}^d$ that

$$f^{(k)}(x)(y_1, y_2, \dots, y_k) = \langle g^{(k-1)}(x)(y_2, y_3, \dots, y_k), y_1 \rangle_{\mathbb{R}^d}, \tag{5.1}$$

and

(iii) we have for all $k \in \{1, 2, \dots, n\}$, $x \in \mathbb{R}^d$ that

$$\|f^{(k)}(x)\|_{L^{(k)}(\mathbb{R}^d, \mathbb{R})} = \|g^{(k-1)}(x)\|_{L^{(k-1)}(\mathbb{R}^d, \mathbb{R}^d)}. \tag{5.2}$$

Proof of Lemma 5.1.1. First, note that the hypothesis that $f \in C^n(\mathbb{R}^d, \mathbb{R})$ establishes item (i). Next we prove item (ii) by induction on $k \in \{1, 2, \dots, n\}$. For the base case $k = 1$ note that for all $x, y \in \mathbb{R}^d$ we have that

$$f'(x)(y) = \langle g(x), y \rangle_{\mathbb{R}^d}. \tag{5.3}$$

This proves (5.1) in the base case $k = 1$. For the induction step $\{1, 2, \dots, n-1\} \ni k \to k+1 \in \{2, 3, \dots, n\}$ let $k \in \{1, 2, \dots, n-1\}$ satisfy for all $x, y_1, y_2, \dots, y_k \in \mathbb{R}^d$ that

$$f^{(k)}(x)(y_1, y_2, \dots, y_k) = \langle g^{(k-1)}(x)(y_2, y_3, \dots, y_k), y_1 \rangle_{\mathbb{R}^d}. \tag{5.4}$$

Next observe that item (i) ensures that for all $x, y_2, y_3, \dots, y_{k+1} \in \mathbb{R}^d$ we have that

$$\begin{aligned} \limsup_{\substack{h \to 0 \\ h \in \mathbb{R} \setminus \{0\}}} \bigg\| & \frac{g^{(k-1)}(x + h y_{k+1})(y_2, y_3, \dots, y_k) - g^{(k-1)}(x)(y_2, y_3, \dots, y_k)}{h} \\ & - g^{(k)}(x)(y_2, y_3, \dots, y_k, y_{k+1}) \bigg\|_{\mathbb{R}^d} = 0. \end{aligned} \tag{5.5}$$

The Cauchy-Schwartz inequality hence implies that for all $x, y_1, y_2, \dots, y_{k+1} \in \mathbb{R}^d$ we have that

$$\begin{aligned} & \limsup_{\substack{h \to 0 \\ h \in \mathbb{R} \setminus \{0\}}} \bigg| \frac{\langle g^{(k-1)}(x + h y_{k+1})(y_2, y_3, \dots, y_k) - g^{(k-1)}(x)(y_2, y_3, \dots, y_k), y_1 \rangle_{\mathbb{R}^d}}{h} \\ & \qquad - \langle g^{(k)}(x)(y_2, y_3, \dots, y_k, y_{k+1}), y_1 \rangle_{\mathbb{R}^d} \bigg| \\ & \leq \limsup_{\substack{h \to 0 \\ h \in \mathbb{R} \setminus \{0\}}} \bigg\| \frac{g^{(k-1)}(x + h y_{k+1})(y_2, y_3, \dots, y_k) - g^{(k-1)}(x)(y_2, y_3, \dots, y_k)}{h} \\ & \qquad - g^{(k)}(x)(y_2, y_3, \dots, y_k, y_{k+1}) \bigg\|_{\mathbb{R}^d} \|y_1\|_{\mathbb{R}^d} = 0. \end{aligned} \tag{5.6}$$

The induction hypothesis (see (5.4)) therefore assures that for all $x, y_1, y_2, \ldots, y_{k+1} \in \mathbb{R}^d$ we have that

$$\begin{aligned}
&\limsup_{\substack{h\to 0\\ h\in\mathbb{R}\setminus\{0\}}} \bigg| \frac{f^{(k)}(x+hy_{k+1})(y_1,y_2,\ldots,y_k) - f^{(k)}(x)(y_1,y_2,\ldots,y_k)}{h} \\
&\qquad - \langle g^{(k)}(x)(y_2,y_3,\ldots,y_{k+1}), y_1\rangle_{\mathbb{R}^d} \bigg| \\
&= \limsup_{\substack{h\to 0\\ h\in\mathbb{R}\setminus\{0\}}} \bigg| \frac{\langle g^{(k-1)}(x+hy_{k+1})(y_2,y_3,\ldots,y_k) - g^{(k-1)}(x)(y_2,y_3,\ldots,y_k), y_1\rangle_{\mathbb{R}^d}}{h} \\
&\qquad - \langle g^{(k)}(x)(y_2,y_3,\ldots,y_{k+1}), y_1\rangle_{\mathbb{R}^d} \bigg| = 0.
\end{aligned} \tag{5.7}$$

This and the assumption that $f \in C^n(\mathbb{R}^d, \mathbb{R})$ demonstrates that for all $x, y_1, y_2, \ldots, y_{k+1} \in \mathbb{R}^d$ we have that

$$f^{(k+1)}(x)(y_1,y_2,\ldots,y_{k+1}) = \langle g^{(k)}(x)(y_2,y_3,\ldots,y_{k+1}), y_1\rangle_{\mathbb{R}^d}. \tag{5.8}$$

Induction thus proves item (ii). Next observe that item (ii) implies that for all $k \in \{1,2,\ldots,n\}$, $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
\|f^{(k)}(x)\|_{L^{(k)}(\mathbb{R}^d,\mathbb{R})} &= \sup_{y_1,y_2,\ldots,y_k\in\mathbb{R}^d\setminus\{0\}} \frac{|f^{(k)}(x)(y_1,y_2,\ldots,y_k)|}{\|y_1\|_{\mathbb{R}^d}\|y_2\|_{\mathbb{R}^d}\ldots\|y_k\|_{\mathbb{R}^d}} \\
&= \sup_{y_1,y_2,\ldots,y_k\in\mathbb{R}^d\setminus\{0\}} \frac{|\langle g^{(k-1)}(x)(y_2,y_3,\ldots,y_k), y_1\rangle_{\mathbb{R}^d}|}{\|y_1\|_{\mathbb{R}^d}\|y_2\|_{\mathbb{R}^d}\ldots\|y_k\|_{\mathbb{R}^d}} \\
&= \sup_{y_2,y_3,\ldots,y_k\in\mathbb{R}^d\setminus\{0\}} \left[\sup_{y_1\in\mathbb{R}^d\setminus\{0\}} \frac{|\langle g^{(k-1)}(x)(y_2,y_3,\ldots,y_k), y_1\rangle_{\mathbb{R}^d}|}{\|y_1\|_{\mathbb{R}^d}\|y_2\|_{\mathbb{R}^d}\ldots\|y_k\|_{\mathbb{R}^d}} \right] \\
&= \sup_{y_2,y_3,\ldots,y_k\in\mathbb{R}^d\setminus\{0\}} \frac{\|g^{(k-1)}(x)(y_2,y_3,\ldots,y_k)\|_{\mathbb{R}^d}}{\|y_2\|_{\mathbb{R}^d}\|y_3\|_{\mathbb{R}^d}\ldots\|y_k\|_{\mathbb{R}^d}} \\
&= \|g^{(k-1)}(x)\|_{L^{(k-1)}(\mathbb{R}^d,\mathbb{R}^d)}.
\end{aligned} \tag{5.9}$$

This establishes item (iii). The proof of Lemma 5.1.1 is thus completed. □

Corollary 5.1.2. *Let $d \in \mathbb{N}$, $\xi, \Xi \in \mathbb{R}^d$, $\varepsilon \in (0, 1/2)$, $\eta, L, c \in (0,\infty)$, $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$, let $(S, \mathcal{S})$ be a measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $F = (F(x,s))_{(x,s)\in\mathbb{R}^d\times S} \colon \mathbb{R}^d \times S \to \mathbb{R}$ be $(\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{S})/\mathcal{B}(\mathbb{R})$-measurable, let $Z_n \colon \Omega \to S$, $n \in \mathbb{N}$, be i.i.d. random variables, let $g \colon \mathbb{R}^d \to \mathbb{R}^d$ be a function,* *assume for all $s \in S$ that $(\mathbb{R}^d \ni x \mapsto F(x,s) \in \mathbb{R}) \in C^3(\mathbb{R}^d, \mathbb{R})$, assume for all $x, y \in \mathbb{R}^d$ that*

$$\mathbb{E}\big[\|(\nabla_x F)(x, Z_1)\|_{\mathbb{R}^d}^2\big] \leq c\big[1 + \|x\|_{\mathbb{R}^d}^2\big], \qquad \langle x - \Xi, g(x) \rangle_{\mathbb{R}^d} \leq -L\|g(x)\|_{\mathbb{R}^d}^2, \tag{5.10}$$

$$g(x) = \mathbb{E}[(\nabla_x F)(x, Z_1)], \quad \langle x - y, g(x) - g(y) \rangle_{\mathbb{R}^d} \leq -L\|x - y\|_{\mathbb{R}^d}^2, \tag{5.11}$$

$$\max_{i \in \{2,3\}} \inf_{\delta \in (0,\infty)} \sup_{u \in [-\delta,\delta]^d} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i} F)(x + u, Z_1)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})}^{1+\delta}\Big] < \infty, \tag{5.12}$$

and

$$\big\|\mathbb{E}\big[(\tfrac{\partial^3}{\partial x^3} F)(x, Z_1)\big]\big\|_{L^{(3)}(\mathbb{R}^d,\mathbb{R})} + \max_{i \in \{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})} < c \tag{5.13}$$

(cf. Corollary 2.2.5), and let $\Theta \colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ be the stochastic process which satisfies for all $n \in \mathbb{N}$ that $\Theta_0 = \xi$ and

$$\Theta_n = \Theta_{n-1} + \tfrac{\eta}{n^{1-\varepsilon}} (\nabla_x F)(\Theta_{n-1}, Z_n). \tag{5.14}$$

Then

(i) we have that $\{x \in \mathbb{R}^d \colon g(x) = 0\} = \{\Xi\}$ and

(ii) there exists $C \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have that

$$|\mathbb{E}[\psi(\Theta_n)] - \psi(\Xi)| \leq C n^{2\varepsilon - 1}. \tag{5.15}$$

Proof of Corollary 5.1.2. Throughout this proof let $G = (G(x,s))_{(x,s) \in \mathbb{R}^d \times S} \colon \mathbb{R}^d \times S \to \mathbb{R}^d$ satisfy for all $x \in \mathbb{R}^d$, $s \in S$ that

$$G(x,s) = (\nabla_x F)(x,s). \tag{5.16}$$

Observe that the hypothesis that $\forall\, s \in S \colon (\mathbb{R}^d \ni x \mapsto F(x,s) \in \mathbb{R}) \in C^3(\mathbb{R}^d, \mathbb{R})$ ensures that for all $s \in S$ we have that

$$\big(\mathbb{R}^d \ni x \mapsto G(x,s) \in \mathbb{R}^d\big) \in C^2(\mathbb{R}^d, \mathbb{R}^d). \tag{5.17}$$

In addition, note that (5.16) and (5.10) imply that

$$\sup_{x \in \mathbb{R}^d} \left(\frac{\mathbb{E}\big[\|G(x, Z_1)\|_{\mathbb{R}^d}^2\big]}{\big[1 + \|x\|_{\mathbb{R}^d}\big]^2} \right) \leq c. \tag{5.18}$$

Next observe that item (iii) in Lemma 5.1.1 (with $d = d$, $n = 3$, $f = (\mathbb{R}^d \ni x \mapsto F(x,s) \in \mathbb{R}) \in C^3(\mathbb{R}^d,\mathbb{R})$, $g = (\mathbb{R}^d \ni x \mapsto G(x,s) \in \mathbb{R}^d)$ for $s \in S$ in the notation of Lemma 5.1.1) assures that for all $i \in \{1,2\}$, $x \in \mathbb{R}^d$, $s \in S$ we have that

$$\|(\tfrac{\partial^i}{\partial x^i}G)(x,s)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)} = \|(\tfrac{\partial^{i+1}}{\partial x^{i+1}}F)(x,s)\|_{L^{(i+1)}(\mathbb{R}^d,\mathbb{R})}. \tag{5.19}$$

This and (5.12) demonstrate that for all $x \in \mathbb{R}^d$ we have that

$$\max_{i\in\{1,2\}} \inf_{\delta\in(0,\infty)} \sup_{u\in[-\delta,\delta]^d} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i}G)(x+u,Z_1)\|^{1+\delta}_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)}\Big] < \infty. \tag{5.20}$$

Jensen's inequality hence proves that for all $x \in \mathbb{R}^d$ we have that

$$\max_{i\in\{1,2\}} \mathbb{E}\Big[\|(\tfrac{\partial^i}{\partial x^i}G)(x,Z_1)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R}^d)}\Big] < \infty. \tag{5.21}$$

Moreover, observe that for all $y_1,y_2,y_3 \in \mathbb{R}^d$ we have that $(L^{(3)}(\mathbb{R}^d,\mathbb{R}) \ni A \mapsto A(y_1,y_2,y_3) \in \mathbb{R})$ is a continuous linear function. This ensures that for all vectors $y_1,y_2,y_3 \in \mathbb{R}^d$ and all random variables $A\colon \Omega \to L^{(3)}(\mathbb{R}^d,\mathbb{R})$ with $\mathbb{E}[\|A\|_{L^{(3)}(\mathbb{R}^d,\mathbb{R})}] < \infty$ we have that $\mathbb{E}[|A(y_1,y_2,y_3)|] < \infty$ and

$$\mathbb{E}[A](y_1,y_2,y_3) = \mathbb{E}[A(y_1,y_2,y_3)]. \tag{5.22}$$

Combining this, Corollary 2.2.5, and item (ii) in Lemma 5.1.1 (with $d = d$, $n = 3$, $f = (\mathbb{R}^d \ni x \mapsto F(x,Z_1(\omega)) \in \mathbb{R}) \in C^3(\mathbb{R}^d,\mathbb{R})$, $g = (\mathbb{R}^d \ni x \mapsto G(x,Z_1(\omega)) \in \mathbb{R}^d)$ for $\omega \in \Omega$ in the notation of Lemma 5.1.1) implies that for all $x,y_1,y_2,y_3 \in \mathbb{R}^d$ we have that

$$\begin{aligned}\mathbb{E}\big[(\tfrac{\partial^3}{\partial x^3}F)(x,Z_1)\big](y_1,y_2,y_3) &= \mathbb{E}\big[(\tfrac{\partial^3}{\partial x^3}F)(x,Z_1)(y_1,y_2,y_3)\big]\\ &= \mathbb{E}\big[\langle(\tfrac{\partial^2}{\partial x^2}G)(x,Z_1)(y_2,y_3),y_1\rangle_{\mathbb{R}^d}\big].\end{aligned} \tag{5.23}$$

Moreover, note that (5.21) assures that for all $x,y_1,y_2,y_3 \in \mathbb{R}^d$ we have that

$$\begin{aligned}&\mathbb{E}\big[|\langle(\tfrac{\partial^2}{\partial x^2}G)(x,Z_1)(y_2,y_3),y_1\rangle_{\mathbb{R}^d}|\big] \le \mathbb{E}\big[\|(\tfrac{\partial^2}{\partial x^2}G)(x,Z_1)(y_2,y_3)\|_{\mathbb{R}^d}\|y_1\|_{\mathbb{R}^d}\big]\\ &\le \mathbb{E}\big[\|(\tfrac{\partial^2}{\partial x^2}G)(x,Z_1)\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}\big]\|y_2\|_{\mathbb{R}^d}\|y_3\|_{\mathbb{R}^d}\|y_1\|_{\mathbb{R}^d} < \infty.\end{aligned} \tag{5.24}$$

This and (5.23) prove that for all $x,y_1,y_2,y_3 \in \mathbb{R}^d$ we have that

$$\mathbb{E}\big[(\tfrac{\partial^3}{\partial x^3}F)(x,Z_1)\big](y_1,y_2,y_3) = \big\langle\mathbb{E}\big[(\tfrac{\partial^2}{\partial x^2}G)(x,Z_1)(y_2,y_3)\big],y_1\big\rangle_{\mathbb{R}^d}. \tag{5.25}$$

This reveals that for all $x \in \mathbb{R}^d$ it holds that

$$
\begin{aligned}
&\big\|\mathbb{E}\big[(\tfrac{\partial^3}{\partial x^3}F)(x, Z_1)\big]\big\|_{L^{(3)}(\mathbb{R}^d,\mathbb{R})} = \sup_{y_1,y_2,y_3\in\mathbb{R}^d\setminus\{0\}} \frac{\big|\mathbb{E}\big[(\frac{\partial^3}{\partial x^3}F)(x, Z_1)\big](y_1, y_2, y_3)\big|}{\|y_1\|_{\mathbb{R}^d}\|y_2\|_{\mathbb{R}^d}\|y_3\|_{\mathbb{R}^d}} \\
&= \sup_{y_1,y_2,y_3\in\mathbb{R}^d\setminus\{0\}} \frac{\big|\big\langle \mathbb{E}\big[(\frac{\partial^2}{\partial x^2}G)(x, Z_1)(y_2, y_3)\big], y_1\big\rangle_{\mathbb{R}^d}\big|}{\|y_1\|_{\mathbb{R}^d}\|y_2\|_{\mathbb{R}^d}\|y_3\|_{\mathbb{R}^d}} \\
&= \sup_{y_2,y_3\in\mathbb{R}^d\setminus\{0\}} \left[\sup_{y_1\in\mathbb{R}^d\setminus\{0\}} \frac{\big|\big\langle \mathbb{E}\big[(\frac{\partial^2}{\partial x^2}G)(x, Z_1)(y_2, y_3)\big], y_1\big\rangle_{\mathbb{R}^d}\big|}{\|y_1\|_{\mathbb{R}^d}\|y_2\|_{\mathbb{R}^d}\|y_3\|_{\mathbb{R}^d}} \right] \\
&= \sup_{y_2,y_3\in\mathbb{R}^d\setminus\{0\}} \frac{\big\|\mathbb{E}\big[(\frac{\partial^2}{\partial x^2}G)(x, Z_1)(y_2, y_3)\big]\big\|_{\mathbb{R}^d}}{\|y_2\|_{\mathbb{R}^d}\|y_3\|_{\mathbb{R}^d}} = \big\|\mathbb{E}\big[(\tfrac{\partial^2}{\partial x^2}G)(x, Z_1)\big]\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)}.
\end{aligned}
\tag{5.26}
$$

This and (5.13) demonstrate that for all $x \in \mathbb{R}^d$ we have that

$$
\big\|\mathbb{E}\big[(\tfrac{\partial^2}{\partial x^2}G)(x, Z_1)\big]\big\|_{L^{(2)}(\mathbb{R}^d,\mathbb{R}^d)} + \max_{i\in\{1,2\}} \|\psi^{(i)}(x)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})} < c. \tag{5.27}
$$

Moreover, note that combining (5.14) and (5.16) ensures that for all $n \in \mathbb{N}$ we have that $\Theta_0 = \xi$ and

$$
\Theta_n = \Theta_{n-1} + \tfrac{\eta}{n^{1-\varepsilon}} G(\Theta_{n-1}, Z_n). \tag{5.28}
$$

Next observe that, e.g., [53, Lemma 4.4] proves that G is $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{S})/\mathcal{B}(\mathbb{R}^d)$-measurable. Corollary 4.7.1 (with $d = d$, $\xi = \xi$, $\Xi = \Xi$, $\varepsilon = \varepsilon$, $\eta = \eta$, $L = L$, $c = c$, $\psi = \psi$, $(S, \mathcal{S}) = (S, \mathcal{S})$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $Z_n = Z_n$, $G = G$, $g = g$, $\Theta = \Theta$ for $n \in \mathbb{N}$ in the notation of Corollary 4.7.1), (5.17), (5.18), (5.20), (5.27), (5.28), (5.11), and (5.10) therefore assure that $\{x \in \mathbb{R}^d \colon g(x) = 0\} = \{\Xi\}$ and that there exists $C \in \mathbb{R}$ such that for all $n \in \mathbb{N}$ we have that

$$
|\mathbb{E}[\psi(\Theta_n)] - \psi(\Xi)| \leq C n^{2\varepsilon-1}. \tag{5.29}
$$

The proof of Corollary 5.1.2 is thus completed. □

5.2 Weak error estimates for SGD optimization methods in the case of objective functions with linearly growing derivatives

Corollary 5.2.1. *Let $d \in \mathbb{N}$, $\xi, \Xi \in \mathbb{R}^d$, $\varepsilon \in (0, 1/2)$, $\eta, L, c \in (0, \infty)$, $\psi \in C^2(\mathbb{R}^d, \mathbb{R})$, let $(S, \mathcal{S})$ be a measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space,* *let* $F = (F(\theta, s))_{(\theta,s)\in\mathbb{R}^d\times S}\colon \mathbb{R}^d \times S \to \mathbb{R}$ *be* $(\mathcal{B}(\mathbb{R}^d)\otimes\mathcal{S})/\mathcal{B}(\mathbb{R})$*-measurable, let* $Z_n\colon \Omega \to S$, $n \in \mathbb{N}$, *be i.i.d. random variables, assume for all* $s \in S$ *that* $(\mathbb{R}^d \ni \theta \mapsto F(\theta, s) \in \mathbb{R}) \in C^3(\mathbb{R}^d, \mathbb{R})$, *assume for all* $\theta, \vartheta \in \mathbb{R}^d$ *that*

$$\mathbb{E}\big[\|(\nabla_\theta F)(\theta, Z_1)\|_{\mathbb{R}^d}^2\big] \le c\big[1 + \|\theta\|_{\mathbb{R}^d}\big]^2, \tag{5.30}$$

$$\max_{i\in\{2,3\}} \inf_{\delta\in(0,\infty)} \sup_{u\in[-\delta,\delta]^d} \mathbb{E}\big[|F(\theta, Z_1)| + \|(\tfrac{\partial^i}{\partial\theta^i}F)(\theta+u, Z_1)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})}^{1+\delta}\big] < \infty, \tag{5.31}$$

$$\langle \theta - \vartheta, \mathbb{E}[(\nabla_\theta F)(\theta, Z_1)] - \mathbb{E}[(\nabla_\theta F)(\vartheta, Z_1)]\rangle_{\mathbb{R}^d} \ge L\|\theta - \vartheta\|_{\mathbb{R}^d}^2, \tag{5.32}$$

$$\big\|\mathbb{E}\big[(\tfrac{\partial^3}{\partial\theta^3}F)(\theta, Z_1)\big]\big\|_{L^{(3)}(\mathbb{R}^d,\mathbb{R})} + \max_{i\in\{1,2\}} \|\psi^{(i)}(\theta)\|_{L^{(i)}(\mathbb{R}^d,\mathbb{R})} < c, \tag{5.33}$$

and $\|\mathbb{E}[(\nabla_\theta F)(\theta, Z_1)]\|_{\mathbb{R}^d} \le c\|\theta - \Xi\|_{\mathbb{R}^d}$ *(cf. Corollary 2.2.5), and let* $\Theta\colon \mathbb{N}_0 \times \Omega \to \mathbb{R}^d$ *be the stochastic process which satisfies for all* $n \in \mathbb{N}$ *that* $\Theta_0 = \xi$ *and*

$$\Theta_n = \Theta_{n-1} - \tfrac{\eta}{n^{1-\varepsilon}}(\nabla_\theta F)(\Theta_{n-1}, Z_n). \tag{5.34}$$

Then

(i) we have that $\{\theta \in \mathbb{R}^d\colon (\mathbb{E}[F(\theta, Z_1)] = \inf_{\vartheta\in\mathbb{R}^d} \mathbb{E}[F(\vartheta, Z_1)])\} = \{\Xi\}$ *and*

(ii) there exists $C \in \mathbb{R}$ *such that for all* $n \in \mathbb{N}$ *we have that*

$$|\mathbb{E}[\psi(\Theta_n)] - \psi(\Xi)| \le Cn^{2\varepsilon-1}. \tag{5.35}$$

Proof of Corollary 5.2.1. Throughout this proof let $f\colon \mathbb{R}^d \to \mathbb{R}$ satisfy for all $\theta \in \mathbb{R}^d$ that

$$f(\theta) = \mathbb{E}[F(\theta, Z_1)]. \tag{5.36}$$

Observe that (5.30) ensures that for all $\theta \in \mathbb{R}^d$ we have that

$$\mathbb{E}\big[\|(\nabla_\theta F)(\theta, Z_1)\|_{\mathbb{R}^d}\big] \le \sqrt{\mathbb{E}\big[\|(\nabla_\theta F)(\theta, Z_1)\|_{\mathbb{R}^d}^2\big]} \le \sqrt{c}\big[1 + \|\theta\|_{\mathbb{R}^d}\big]. \tag{5.37}$$

This and (5.31) imply that for all $\theta \in \mathbb{R}^d$ we have that

$$\mathbb{E}\big[|F(\theta, Z_1)| + \|(\nabla_\theta F)(\theta, Z_1)\|_{\mathbb{R}^d}\big] < \infty. \tag{5.38}$$

Next note that (5.30) and Lemma 3.6.1 (with $n = 2$, $p = 2$ in the notation of Lemma 3.6.1) demonstrate that for all $\theta \in \mathbb{R}^d$ we have that

$$\begin{aligned}
&\mathbb{E}\big[\|(\nabla_\theta F)(\theta, Z_1) - \mathbb{E}[(\nabla_\theta F)(\theta, Z_1)]\|_{\mathbb{R}^d}^2\big] \\
&= \mathbb{E}\big[\|(\nabla_\theta F)(\theta, Z_1)\|_{\mathbb{R}^d}^2\big] - \|\mathbb{E}[(\nabla_\theta F)(\theta, Z_1)]\|_{\mathbb{R}^d}^2 \\
&\le \mathbb{E}\big[\|(\nabla_\theta F)(\theta, Z_1)\|_{\mathbb{R}^d}^2\big] \le c\big[1 + \|\theta\|_{\mathbb{R}^d}\big]^2 \le 2c(1 + \|\theta\|_{\mathbb{R}^d}^2).
\end{aligned} \tag{5.39}$$

Combining this, (5.38), and [53, Lemma 4.8] (with $d = d$, $p = 2$, $\kappa = 2c$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $(S, \mathcal{S}) = (S, \mathcal{S})$, $X = Z_1$, $F = F$, $f = f$ in the notation of [53, Lemma 4.8]) ensures that for all $\theta \in \mathbb{R}^d$ we have that

$$f \in C^1(\mathbb{R}^d, \mathbb{R}) \qquad \text{and} \qquad (\nabla f)(\theta) = \mathbb{E}[(\nabla_\theta F)(\theta, Z_1)]. \tag{5.40}$$

This and the assumption that $\forall\, \theta \in \mathbb{R}^d\colon \|\mathbb{E}[(\nabla_\theta F)(\theta, Z_1)]\|_{\mathbb{R}^d} \leq c\|\theta - \Xi\|_{\mathbb{R}^d}$ prove that for all $\theta \in \mathbb{R}^d$ we have that

$$\|(\nabla f)(\theta)\|_{\mathbb{R}^d} \leq c\|\theta - \Xi\|_{\mathbb{R}^d}. \tag{5.41}$$

This reveals that

$$(\nabla f)(\Xi) = 0. \tag{5.42}$$

Combining this with (5.40) and (5.32) assures that for all $\theta \in \mathbb{R}^d$ we have that

$$\langle \theta - \Xi, (\nabla f)(\theta) \rangle_{\mathbb{R}^d} \geq L\|\theta - \Xi\|^2_{\mathbb{R}^d}. \tag{5.43}$$

This proves that for all $\theta \in \mathbb{R}^d$ we have that

$$\langle \theta, (\nabla f)(\theta + \Xi) \rangle_{\mathbb{R}^d} \geq L\|\theta\|^2_{\mathbb{R}^d}. \tag{5.44}$$

The fundamental theorem of calculus hence demonstrates that for all $\theta \in \mathbb{R}^d$ we have that

$$\begin{aligned} f(\theta) &= f(\Xi) + \big[f(\Xi + t(\theta - \Xi))\big]_{t=0}^{t=1} \\ &= f(\Xi) + \int_0^1 f'(\Xi + t(\theta - \Xi))(\theta - \Xi)\, dt \\ &= f(\Xi) + \int_0^1 \langle (\nabla f)(\Xi + t(\theta - \Xi)), t(\theta - \Xi) \rangle_{\mathbb{R}^d} \frac{1}{t}\, dt \\ &\geq f(\Xi) + \int_0^1 L\, \|t(\theta - \Xi)\|^2_{\mathbb{R}^d}\, \frac{1}{t}\, dt \\ &= f(\Xi) + L\, \|\theta - \Xi\|^2_{\mathbb{R}^d} \int_0^1 t\, dt = f(\Xi) + \tfrac{L}{2}\, \|\theta - \Xi\|^2_{\mathbb{R}^d}. \end{aligned} \tag{5.45}$$

The hypothesis that $L \in (0, \infty)$ therefore ensures that for all $\theta \in \mathbb{R}^d \backslash \{\Xi\}$ we have that

$$f(\theta) \geq f(\Xi) + \tfrac{L}{2}\, \|\theta - \Xi\|^2_{\mathbb{R}^d} > f(\Xi). \tag{5.46}$$

This establishes item (i). Moreover, observe that (5.41) and (5.43) ensure that for all $\theta \in \mathbb{R}^d$, $r \in (0,\infty)$ we have that

$$\begin{aligned} 2\langle \theta - \Xi, -(\nabla f)(\theta)\rangle_{\mathbb{R}^d} + r\|(\nabla f)(\theta)\|^2_{\mathbb{R}^d} &\leq -2L\|\theta - \Xi\|^2_{\mathbb{R}^d} + rc^2\|\theta - \Xi\|^2_{\mathbb{R}^d} \\ &= (rc^2 - 2L)\|\theta - \Xi\|^2_{\mathbb{R}^d}. \end{aligned} \tag{5.47}$$

This reveals that

$$\inf_{r\in(0,\infty)} \left(\sup_{\theta\in\mathbb{R}^d\setminus\{\Xi\}} \left[\frac{2\langle \theta - \Xi, -(\nabla f)(\theta)\rangle_{\mathbb{R}^d} + r\|(\nabla f)(\theta)\|^2_{\mathbb{R}^d}}{\|\theta - \Xi\|^2_{\mathbb{R}^d}} \right] \right) < 0. \tag{5.48}$$

Combining this with (5.42) and, e.g., [53, Proposition 2.16] (with $d = d$, $\vartheta = \Xi$, $\langle \cdot,\cdot \rangle = \langle \cdot,\cdot \rangle_{\mathbb{R}^d}$, $\|\cdot\| = \|\cdot\|_{\mathbb{R}^d}$, $g = -(\nabla f)$ in the notation of [53, Proposition 2.16]) prove that there exists $M \in (0,\infty)$ which satisfies for all $\theta \in \mathbb{R}^d$ that

$$\langle \theta - \Xi, (\nabla f)(\theta)\rangle_{\mathbb{R}^d} \geq M \max\{\|\theta - \Xi\|^2_{\mathbb{R}^d}, \|(\nabla f)(\theta)\|^2_{\mathbb{R}^d}\}. \tag{5.49}$$

This, (5.40), and (5.32) assure that for all $\theta, \vartheta \in \mathbb{R}^d$ we have that

$$\begin{aligned} \langle \theta - \vartheta, -(\nabla f)(\theta) + (\nabla f)(\vartheta)\rangle_{\mathbb{R}^d} &= -\langle \theta - \vartheta, (\nabla f)(\theta) - (\nabla f)(\vartheta)\rangle_{\mathbb{R}^d} \\ &\leq -L\|\theta - \vartheta\|^2_{\mathbb{R}^d} \\ &\leq -\min\{L, M\}\|\theta - \vartheta\|^2_{\mathbb{R}^d} \end{aligned} \tag{5.50}$$

and

$$\begin{aligned} \langle \theta - \Xi, -(\nabla f)(\theta)\rangle_{\mathbb{R}^d} &= -\langle \theta - \Xi, (\nabla f)(\theta)\rangle_{\mathbb{R}^d} \leq -M\|(\nabla f)(\theta)\|^2_{\mathbb{R}^d} \\ &\leq -\min\{L, M\}\|(\nabla f)(\theta)\|^2_{\mathbb{R}^d}. \end{aligned} \tag{5.51}$$

Corollary 5.1.2 (with $d = d$, $\xi = \xi$, $\Xi = \Xi$, $\varepsilon = \varepsilon$, $\eta = \eta$, $L = \min\{L, M\} \in (0,\infty)$, $c = c$, $\psi = \psi$, $(S, \mathcal{S}) = (S, \mathcal{S})$, $(\Omega, \mathcal{F}, \mathbb{P}) = (\Omega, \mathcal{F}, \mathbb{P})$, $Z_n = Z_n$, $F = -F$, $g = (\mathbb{R}^d \ni \theta \mapsto -(\nabla f)(\theta) \in \mathbb{R}^d)$, $\Theta = \Theta$ for $n \in \mathbb{N}$ in the notation of Corollary 5.1.2) therefore establishes item (ii). The proof of Corollary 5.2.1 is thus completed. □

Acknowledgments

This article is to a large extent based on the master thesis of AB written in 2017 at ETH Zurich under the supervision of AJ. Special thanks are due to Philipp Grohs for several instructive suggestions. Mike Giles is also gratefully acknowledged for several useful comments. This work has been funded by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy EXC 2044-390685587, Mathematics Münster: Dynamics-Geometry-Structure and by the Swiss National Science Foundation (SNSF) through the research grant 200020_175699.

Bibliography

[1] Aliprantis, C. D., and Border, K. C. *Infinite dimensional analysis*, third ed. Springer, Berlin, 2006.

[2] Allen-Zhu, Z., Li, Y., and Song, Z. A Convergence Theory for Deep Learning via Over-Parameterization. *arXiv:1811.03962* (2018), 37 pages.

[3] Bach, F. Adaptivity of averaged stochastic gradient descent to local strong convexity for logistic regression. *J. Mach. Learn. Res. 15* (2014), 595–627.

[4] Bach, F., and Moulines, E. Non-asymptotic analysis of stochastic approximation algorithms for machine learning. *Advances in Neural Information Processing Systems 24* (2011), 451–459.

[5] Bach, F., and Moulines, E. Non-strongly-convex smooth stochastic approximation with convergence rate $O(1/n)$. *Proceedings of the 26th International Conference on Neural Information Processing Systems 1* (2013), 773–781.

[6] Bayer, C., Horvath, B., Muguruza, A., Stemper, B., and Tomas, M. On deep calibration of (rough) stochastic volatility models. *arXiv:1908.08806* (2019), 32 pages.

[7] Beck, C., Becker, S., Cheridito, P., Jentzen, A., and Neufeld, A. Deep splitting method for parabolic PDEs. *arXiv:1907.03452* (2019), 40 pages.

[8] Beck, C., Becker, S., Grohs, P., Jaafari, N., and Jentzen, A. Solving stochastic differential equations and Kolmogorov equations by means of deep learning. *arXiv:1806.00421* (2018), 56 pages.

[9] Beck, C., E, W., and Jentzen, A. Machine Learning Approximation Algorithms for High-Dimensional Fully Nonlinear Partial Differential Equations

and Second-order Backward Stochastic Differential Equations. *J. Nonlinear Sci. 29*, 4 (2019), 1563–1619.

[10] Becker, S., Cheridito, P., and Jentzen, A. Deep optimal stopping. *J. Mach. Learn. Res. 20*, 74 (2019), 1–25.

[11] Becker, S., Cheridito, P., and Jentzen, A. Pricing and hedging American-style options with deep learning. *arXiv:1912.11060* (2019), 12 pages.

[12] Becker, S., Cheridito, P., Jentzen, A., and Welti, T. Solving high-dimensional optimal stopping problems using deep learning. *arXiv:1908.01602* (2019), 42 pages.

[13] Bengio, Y., Boulanger-Lewandowski, N., and Pascanu, R. Advances in optimizing recurrent networks. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 8624–8628.

[14] Bercu, B., and Fort, J.-C. Generic stochastic gradient methods. *Wiley Encyclopedia of Operations Research and Management Science* (2013), 1–8.

[15] Bordes, A., Bottou, L., and Gallinari, P. SGD-QN: Careful quasi-Newton stochastic gradient descent. *J. Mach. Learn. Res. 10* (2009), 1737–1754.

[16] Bottou, L. Large-scale machine learning with stochastic gradient descent. *Proceedings of COMPSTAT 2010* (2010), 177–186.

[17] Bottou, L., and Bousquet, O. The tradeoffs of large scale learning. *Optimization for Machine Learning, MIT Press* (2011), 351–368.

[18] Bottou, L., Curtis, F. E., and Nocedal, J. Optimization methods for large-scale machine learning. *SIAM Rev. 60*, 2 (2018), 223–311.

[19] Bottou, L., and LeCun, Y. Large scale online learning. *Advances in Neural Information Processing Systems 16* (2004), 217–224.

[20] Brutzkus, A., Globerson, A., Malach, E., and Shalev-Shwartz, S. SGD Learns Over-parameterized Networks that Provably Generalize on Linearly Separable Data. *arXiv:1710.10174* (2017), 17 pages.

[21] Buehler, H., Gonon, L., Teichmann, J., and Wood, B. Deep hedging. *Quant. Finance 19*, 8 (2019), 1271–1291.

[22] Chau, N. H., Moulines, E., Rásonyi, M., Sabanis, S., and Zhang, Y. On stochastic gradient Langevin dynamics with dependent data streams: the fully non-convex case. *arXiv:1905.13142* (2019), 31 pages.

[23] Cheridito, P., Jentzen, A., and Rossmannek, F. Non-convergence of stochastic gradient descent in the training of deep neural networks. *arXiv:2006.07075* (2020), 12 pages.

[24] Chouiekh, A., and Haj, E. H. I. E. ConvNets for Fraud Detection analysis. *Procedia Computer Science 127* (2018), 133–138.

[25] Coleman, R. *Calculus on normed vector spaces.* Universitext. Springer, New York, 2012.

[26] Dahl, G. E., Sainath, T. N., and Hinton, G. E. Improving deep neural networks for lvcsr using rectified linear units and dropout. In *IEEE International Conference on Acoustics, Speech and Signal Processing* (2013), pp. 8609–8613.

[27] Dahl, G. E., Yu, D., Deng, L., and Acero, A. Context-dependent pre-trained deep neural networks for large-vocabulary speech recognition. *IEEE Transactions on audio, speech, and language processing 20*, 1 (2012), 30–42.

[28] Darken, C., Chang, J., and Moody, J. Learning rate schedules for faster stochastic gradient search. *Neural Networks for Signal Processing II Proceedings of the 1992 IEEE Workshop* (1992), 1–11.

[29] Dauphin, Y., de Vries, H., and Bengio, Y. Equilibrated adaptive learning rates for non-convex optimization. *Advances in Neural Information Processing Systems 28* (2015), 1504–1512.

[30] Dauphin, Y. N., Pascanu, R., Gulcehre, C., Cho, K., Ganguli, S., and Bengio, Y. Identifying and attacking the saddle point problem in high-dimensional non-convex optimization. *Proceedings of the 27th International Conference on Neural Information Processing Systems* (2014), 2933–2941.

[31] Dean, J., Corrado, G. S., Monga, R., Chen, K., Devin, M., Le, Q. V., Mao, M. Z., Ranzato, M. A., Senior, A., Tucker, P., Yang, K., and Ng, A. Y. Large scale distributed deep networks. *Advances in Neural Information Processing Systems 25* (2012), 1–11.

[32] Défossez, A., and Bach, F. AdaBatch: Efficient Gradient Aggregation Rules for Sequential and Parallel Stochastic Gradient Methods. *arXiv:1711.01761* (2017), 26 pages.

[33] Deng, L., Li, J., Huang, J.-T., Yao, K., Yu, D., Seide, F., Seltzer, M., Zweig, G., He, X., and Williams, J. Recent advances in deep learning for speech research at Microsoft. *Acoustics, Speech, and Signal Processing (ICASSP)* (2013).

[34] Dereich, S., and Müller-Gronbach, T. General multilevel adaptations for stochastic approximation algorithms of Robbins-Monro and Polyak-Ruppert type. *Numer. Math. 142*, 2 (2019), 279–328.

[35] Dieuleveut, A., Durmus, A., and Bach, F. Bridging the gap between constant step size stochastic gradient descent and Markov chains. *arXiv:1707.06386* (2017), 49 pages.

[36] Dozat, T. Incorporating Nesterov Momentum into Adam. *ICLR Workshop* (2016), 2013–2016.

[37] Duchi, J., Hazan, E., and Singer, Y. Adaptive subgradient methods for online learning and stochastic optimization. *J. Mach. Learn. Res. 12* (2011), 2121–2159.

[38] Durrett, R. *Probability: Theory and Examples.* Cambridge University Press, 2010.

[39] E, W., Han, J., and Jentzen, A. Deep learning-based numerical methods for high-dimensional parabolic partial differential equations and backward stochastic differential equations. *Commun. Math. Stat. 5*, 4 (2017), 349–380.

[40] E, W., and Yu, B. The deep Ritz method: a deep learning-based numerical algorithm for solving variational problems. *Commun. Math. Stat. 6*, 1 (2018), 1–12.

[41] Fehrman, B., Gess, B., and Jentzen, A. Convergence rates for the stochastic gradient descent method for non-convex objective functions. *Accepted in J. Mach. Learn. Res., arXiv:1904.01517* (2019), 52 pages.

[42] Ghadimi, S., Lan, G., and Zhang, H. Mini-batch stochastic approximation methods for nonconvex stochastic composite optimization. *Math. Program. 155*, 1-2, Ser. A (2016), 267–305.

[43] Graves, A. Generating Sequences With Recurrent Neural Networks. *arXiv:1308.0850* (2013), 43 pages.

[44] Graves, A., Mohamed, A.-r., and Hinton, G. Speech recognition with deep recurrent neural networks. *Acoustics, Speech and Signal Processing (ICASSP)* (2013), 6645–6649.

[45] Han, J., Jentzen, A., and E, W. Solving high-dimensional partial differential equations using deep learning. *Proceedings of the National Academy of Sciences 115*, 34 (2018), 8505–8510.

[46] Henry-Labordère, P. Deep Primal-Dual Algorithm for BSDEs: Applications of Machine Learning to CVA and IM. (November 15, 2017), 16 pages. Available at SSRN: https://ssrn.com/abstract=3071506.

[47] Hinton, G., Deng, L., Yu, D., Dahl, G. E., Mohamed, A.-r., Jaitly, N., Senior, A., Vanhoucke, V., Nguyen, P., and Sainath, T. N. Deep neural networks for acoustic modeling in speech recognition: The shared views of four research groups. *Signal Processing Magazine, IEEE 29*, 6 (2012), 82–97.

[48] Hinton, G. E., and Salakhutdinov, R. R. Reducing the dimensionality of data with neural networks. *Science 313*, 5786 (2006), 504–507.

[49] Hu, B., Lu, Z., Li, H., and Chen, Q. Convolutional neural network architectures for matching natural language sentences. *Proceedings of the 27th International Conference on Neural Information Processing Systems 2* (2014), 2042–2050.

[50] Huang, G., Liu, Z., van der Maaten, L., and Weinberger, K. Q. Densely connected convolutional networks. *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition* (2017), 2261–2269.

[51] Inoue, M., Park, H., and Okada, M. On-line learning theory of soft committee machines with correlated hidden units steepest gradient descent and natural gradient descent. *Journal of the Physical Society of Japan 72*, 4 (2003), 805–810.

[52] Ioffe, S., and Szegedy, C. Batch normalization: Accelerating deep network training by reducing internal covariate shift. *Proceedings of the 32nd International Conference on Machine Learning* (2015), 448–456.

[53] Jentzen, A., Kuckuck, B., Neufeld, A., and von Wurstemberger, P. Strong error analysis for stochastic gradient descent optimization algorithms. *Accepted in IMA J. Num. Anal., arXiv:1801.09324* (2018), 75 pages.

[54] Jentzen, A., Salimova, D., and Welti, T. Strong convergence for explicit space-time discrete numerical approximation methods for stochastic Burgers equations. *J. Math. Anal. Appl. 469*, 2 (2019), 661–704.

[55] Jentzen, A., and von Wurstemberger, P. Lower error bounds for the stochastic gradient descent optimization algorithm: Sharp convergence rates for slowly and fast decaying learning rates. *J. Complexity 57* (2020), 101438, 16.

[56] Johnson, R., and Zhang, T. Accelerating stochastic gradient descent using predictive variance reduction. *Advances in Neural Information Processing Systems 26* (2013), 315–323.

[57] Kalchbrenner, N., Grefenstette, E., and Blunsom, P. A convolutional neural network for modelling sentences. *Proceedings of the 52nd Annual Meeting of the Association for Computational Linguistics* (2014), 655–665.

[58] Kingma, D. P., and Ba, J. Adam: A method for stochastic optimization. *arXiv:1412.6980* (2014), 15 pages.

[59] Klenke, A. *Probability theory*, second ed. Universitext. Springer, London, 2014. A comprehensive course.

[60] Kloeden, P. E., and Platen, E. *Numerical solution of stochastic differential equations*, vol. 23 of *Applications of Mathematics (New York)*. Springer-Verlag, Berlin, 1992.

[61] Königsberger, K. *Analysis. 2*, fifth ed. Springer-Lehrbuch. [Springer Textbook]. Springer-Verlag, Berlin, 2004.

[62] Krizhevsky, A., Sutskever, I., and Hinton, G. E. ImageNet classification with deep convolutional neural networks. *Commun. ACM 60*, 6 (2017), 84–90.

[63] Lan, G., and Zhou, Y. An optimal randomized incremental gradient method. *Math. Program. 171*, 1-2, Ser. A (2018), 167–215.

[64] Langford, J., Li, L., and Zhang, T. Sparse online learning via truncated gradient. *J. Mach. Learn. Res. 10* (2009), 777–801.

[65] Le Roux, N., Schmidt, M., and Bach, F. A stochastic gradient method with an exponential convergence rate for strongly-convex optimization with finite training sets. *Proceedings of the 25th International Conference on Neural Information Processing Systems 2* (2012), 2663–2671.

[66] LeCun, Y., Bottou, L., Bengio, Y., and Haffner, P. Gradient-based learning applied to document recognition. *Proceedings of the IEEE 86*, 11 (1998), 2278–2324.

[67] LeCun, Y., Bottou, L., Orr, G. B., and Müller, K. R. Efficient backprop. *Neural Networks: Tricks of the Trade* (1998), 9–50.

[68] Lei, Y., Hu, T., Li, G., and Tang, K. Stochastic gradient descent for nonconvex learning without bounded gradient assumptions. *IEEE Transactions on Neural Networks and Learning Systems* (2019), 1–7.

[69] Li, Q., Tai, C., and E, W. Stochastic modified equations and adaptive stochastic gradient algorithms. *Proceedings of the 34th International Conference on Machine Learning* (2017), 2101–2110.

[70] Li, Y., and Liang, Y. Learning overparameterized neural networks via stochastic gradient descent on structured data. *Proceedings of the 32nd International Conference on Neural Information Processing Systems* (2018), 8168–8177.

[71] Liu, S., Borovykh, A., Grzelak, L. A., and Oosterlee, C. W. A neural network-based framework for financial model calibration. *J. Math. Ind. 9* (2019), Paper No. 9.

[72] Lovas, A., Lytras, I., Rásonyi, M., and Sabanis, S. Taming neural networks with TUSLA: Non-convex learning via adaptive stochastic gradient Langevin algorithms. *arXiv:2006.14514* (2020), 29 pages.

[73] Mcmahan, H. B., and Streeter, M. Delay-tolerant algorithms for asynchronous distributed online learning. *Advances in Neural Information Processing Systems 27* (2014), 1–9.

[74] Mishra, S. A machine learning framework for data driven acceleration of computations of differential equations. *Mathematics in Engineering 1*, 1 (2019), 118–146.

[75] Mizutani, E., and Dreyfus, S. An analysis on negative curvature induced by singularity in multi-layer neural-network learning. *Advances in Neural Information Processing Systems 23* (2010), 1669–1677.

[76] Müller-Gronbach, T., and Ritter, K. Minimal errors for strong and weak approximation of stochastic differential equations. In *Monte Carlo and quasi-Monte Carlo methods 2006*. Springer, Berlin, 2008, pp. 53–82.

[77] Nabian, M. A., and Meidani, H. A deep learning solution approach for high-dimensional random differential equations. *Probabilistic Engineering Mechanics 57* (2019), 14–25.

[78] Nguyen, L. M., Nguyen, N. H., Phan, D. T., Kalagnanam, J. R., and Scheinberg, K. When does stochastic gradient algorithm work well? *arXiv:1801.06159* (2018), 21 pages.

[79] Niu, F., Recht, B., Christopher, R., and Wright, S. J. HOGWILD!: A Lock-Free Approach to Parallelizing Stochastic Gradient Descent. *Proceedings of the 24th International Conference on Neural Information Processing Systems* (2011), 693–701.

[80] Pascanu, R., and Bengio, Y. Revisiting natural gradient for deep networks. *International Conference on Learning Representations* (2014).

[81] Pillaud-Vivien, L., Rudi, A., and Bach, F. Exponential convergence of testing error for stochastic gradient methods. *Proceedings of the 31st Conference On Learning Theory 75* (2018), 250–296.

[82] Polyak, B. T. A new method of stochastic approximation type. *Avtomat. i Telemekh. 51*, 7 (1998), 937–1008.

[83] Polyak, B. T., and Juditsky, A. B. Acceleration of stochastic approximation by averaging. *Automat. Remote Control 30*, 4 (1992), 838–855.

[84] Polyak, B. T., and Tsypkin, Y. Z. Optimal pseudogradient adaptation algorithms. *Avtomat. i Telemekh. 8* (1980), 74–84.

[85] Rakhlin, A., Shamir, O., and Sridharan, K. Making gradient descent optimal for strongly convex stochastic optimization. *Proceedings of the 29th International Conference on Machine Learning* (2012), 1571–1578.

[86] Rattray, M., Saad, D., and Amari, S. I. Natural gradient descent for on-line learning. *Phys. Rev. Lett. 81*, 24 (1998), 5461–5464.

[87] Rössler, A. *Runge-Kutta Methods for the Numerical Solution of Stochastic Differential Equations.* Shaker Verlag, Aachen, 2003. Dissertation, Technische Universität Darmstadt, Darmstadt, Germany.

[88] Roy, A., Sun, J., Mahoney, R., Alonzi, L., Adams, S., and Beling, P. Deep learning detecting fraud in credit card transactions. In *Systems and Information Engineering Design Symposium* (2018), pp. 129–134.

[89] Ruder, S. An overview of gradient descent optimization algorithms. *arXiv:1609.04747* (2016), 12 pages.

[90] Sankararaman, K. A., Soham De, Z. X., Huang, W. R., and Goldstein, T. The Impact of Neural Network Overparameterization on Gradient Confusion and Stochastic Gradient Descent. *arXiv:1904.06963* (2019), 37 pages.

[91] Schaul, T., Zhang, S., and LeCun, Y. No more pesky learning rates. *Proceedings of the 30th International Conference on Machine Learning 28*, 3 (2013), 343–351.

[92] Schraudolph, N. N. Local gain adaptation in stochastic gradient descent. *Artificial Neural Networks, 1999. ICANN 99. Ninth International Conference on (Conf. Publ. No. 470) 2* (1999), 569–574.

[93] Schraudolph, N. N., Yu, J., and Günter, S. A stochastic quasi-newton method for online convex optimization. *Proceedings of the 9th International Conference on Artificial Intelligence and Statistics (AISTAT)* (2007), 433–440.

[94] Shalev-Shwartz, S., Shinger, Y., Srebro, N., and Cotter, A. Pegasos: primal estimated sub-gradient solver for SVM. *Math. Program. 127*, 1 (2011), 3–30.

[95] Simonyan, K., and Zisserman, A. Very deep convolutional networks for large-scale image recognition. *arXiv:1409.1556* (2014), 14 pages.

[96] Sirignano, J., and Cont, R. Universal features of price formation in financial markets: perspectives from deep learning. *Quant. Finance 19*, 9 (2019), 1449–1459.

[97] Sirignano, J., and Spiliopoulos, K. DGM: a deep learning algorithm for solving partial differential equations. *J. Comput. Phys. 375* (2018), 1339–1364.

[98] Sohl-Dickstein, J., Poole, B., and Ganguli, S. Fast large-scale optimization by unifying stochastic gradient and quasi-Newton methods. *Proceedings of the 31st International Conference on Machine Learning* (2014), 604–612.

[99] Sutskever, I., Martens, J., Dahl, G., and Hinton, G. On the importance of initialization and momentum in deep learning. *Proceedings of the 30th International Conference on Machine Learning* (2013), 1139–1147.

[100] Taigman, Y., Yang, M., Ranzato, M., and Wolf, L. Deepface: Closing the gap to human-level performance in face verification. In *IEEE Conference on Computer Vision and Pattern Recognition* (2014), pp. 1701–1708.

[101] Tang, C., and Monteleoni, C. On the convergence rate of stochastic gradient descent for strongly convex functions. In *Regularization, optimization, kernels, and support vector machines*, Chapman & Hall/CRC Mach. Learn. Pattern Recogn. Ser. CRC Press, Boca Raton, FL, 2015, pp. 159–175.

[102] Teschl, G. *Ordinary differential equations and dynamical systems*, vol. 140 of *Graduate Studies in Mathematics*. American Mathematical Society, Providence, RI, 2012.

[103] Wang, R., Fu, B., Fu, G., and Wang, M. Deep & cross network for ad click predictions. *Proceedings of the 26th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2017), 1–7.

[104] Wang, W., Yang, J., Xiao, J., Li, S., and Zhou, D. Face recognition based on deep learning. In *Human Centered Computing* (2015), pp. 812–820.

[105] Woodworth, B. E., and Srebro, N. Tight complexity bounds for optimizing composite objectives. *Advances in Neural Information Processing Systems 29* (2016), 3639–3647.

[106] Wu, C., Karanasou, P., Gales, M. J., and Sim, K. C. Stimulated deep neural network for speech recognition. In *Interspeech 2016* (2016), pp. 400–404.

[107] Xu, W. Towards optimal one pass large scale learning with averaged stochastic gradient descent. *arXiv:1107.2490* (2011), 19 pages.

[108] Zaremba, W., and Sutskever, I. Learning to execute. *arXiv:1410.4615* (2014), 25 pages.

[109] Zeiler, M. D. ADADELTA: An adaptive learning rate method. *arXiv:1212.5701* (2012), 6 pages.

[110] Zhai, S., Chang, K.-h., Zhang, R., and Zhang, Z. M. DeepIntent: Learning attentions for online advertising with recurrent neural networks. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining* (2016), 1295–1304.

[111] Zhang, S., Choromanska, A., and LeCun, Y. Deep learning with elastic averaging SGD. *Proceedings of the 28th International Conference on Neural Information Processing Systems* (2015), 685–693.

[112] Zhang, T. Solving large scale linear prediction problems using stochastic gradient descent algorithms. *Proceedings of the 21st International Conference on Machine Learning* (2004).