



Monte Carlo Methods and Deep Neural Network Approximations for High-Dimensional Markov Decision Processes and Elliptic Partial Differential Equations

Dissertation

zur Erlangung des akademischen Grades
Doktor der Naturwissenschaften (Dr. rer. nat.)

vorgelegt von

Konrad Malte Kleinberg

August 2026

Betreuer: Prof. Dr. Thomas Kruse

Acknowledgments

PhD studies are among the most difficult challenges I have encountered in my life so far. I owe my sincere gratitude to the people around me who provided me with the necessary academic and, even more important, social support. I deeply value the knowledge, time, and encouragement you have shared with me.

First of all, I thank my advisor Prof. Dr. Thomas Kruse for the continuous support I received throughout my studies. It is remarkable that you always had the advice I needed no matter the situation. Thank you very much for your guidance, communication, and patience!

I express my deepest gratitude to Prof. Dr. Michael Günther, Prof. Dr. Matthias Ehrhardt, apl. Prof. Dr. Andreas Bartel, and PD Dr. Long Teng for all their professional advice, wisdom, and honest encouragement.

I thank Prof. Dr. Arnulf Jentzen very much for our collaborations. Your expertise greatly broadened my knowledge beyond mathematics.

Next, I want to thank the German Research Foundation (DFG) for funding the SPP2298, which I am very grateful to have been a part of. The support allowed me to connect with fellow researchers who enriched my academic experience.

I extend my gratitude to Prof. Dr. Lars Grüne and to Mario Sperl, whom I met through the SPP2298. Thank you for the fruitful collaboration and the pleasant time I had during my stay in Bayreuth.

To all current and former members of the Applied and Computational Mathematics group, whom I had the pleasure of meeting, I express my heartfelt thanks. I thank you especially for all the moments in which we could wind down together, whether during a much needed coffee break or a night out. Trust me when I say that each and every one of you has left a lasting impression on me, and I am deeply grateful for that. For my studies, I could not have imagined a better environment.

Next, I would like to thank my friends. From the beginning of my time in Wuppertal, I felt your acceptance and support. I thank you for all the time we spent together.

I sincerely thank my family, especially my parents, Ulla and Joachim, and my sister, Kamilla, for their unconditional and enduring love and support.

Finally, I thank my partner, Michelle, for being there for me. I thank you for the many things I have learned through our time together and for helping me trust myself.

Contents

1. Introduction	1
1.1. Motivation	1
1.2. Two High-Dimensional Problems	2
1.3. A Common Approximation Strategy	3
1.4. Literature	4
1.5. Contributions of the Dissertation	5
1.6. Organization	7
 I. Monte Carlo Methods and Deep Neural Network Approximations for High-Dimensional Markov Decision Processes	 9
2. Nonlinear Monte Carlo Methods for Bellman Equations of High-Dimensional Markov Decision Processes	11
2.1. Existence and Uniqueness of Solutions of Functional Fixed-Point Equations . .	16
2.2. Full-History Recursive Multilevel Fixed-Point (MLFP) Approximations	20
2.2.1. Mathematical Description of MLFP Approximations	20
2.2.2. Measurability and Distributional Properties for MLFP Approximations .	21
2.2.3. Integrability Properties for MLFP Approximations	23
2.2.4. Recursive Error Bounds for MLFP Approximations	25
2.2.5. Non-Recursive Error Bounds for MLFP Approximations	27
2.3. Computational Complexity Analysis for MLFP Approximations	32
2.3.1. MLFP Approximations for Functional Fixed-Point Equations	32
2.3.2. MLFP Approximations for Bellman Equations of Optimal Control Problems	35
2.3.3. MLFP Approximations for Bellman Equations of Optimal Stopping Problems	37
 3. Artificial Neural Network (ANN) Calculus	 41
3.1. Structured Descriptions of ANNs	41
3.2. Compositions, Affine Linear Transformations, Powers, and Extensions of ANNs	42
3.3. Parallelizations of ANNs	44
3.4. Linear Combinations of ANNs	46
 4. Deep Neural Network Approximations for Bellman Equations of High-Dimensional Markov Decision Processes	 49
4.1. A Stability Result for Solutions of Certain Stochastic Fixed-Point Equations . .	51
4.2. Auxiliary ANN Representations and Approximations	60
4.2.1. ANN Representations and Approximations for the Maximum Function .	60
4.2.2. ANN Representations for MLFP Approximations	65
4.3. ANN Approximations for Functional Fixed-Point Equations	73
4.3.1. ANN Approximations for Functional Fixed-Point Equations with General Activation Functions	73
4.3.2. ANN Approximations for Functional Fixed-Point Equations with Specific Activation Functions and One-Step Transitions	80

4.3.3. ANN Approximations for Bellman Equations with Specific Activation Functions and One-Step Transitions	83
---	----

II. Monte Carlo and Deep Neural Network Approximations of Elliptic PDEs with Drift and Killing 89

5. Walk-on-Spheres Monte Carlo Methods for Elliptic PDEs with Drift and Killing	91
5.1. A Girsanov-Based Feynman-Kac Representation	93
5.2. A Modified Walk-on-Spheres Algorithm	96
5.3. Error Analysis for the Modified Walk-on-Spheres Algorithm	102
5.3.1. Bias Estimates	102
5.3.2. Global L^∞ -Error Analysis	105
6. Deep Neural Network Approximation of Elliptic PDEs with Drift and Killing	115
6.1. ANN Approximations of Modified Walk-on-Spheres Algorithms	115
6.2. ANN Approximations for Solutions of Elliptic PDEs	118
7. Conclusion and Outlook	123

Chapter 1.

Introduction

1.1. Motivation

Many problems in stochastic control and *partial differential equations* (PDEs) revolve around the approximation of functions defined on high-dimensional state spaces. In these cases, the dimension is not only a parameter in the mathematical formulation of the problem, but rather has a decisive influence in the viability of numerical approximation methods.

Classical deterministic algorithms frequently rely on a discretization of the domain the target quantity is defined on. A grid with $n \in \mathbb{N} = \{1, 2, \dots\}$ points along each coordinate direction of a space of dimension $d \in \mathbb{N}$ contains n^d points and, hence, the corresponding computational effort and storage requirement may exhibit an exponential growth with respect to the dimension. Even if a numerical approximation scheme converges for each fixed dimension it still might become infeasible for high-dimensional problem instances. This rapid decline in computational feasibility with increasing dimension is commonly associated with the *curse of dimensionality* (cf., e.g., Bellman [14], Novak & Woźniakowski [82, Chapter 1], and Novak & Ritter [81]). This motivates the development of numerical approximation algorithms that provide control over both the approximation accuracy and the required computational complexity.

Monte Carlo methods offer a possible alternative to a grid discretization of the state space. They explore the domain of the target function by using random samples and can therefore be free from grid-based discretization, which makes them attractive for application in high-dimensional settings (see, e.g., Müller-Gronbach et al. [75]). Nevertheless, replacing grid-based methods with sampling-based variants does not automatically ensure scalability to high dimensions. This scalability of sampling-based approaches has to be established through a rigorous analysis of their approximation error and total computational complexity, including the generation of all random samples appearing in the approximation.

Artificial neural networks (ANNs) provide a flexible class of approximators for functions defined on high-dimensional domains that does not necessarily require a grid-based discretization. While universal approximation theorems qualitatively demonstrate their expressive power (see, e.g., Cybenko [29] and Hornik et al. [47]), they do not provide quantitative estimates on how the required network complexity depends on the approximation accuracy and the input dimension by themselves. For high-dimensional applications, quantitative estimates on both the approximation error and network complexity are necessary (see, e.g., Mhaskar [72]). If the components of a numerical approximation scheme can be represented or suitably approximated by neural networks, the structure of the approximation scheme can be used to study the artificial neural network complexity required to approximate the target function. This enables a transfer from an algorithmic to a representational perspective. In particular, the focus here is on representation and approximation results rather than on the analysis of neural network training algorithms.

In view of these considerations, this thesis investigates Monte Carlo and ANN approximations for two problem classes emerging from *Markov decision processes* (MDPs) and PDEs.

1.2. Two High-Dimensional Problems

This thesis considers two classes of high-dimensional approximation problems. The first problem concerns the numerical approximation of solutions of functional fixed-point equations motivated by MDPs. The second problem arises from a class of elliptic PDEs. We now briefly explain each problem as well as the associated target quantity considered in this thesis.

Markov Decision Processes

MDPs serve as mathematical models for tasks requiring sequential decision making under uncertainty and as such constitute a class of models in stochastic optimal control (see, e.g., [18, 85]). Moreover, they provide the mathematical foundations for *reinforcement learning* (see, e.g., [17, 19, 100]). In each time epoch, the system is described by a state and an action is selected from a set of admissible actions. The evolution to the next state is governed by a random transition that depends on the current state and the chosen action. Each pair of state and action is associated with an immediate reward. The overall objective is to sequentially select the actions in order to maximize the expected accumulated reward.

We consider in this thesis discrete time, discounted, infinite horizon Markov decision processes with a finite action set A on a measurable state space $\mathbb{X}$. Initially, the state space $\mathbb{X}$ is allowed to be a general measurable space. In view of high-dimensional approximation, the main case of interest is the scenario where the state space is $\mathbb{X} = \mathbb{R}^d$, $d \in \mathbb{N}$. Under optimal choices of actions, the total discounted accumulated reward obtained from a starting state $x \in \mathbb{X}$ is represented by the value $V(x)$, where $V: \mathbb{X} \rightarrow \mathbb{R}$ denotes the so-called *value function*. Under appropriate conditions, the value function satisfies a fixed-point equation, the so-called *Bellman equation* (see, e.g., [14, 18]):

$$V(x) = \max_{a \in A} (g(x, a) + \delta \mathbb{E}[V(X^{x,a})]), \quad x \in \mathbb{X}. \quad (1.1)$$

Here $g(x, a)$ denotes the reward associated to the state and action pair $(x, a) \in \mathbb{X} \times A$, $\delta \in (0, 1)$ is the discount factor, and $X^{x,a}$ denotes the random next state obtained from choosing action a in state x . Directly related to the value function V is the so-called Q -function, $Q: \mathbb{X} \times A \rightarrow \mathbb{R}$, given by the relation $Q(x, a) = g(x, a) + \delta \mathbb{E}[V(X^{x,a})]$, $x \in \mathbb{X}$, $a \in A$. Together with (1.1) this yields for all $x \in \mathbb{X}$ that $V(x) = \max_{a \in A} Q(x, a)$ and, hence, Q satisfies the Bellman equation

$$Q(x, a) = g(x, a) + \delta \mathbb{E}[\max_{b \in A} Q(X^{x,a}, b)], \quad x \in \mathbb{X}, \quad a \in A. \quad (1.2)$$

This fixed-point structure motivates the more general analysis of functions $u: \mathbb{X} \rightarrow \mathbb{R}^{|A|}$ that satisfy the functional fixed-point equation

$$(u(x))(a) = \mathbb{E}[f(X^{x,a}, u(X^{x,a}))], \quad x \in \mathbb{X}, \quad a \in A, \quad (1.3)$$

where $f: \mathbb{X} \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$. The Bellman equation (1.2) for the Q -function is recovered by a specific choice of f in terms of g and δ . In particular, for $f(x, r) = \delta \max_{a \in A} (g(x, a) + r(a))$ it holds that $(x, a) \mapsto g(x, a) + (u(x))(a)$ satisfies (1.2). Discrete time optimal stopping problems arise as a further specialization by considering the two actions “stop” and “continue”.

The objective to approximate functions u that satisfy the functional fixed-point equation (1.3) where u is defined on a high-dimensional state space, such as $\mathbb{R}^d$, is challenging due to the curse of dimensionality. The relevant question therefore is whether these functions can be numerically approximated to a prescribed approximation accuracy while the computational effort remains controlled in the dimension of the state space.

Elliptic Partial Differential Equations on Bounded Domains

The second problem class we study in this thesis concerns the solutions of linear drift-diffusion elliptic PDEs with constant coefficients on high-dimensional bounded domains of the following type:

$$\begin{aligned} \frac{1}{2} \operatorname{tr}(\sigma \sigma^\top D^2 u) + \langle b, \nabla u \rangle - ru + f &= 0 & \text{on } D, \\ u|_{\partial D} &= g & \text{on } \partial D. \end{aligned} \quad (1.4)$$

Here $d \in \mathbb{N}$, $D \subseteq \mathbb{R}^d$ is open, bounded, and convex, $u: \overline{D} \rightarrow \mathbb{R}$ is the unknown, $\sigma \in \mathbb{R}^{d \times d}$, $b \in \mathbb{R}^d$, $r \in [0, \infty)$ are the coefficients, $f: D \rightarrow \mathbb{R}$ denotes the source, and $g: \partial D \rightarrow \mathbb{R}$ describes the boundary data. Under suitable assumptions, solutions of the PDE (1.4) admit a stochastic representation via functions $w: D \rightarrow \mathbb{R}$,

$$w(x) = \mathbb{E}[h_0(\tau^x, x + B_{\tau^x})] + \mathbb{E}\left[\int_0^{\tau^x} h_1(s, x + B_s) \, ds\right], \quad x \in D, \quad (1.5)$$

where $h_0, h_1: [0, \infty) \times \overline{D} \rightarrow \mathbb{R}$, B denotes a standard d -dimensional Brownian motion starting in 0, and $\tau^x, x \in D$, denotes the first exit time of the process $x + B$ from the domain D . The analysis for this problem class is conducted for expectations of the form (1.5).

The challenge of the numerical approximation of functions of the form (1.5) as target quantities again lies in the dimensionality of the domain $D \subseteq \mathbb{R}^d$. The objective considered here is to establish a global approximation on the entire, possibly high-dimensional, domain D , with quantitative error estimates and complexity bounds that depend at most polynomially on the reciprocal of the approximation accuracy and on the dimension of the domain.

Although both problem classes differ in their origin as well as in the mathematical objects describing them, they share the core challenge of approximating target functions defined on high-dimensional spaces. For both approximation tasks, the aim is to establish quantitative error and complexity results. To achieve this, we employ a common approximation strategy.

1.3. A Common Approximation Strategy

The numerical approximation for both problem classes, as presented here, follows a common two-stage strategy. In the first stage, problem specific Monte Carlo methods are constructed to approximate the respective target functions and quantitative error and complexity analyses for the resulting estimators are conducted. The central aim is to determine how the required computational effort depends on the prescribed approximation accuracy and on the dimension of the underlying state space or domain.

For the first problem class related to MDPs, the Monte Carlo approximation is based on the fixed-point structure of the target function given in (1.3) and incorporates ideas from the *full-history recursive multilevel Picard* (MLP) approximation methods (see, e.g., [31, 32, 51, 54]). Estimators for Q -functions of associated MDPs are obtained by suitable choices of the nonlinearity in (1.3).

For the second problem class, the Monte Carlo approximations are constructed for expectations of the form (1.5), which arise from suitable stochastic representations of solutions of elliptic PDEs of the type (1.4). More precisely, they are built on the so-called *Walk-on-Spheres* algorithm (see [74]), which constitutes a numerically efficient way to approximate exit points of Brownian motion from the underlying domain.

The concrete construction and the corresponding error analysis of the Monte Carlo estimators for the respective problems differ substantially. Nevertheless, in both cases probabilistic representations or fixed-point structures are exploited. This leads to probabilistic methods, for which an error and complexity analysis is carried out explicitly. Besides building a basis

for the subsequent ANN approximation results, the constructed Monte Carlo methods are therefore of independent interest.

In the second stage, the constructed Monte Carlo methods are used to derive approximation results for ANNs. While the Monte Carlo analysis of the functional fixed-point equation (1.3) is formulated for general measurable state spaces $\mathbb{X}$, subsequent ANN approximation results specialize to the case of Euclidean state spaces $\mathbb{R}^d$. To this end, ANNs are shown to structurally represent and approximate the respective Monte Carlo estimators while quantitatively controlling the network complexity. Together with the error and complexity analyses of the respective Monte Carlo methods, this yields ANN approximations for both problem classes.

1.4. Literature

In this section, we present the major lines of research that the material developed in this thesis connects to.

The curse of dimensionality as described above constitutes a central difficulty in the numerical approximation of high-dimensional problems. In this context, a fundamental objective is to construct approximation methods that are polynomially tractable in the sense that the computational complexity required to achieve a prescribed approximation accuracy grows at most polynomially in the underlying problem dimension and in the reciprocal of the approximation accuracy (see, e.g., Novak & Woźniakowski [82, Chapter 1], and Novak & Ritter [81]).

An important class of probabilistic approximation methods developed for this purpose is given by the above mentioned MLP methods. These approximation schemes have been developed for certain high-dimensional semilinear PDEs and have been shown to overcome the curse of dimensionality in the sense of polynomial tractability (see, e.g., [31, 32, 51, 54]). Subsequent works extended the MLP methodology to more general classes of high-dimensional PDEs (see, e.g., [6, 7, 36, 48, 49, 52, 53, 79, 80]). Related multilevel Monte Carlo methods have also been developed for entropy-regularized MDPs (see, e.g., [71]).

The broader approximation strategy employed in this thesis builds on a line of research combining probabilistic numerical approximations with representation and approximation properties of ANNs to attain tractable approximations in high-dimensional situations. This approach was first developed for linear parabolic PDEs (see, e.g., [16, 40, 58]) and subsequently extended to nonlinear equations, including settings where MLP approximations provide the underlying numerical approximation scheme (see, e.g., [1, 2, 28, 50, 76, 77, 78])¹. Additionally, this strategy was applied, for example, in stochastic differential games (see, e.g., [86]).

The use of deep neural networks (DNNs) in reinforcement learning for the approximation of value functions, Q -functions, and optimal strategies has shown substantial empirical success, including high-dimensional tasks such as game playing, robotics, and control (see, e.g., [73], [98], [68], and [62]). Motivated by this development, several works investigated the effectiveness of deep reinforcement learning, i.e., the combination of DNNs and reinforcement learning methods, in high dimensions (see, e.g., [33], [25], and [105]). A comprehensive theoretical understanding of its performance for high-dimensional problems remains incomplete. In related joint work carried out during this PhD project, Grüne et al. [43] considered high-dimensional deterministic optimal control problems. More specifically, building on the compositional framework of Kang & Gong [60], we established under suitable assumptions that optimal controls can be approximated by ANNs without the curse of dimensionality.

Probabilistic methods also play an important role in the numerical approximation of solutions to elliptic boundary value problems. In particular, the Walk-on-Spheres algorithm, originally introduced by Müller [74] for the Dirichlet problem, provides a Monte Carlo method for approximating exit locations of Brownian motion from a bounded domain. Subsequently,

¹See also [8] for a survey on deep-learning based approximation methods for PDEs.

Walk-on-Spheres based numerical methods were extended to cover more general elliptic boundary value problems, including equations with drift coefficients (see, e.g., [21, 67, 91, 92, 93, 94]). More recently, Walk-on-Spheres based Monte Carlo approximations have been combined with ANN representation and approximation to obtain global approximation results for high-dimensional elliptic PDEs. In particular, the works [39] and [20] demonstrated that ANNs can approximate solutions of drift-free Poisson problems and that, under suitable assumptions, the required network complexity grows at most polynomially in the dimension of the domain and in the reciprocal of the prescribed approximation accuracy. The methodology developed in [20] constitutes a starting point for the Walk-on-Spheres based constructions considered in this thesis.

1.5. Contributions of the Dissertation

The contributions of this thesis revolve around the two classes of high-dimensional approximation problems presented in Section 1.2. The first concerns the approximation of solutions of functional fixed-point equations of type (1.3) and their applications to Bellman equations of MDPs, while the other concerns the approximation of functions of the form (1.5) related to solutions of a class of elliptic PDEs on bounded domains.

For the first problem class, the thesis develops nonlinear Monte Carlo methods, so-called *full-history recursive multilevel fixed-point* (MLFP) methods, to approximate functions which satisfy the functional fixed-point relation (1.3). To be more specific, let A be a nonempty finite set, let $\mathfrak{D}$ be a nonempty index set, consider a family of spaces $\mathbb{X}_d$, $d \in \mathfrak{D}$, for $d \in \mathfrak{D}$ let $f_d: \mathbb{X}_d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ denote the nonlinearity, for $d \in \mathfrak{D}$ let $X_d: \mathbb{X}_d \times \Omega \rightarrow \mathbb{X}_d^{|A|}$ denote a random field on a probability space $(\Omega, \mathcal{F}, \mathbb{P})$, and for $d \in \mathfrak{D}$ let $u_d: \mathbb{X}_d \rightarrow \mathbb{R}^{|A|}$ satisfy the functional fixed-point equation (1.3), i.e., $(u_d(x))(a) = \mathbb{E}[f_d(X_d^{x,a}, u_d(X_d^{x,a}))]$. Under suitable integrability and regularity assumptions, the introduced MLFP estimators $\mathcal{U}_{d,\varepsilon}: \mathbb{X}_d \times \Omega \rightarrow \mathbb{R}^{|A|}$, $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$, approximate the family $(u_d)_{d \in \mathfrak{D}}$ in the L^2 -sense that for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ it holds that

$$\sup_{x \in \mathbb{X}_d} \left(\mathbb{E} \left[\max_{a \in A} |(u_d(x))(a) - (\mathcal{U}_{d,\varepsilon}(x))(a)|^2 \right] \right)^{1/2} \leq \varepsilon. \quad (1.6)$$

The complexity of the MLFP approximations is bounded in the following sense. There exists a constant $c \in \mathbb{R}$ such that for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ it holds that $\text{cost}(d, \varepsilon) \leq c \mathfrak{R}_d \varepsilon^{-c}$, where for all $d \in \mathfrak{D}$, $x \in \mathbb{X}_d$ the number $\mathfrak{R}_d \in \mathbb{R}$ is understood as an upper bound for the computational cost to simulate one realization of the random variable $(X_d^{x,a})_{a \in A}$ in $\mathbb{X}_d^{|A|}$ and for every $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ the number $\text{cost}(d, \varepsilon) \in \mathbb{R}$ is understood as an upper bound of the computational cost required to generate realizations of all random variables $(X_d^{y,a})_{a \in A}$ for $y \in \mathbb{X}_d$ to compute one realization of the MLFP approximation $\mathcal{U}_{d,\varepsilon}(x)$ in $\mathbb{R}^{|A|}$. In particular, the computational complexity of the MLFP scheme grows at most polynomially in the reciprocal of the approximation accuracy and at most linearly in the computational cost associated to the generation of $\mathbb{X}_d^{|A|}$ -valued random variables. More specifically, in the setting $\mathfrak{D} = \mathbb{N}$ and $\mathbb{X}_d = \mathbb{R}^d$ it holds that under the assumption that the computational cost $\mathfrak{R}_d$ to simulate $(\mathbb{R}^d)^{|A|}$ -valued random variables grows at most polynomially in the dimension $d \in \mathbb{N}$ the complexity of the MLFP methods grows at most polynomially in the dimension d and the reciprocal ε^{-1} of the prescribed approximation accuracy. In this sense the MLFP approximations overcome the curse of dimensionality.

By the specific choice of the nonlinearity f_d to reconstruct the Q -function, the above results specialize to the approximation of the Q -function of associated discounted infinite time horizon MDPs with common finite action set A and state space $\mathbb{X}_d$. A further specialization to discrete time optimal stopping problems arises by considering the actions “stop” and “continue”. For the precise statements of these results, we refer to Chapter 2 below.

In the second stage of approximation for the first problem class, the thesis develops ANN approximation results based on the MLFP methods from stage one. Since the neural networks we consider in this thesis require a finite dimensional input, we specialize the setup above to the case $\mathfrak{D} = \mathbb{N}$ and $\mathbb{X}_d = \mathbb{R}^d$. Under suitable choices of neural network activation functions and the additional assumption that the nonlinearities $f_d: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ as well as the random fields $X_d: \mathbb{R}^d \times \Omega \rightarrow (\mathbb{R}^d)^{|A|}$ can be suitably approximated by ANNs, the solutions $u_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ of the functional fixed-point equations (1.3) can be approximated in the L^2 -sense by ANNs that grow at most polynomially in the dimension $d \in \mathbb{N}$ and the reciprocal ε^{-1} of the prescribed approximation accuracy. More precisely, there exists a constant $c \in \mathbb{R}$ such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ there exists a fully connected feedforward network $\Psi: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ (for the sake of simplicity, we identify here the neural network Ψ with its realization function) such that²

$$\left(\int_{\mathbb{R}^d} \|u_d(x) - \Psi(x)\|^2 \mu(dx)\right)^{1/2} \leq \varepsilon \quad (1.7)$$

and that the number of parameters describing the ANN Ψ is bounded by $cd^c \varepsilon^{-c}$. Similarly as for the MLFP method itself, the specialization of the nonlinearity f_d as before yields an analogous ANN approximation for the Q -functions of associated MDPs. For the precise formulation of the results, we refer to Chapter 4.

The Monte Carlo results of Chapter 2 were originally presented in the joint work [9] with C. Beck, A. Jentzen, and T. Kruse, while the ANN approximation results of Chapter 4 are based on the joint work [56] with A. Jentzen and T. Kruse.

For the second problem class, the thesis develops Monte Carlo methods and subsequently ANN approximations for functions of the form (1.5). First, we construct Monte Carlo approximators $\mathcal{W}: D \times \Omega \rightarrow \mathbb{R}$ based on the Walk-on-Spheres algorithm. Compared to the estimators in [20], the Monte Carlo methods established in this thesis incorporate additional sampled random times that appear in stochastic representations of solutions of the PDE (1.4) as a consequence of compensating for the drift and killing coefficients. We conduct an error analysis in the L^∞ -sense for the resulting Monte Carlo schemes as approximators for functions of the form (1.5). In particular, we demonstrate that under suitable assumptions the Monte Carlo approximator achieves $\mathbb{E}[\sup_{x \in D} |w(x) - \mathcal{W}(x)|] \leq \varepsilon$ for a prescribed approximation accuracy $\varepsilon \in (0, 1)$ while the complexity parameters of the estimator can be chosen such that they are bounded by $c\varepsilon^{-q}$. Here $q \in [0, \infty)$ is a constant that does not depend on any variables or parameters of the problem, and $c \in [0, \infty)$ is a constant that depends only on the dimension d of the underlying space, the diameter $\text{diam}(D)$ of the domain D , upper bounds and Lipschitz constants of the functions h_0 , h_1 , and constants quantifying suitable approximations of the distance function $\text{dist}(\cdot, \partial D)$ to the boundary of the domain D . Moreover, this dependency is at most polynomial. In particular, if the above problem quantities are bounded at most polynomially in the dimension $d \in \mathbb{N}$ of the domain D , the Monte Carlo scheme overcomes the curse of dimensionality. The above results reflect the contribution of Chapter 5 below.

In the second approximation stage for this problem class, we establish ANN approximation results based on the Walk-on-Spheres Monte Carlo methods developed above for functions of the form (1.5) in the special case $h_1 \equiv 0$, i.e., for functions $w: D \rightarrow \mathbb{R}$, $w(x) = \mathbb{E}[h_0(\tau^x, x + B_{\tau^x})]$. Under suitable assumptions concerning the ANN representation of h_0 and the ANN approximation of the distance function $\text{dist}(\cdot, \partial D)$ to the boundary of the domain D , we prove that w can be approximated in the L^∞ -sense by an ANN Ψ to a prescribed accuracy $\varepsilon \in (0, 1)$ such that the number of parameters of the approximating ANN Ψ is bounded by $\bar{c}\varepsilon^{-\bar{q}}$. Here $\bar{q} \in [0, \infty)$ is a constant that does not depend on any variables or parameters of the problem, and $\bar{c} \in [0, \infty)$ is a constant that depends only on the diameter of the domain, the dimension of the underlying space, the number of parameters, the Lipschitz constant, and

²For all $n \in \mathbb{N}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\|x\| = [\sum_{i=1}^n (x_i)^2]^{1/2}$ denote the standard norm.

an upper bound on $[0, \infty) \times \overline{D}$ of the realization of the ANN representing h_0 , and the number of parameters and a uniform Lipschitz constant of the ANNs that approximate the distance to the boundary. Moreover, this dependency is at most polynomial. In particular, under similar polynomial bounds on the problem quantities, the number of parameters of the approximating ANN grows at most polynomially in the dimension d and the reciprocal ε^{-1} of the accuracy ε . For the precise formulation of the ANN results for the second problem class, we refer to Chapter 6.

Through the stochastic representation described above, these approximation results yield corresponding approximation results for solutions of PDEs of the type (1.4). Both the Monte Carlo analysis of Chapter 5 and the ANN approximation results of Chapter 6 are based on the joint work [64] with T. Kruse.

1.6. Organization

The remainder of this dissertation is structured as follows. In Part I the numerical approximation for the first problem class, related to MDPs, is conducted. To this end, the MLFP approximations are developed and analyzed in Chapter 2. Chapter 3 presents the ANN framework and notation that are used subsequently to establish ANN representations and approximations. In Chapter 4 we employ the ANN calculus from Chapter 3 in combination with the Monte Carlo analysis from Chapter 2 in order to prove ANN approximation results for functional fixed-point equations of the form Equation (1.3).

Part II establishes the Walk-on-Spheres Monte Carlo and ANN approximations for the second problem class considered in this thesis. We develop in Chapter 5 the Walk-on-Spheres Monte Carlo approximation and the corresponding error and complexity analysis and combine these results in Chapter 6 with the ANN framework from Chapter 3 to obtain ANN approximation results for the second problem class.

In Chapter 7, we conclude the thesis with a brief summary and an outlook on possible research directions.

Part I.

Monte Carlo Methods and Deep Neural Network Approximations for High-Dimensional Markov Decision Processes

This part of the thesis concerns the numerical approximation of solutions to functional fixed-point equations of the form (1.3), motivated by problems in stochastic optimal control. In particular, the considered class includes the fixed-point equation (1.2) satisfied by the Q -function of a discounted, discrete time MDP with infinite time horizon and a finite action set as a special case. The value function of the associated MDP can in turn be recovered from the Q -function. Further specializations include discrete time optimal stopping problems. While the Monte Carlo approximation is formulated for more general state spaces, the approximation by means of ANNs is conducted on finite dimensional Euclidean spaces.

Methodologically, we first introduce the MLFP approximation scheme and establish quantitative error estimates and computational complexity bounds. This analysis is conducted on general state spaces. We subsequently specialize to Euclidean spaces and establish ANN approximation results on the basis of the previously constructed MLFP approximation schemes.

Chapter 2 and Chapter 4 are based on the joint works [9] with C. Beck, A. Jentzen, and T. Kruse and [56] with A. Jentzen and T. Kruse, respectively. The former establishes the Monte Carlo analysis, while the latter builds on these results to obtain ANN approximation results. The general ANN terminology and calculus utilized in Chapter 4 of Part I, and later in Chapter 6 of Part II, is collected in Chapter 3.

Chapter 2.

Nonlinear Monte Carlo Methods for Bellman Equations of High-Dimensional Markov Decision Processes

This chapter is based on the developments originally presented in the joint work [9] with C. Beck, A. Jentzen, and T. Kruse.

As discussed in the general introduction in Chapter 1, MDPs provide a mathematical framework for problems involving sequential decision making under uncertainty and as such constitute a foundation of many reinforcement learning methods (see, e.g., Bertsekas & Shreve [18], Powell [84], and Puterman [85]).

The stochastic dynamic programming principle (see, e.g., Bellman [14]) reduces the sequential decision making problem to the problem of determining an optimal action at the current state by considering both immediate reward and expected optimal rewards attainable in future states. This results in the characterization of the value function of an associated MDP as the solution of a nonlinear functional fixed-point equation, the Bellman equation (cf. (1.1)). This dynamic programming principle constitutes the basis of many algorithms in reinforcement learning such as temporal difference learning (see, e.g., Sutton [99]), Q -learning (see, e.g., Watkins [103] and Watkins & Dayan [104]), and the SARSA algorithm (see, e.g., Rummery & Niranjan [88]).

The approximative solution of Bellman equations in high dimensional state spaces is a notoriously difficult challenge due to the *curse of dimensionality* (cf., e.g., Bellman [14], Novak & Woźniakowski [82, Chapter 1], and Novak & Ritter [81]). For certain classes of discounted discrete time MDPs with continuous state spaces, the phenomenon is shown in lower bounds on the computational complexity of deterministic numerical approximation methods (cf., e.g., Chow & Tsitsiklis [26, 27]).

In this chapter, we consider randomized numerical approximation methods. More specifically, we introduce *nonlinear Monte Carlo methods* for discounted discrete time MDPs with infinite time horizon and finite control set that are polynomially tractable in the sense that the computational effort of the algorithm to approximatively compute the solution of the Bellman equation grows at most polynomially in the reciprocal ε^{-1} of the prescribed approximation accuracy $\varepsilon \in (0, 1]$ and the dimension $d \in \mathbb{N}$ of the underlying state space. In particular, the proposed methods do not suffer from the curse of dimensionality in the numerical approximation of solutions of Bellman equations.

More formally, we consider MDPs that are specified by a measurable space $(\mathbb{X}, \mathcal{X})$ (typically $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ with $d \in \mathbb{N}$ but our framework also covers discrete state spaces), a finite control set A , a discount factor $\delta \in (0, 1)$, transition kernels $\kappa_a: \mathbb{X} \times \mathcal{X} \rightarrow [0, 1]$, $a \in A$, and a measurable and bounded reward function $g: \mathbb{X} \times A \rightarrow \mathbb{R}$. The set of strategies $\mathcal{A}$ consists of all measurable functions $\alpha: \mathbb{X} \rightarrow A$ (under appropriate assumptions the restriction to such non-randomized, stationary strategies is without loss of generality, see, e.g., [18, Chapter 9]). Every $\alpha \in \mathcal{A}$ defines a Markov process $X^\alpha = (X_k^\alpha)_{k \in \mathbb{N}_0}$ on a filtered probability space $(\Omega, \mathcal{F}, (\mathbb{F}_k)_{k \in \mathbb{N}_0}, \mathbb{P})$ with state space $(\mathbb{X}, \mathcal{X})$ with the property that for all $k \in \mathbb{N}_0$, $B \in \mathcal{X}$ it holds $\mathbb{P}$ -a.s. that $\mathbb{P}[X_{k+1}^\alpha \in B | \mathbb{F}_k] = \kappa_{\alpha(X_k^\alpha)}(X_k^\alpha, B)$. The expected gain of a strategy $\alpha \in \mathcal{A}$ when starting in

$x \in \mathbb{X}$ is

$$J(x, \alpha) = \mathbb{E} \left[\sum_{k=0}^{\infty} \delta^k g(X_k^\alpha, \alpha(X_k^\alpha)) \middle| X_0^\alpha = x \right]. \quad (2.1)$$

The value function $V: \mathbb{X} \rightarrow \mathbb{R}$ is given by $V(x) = \sup_{\alpha \in \mathcal{A}} J(x, \alpha)$, $x \in \mathbb{X}$. The dynamic programming principle ensures that under appropriate conditions V satisfies for all $x \in \mathbb{X}$ that

$$V(x) = \max_{a \in A} (g(x, a) + \delta \mathbb{E}[V(X^{x,a})]), \quad (2.2)$$

where it holds for all $x \in \mathbb{X}$, $a \in A$ that the random variable $X^{x,a}: \Omega \rightarrow \mathbb{X}$ has distribution $\kappa_a(x, \cdot)$ (see, e.g., [18, Chapter 9] for a proof). An important subclass of these MDPs is given by optimal stopping problems in discrete time with infinite time horizon. For such problems the control set A consists of the two elements “stop” and “continue” and once the control “stop” is taken the Markov process jumps to a hold state from which it cannot leave and generates no further gains.

There is a large number of numerical approximation methods for Bellman equations of the form (2.2) that have been proposed and analyzed in the scientific literature. Deterministic numerical approximation methods for Bellman equations suffer in general from the curse of dimensionality. Indeed, Chow & Tsitsiklis [26, 27] show for all $d \in \mathbb{N}$ that in the case where the state space $\mathbb{X}$ is given by the d -dimensional unit cube $[0, 1]^d$ deterministic numerical approximation methods need at least $O(\varepsilon^{-2d})$ computational operations to approximate the solution of the Bellman equation with precision $\varepsilon \in (0, 1]$. Rust [90] shows that under certain assumptions it is possible to overcome this curse of dimensionality by allowing for randomized algorithms. Rust’s method consists of randomly sampling grid points in the state space and performing value iteration on this stochastic grid. The method is model-based as it requires explicit knowledge of the transition density. In the companion paper [89] Rust points out that convergence of his method might break down in situations where the transition density has spikes. Also Kristensen et al. [66] report a dramatic rise in the method’s variance as the state dimension increases. This is theoretically confirmed by Bray [22] who proves that Rust’s method only overcomes the curse of dimensionality in the special case where the MDP is equivalent to an MDP where all but a vanishingly small fraction of state variables behave like history-independent uniform random variables. Variants of Rust’s method for optimal stopping problems are designed by Broadie & Glasserman in [23] and [24]. The literature on optimal stopping comprises a variety of further randomized algorithms. We refer, for example, to [69], [101], and [102] for regression-based algorithms for optimal stopping problems, we refer, for example, to [87], [4], [44], [30], and [15] for duality-based algorithms for optimal stopping problems, and we refer, for example to [12], [13], and [38] for deep learning-based algorithms for optimal stopping problems.

There is a substantial body of research on sampling methods for MDPs in the literature of reinforcement learning. Most closely related to the present chapter are the works on reinforcement learning with a generative model (i.e., a simulator) that is capable of producing independent realizations of one-step transitions $X^{x,a}$ for arbitrary actions $a \in A$ and states $x \in \mathbb{X}$ (see, e.g., [63], [59], [5], [96], [3], [97]). However, much of this literature focuses on MDPs with finite state spaces, whereas in this chapter we aim at designing methods for more general state spaces, including continuous and possibly high dimensional spaces such as $\mathbb{R}^d$, $d \in \mathbb{N}$. In this context, we also refer to the recent work [71] which develops a class of multilevel Monte Carlo approximations for entropy-regularized MDPs that achieve polynomial sample complexity with complexity bounds independent of the dimension or cardinality of the state and action spaces.

More specifically, in this chapter we introduce nonlinear Monte Carlo algorithms that overcome the curse of dimensionality in the approximation of Bellman equations, that require only very weak regularity assumptions and that are model-free in the sense that they do not

use explicit knowledge of the transition kernel κ but only need access to a generative model. Our approach combines ideas from *Q-learning* and the MLP approximations which have been proven to overcome the curse of dimensionality in the numerical approximation of certain semilinear PDEs (see, e.g., [6, 7, 10, 11, 31, 32, 36, 48, 49, 51, 53, 54]). *Q-learning* is based on the idea to switch the order of expectation and maximization in (2.2). Formally, the *Q*-function satisfies for all $x \in \mathbb{X}$, $a \in A$ that $Q(x, a) = g(x, a) + \delta \mathbb{E}[V(X^{x,a})]$. This together with (2.2) implies for all $x \in \mathbb{X}$ that $V(x) = \max_{a \in A} Q(x, a)$ and, hence, that for all $x \in \mathbb{X}$, $a \in A$ it holds that

$$Q(x, a) = g(x, a) + \delta \mathbb{E} \left[\max_{b \in A} Q(X^{x,a}, b) \right]. \quad (2.3)$$

Under appropriate conditions one can show that Q is the unique solution of this fixed-point equation and that the sequence of fixed-point iterates $Q_n: \mathbb{X} \times A \rightarrow \mathbb{R}$, $n \in \mathbb{N}_0$, which satisfies for all $n \in \mathbb{N}$, $x \in \mathbb{X}$, $a \in A$ that $Q_0(x, a) = g(x, a)$ and

$$Q_n(x, a) = g(x, a) + \delta \mathbb{E} \left[\max_{b \in A} Q_{n-1}(X^{x,a}, b) \right] \quad (2.4)$$

converges to Q . We next employ a central idea of MLP approximations and decompose the iterates into multilevels to obtain for all $n \in \mathbb{N}_0$, $x \in \mathbb{X}$, $a \in A$ that

$$Q_n(x, a) = g(x, a) + \delta \left[\sum_{l=0}^{n-1} \mathbb{E} \left[\left(\max_{b \in A} Q_l(X^{x,a}, b) \right) - \mathbf{1}_{\mathbb{N}}(l) \left(\max_{b \in A} Q_{l-1}(X^{x,a}, b) \right) \right] \right]. \quad (2.5)$$

In this telescope expansion, we apply a fundamental idea of Heinrich [45, 46] and Giles [37] and approximate the expected values by Monte Carlo averages with different degrees of accuracy at different levels $l \in \{1, \dots, n\}$. The convergence of $(Q_l)_{l \in \mathbb{N}}$ ensures that for large $l \in \{1, \dots, n\}$ the difference between Q_l and Q_{l-1} is small and hence we use for large $l \in \{1, \dots, n\}$ less Monte Carlo samples to approximate the expected value $\mathbb{E}[(\max_{b \in A} Q_l(X^{x,a}, b)) - (\max_{b \in A} Q_{l-1}(X^{x,a}, b))]$ than for small $l \in \{1, \dots, n\}$. More specifically, we fix $M \in \mathbb{N}$ and use M^{n-l} independent Monte Carlo samples to approximate the expected value at level $l \in \{1, \dots, n\}$ which leads to the full-history recursive multilevel fixed-point (MLFP) approximation scheme in (2.7) below.

While the preceding motivation and methodology is formulated for the *Q*-function of an MDP, the error and complexity analysis developed in this chapter applies to a more general class of functional fixed-point equations.

We briefly sketch the contribution of this chapter within this introductory section by presenting the following result, Theorem 2.0.1 below, a special case of Theorem 2.3.1, the main result of this chapter. Below Theorem 2.0.1 we explain in words the statement of Theorem 2.0.1 as well as the mathematical objects appearing in Theorem 2.0.1.

Theorem 2.0.1. Let A be a nonempty set, for every $d \in \mathbb{N}$ let $g_d: \mathbb{R}^d \times A \rightarrow \mathbb{R}$ be $(\mathcal{B}(\mathbb{R}^d) \otimes 2^A)/\mathcal{B}(\mathbb{R})$ -measurable, assume

$$\sup_{d \in \mathbb{N}} \sup_{x \in \mathbb{R}^d} \sup_{a \in A} |g_d(x, a)| < \infty, \quad (2.6)$$

let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$, for¹ every $d \in \mathbb{N}$ let $\mathbb{F}_d^\theta \subseteq \mathcal{F}$, $\theta \in \Theta$, be independent sub-sigma-algebras of $\mathcal{F}$, for every $d \in \mathbb{N}$ let $X_d^\theta = (X_d^{\theta, x, a}(\omega))_{(x, a, \omega) \in \mathbb{R}^d \times A \times \Omega}: \mathbb{R}^d \times A \times \Omega \rightarrow \mathbb{R}^d$, $\theta \in \Theta$, be i.i.d. random fields which satisfy for all $d \in \mathbb{N}$, $\theta \in \Theta$ that X_d^θ is $(\mathcal{B}(\mathbb{R}^d) \otimes 2^A \otimes \mathbb{F}_d^\theta)/\mathcal{B}(\mathbb{R}^d)$ -measurable, for every $d \in \mathbb{N}$ let $\delta_d \in [0, 1)$, $\mathfrak{R}_d \in [0, \infty)$, let $M \in \mathbb{N} \cap [\sup_{d \in \mathbb{N}} (4|A|^2(1 - \delta_d)^{-2}), \infty]$, for every $d \in \mathbb{N}$ let $\mathcal{Q}_{d,n}^\theta: \mathbb{R}^d \times A \times \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, satisfy for all $n \in \mathbb{N}_0$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned} \mathcal{Q}_{d,n}^\theta(x, a) &= g_d(x, a) \\ &+ \sum_{l=0}^{n-1} \frac{\delta_d}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \left[\left(\max_{b \in A} \mathcal{Q}_{d,l}^{(\theta, l, i)}(X_d^{(\theta, l, i), x, a}, b) \right) - \mathbb{1}_{\mathbb{N}}(l) \left(\max_{b \in A} \mathcal{Q}_{d, \max\{l-1, 0\}}^{(\theta, -l, i)}(X_d^{(\theta, l, i), x, a}, b) \right) \right], \end{aligned} \quad (2.7)$$

and for every $d \in \mathbb{N}$ let $\mathfrak{C}_{d,n} \in [0, \infty)$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$ that

$$\mathfrak{C}_{d,n} \leq \sum_{l=0}^{n-1} M^{n-l} (\mathfrak{R}_d + |A| \mathfrak{C}_{d,l} + |A| \mathfrak{C}_{d, \max\{l-1, 0\}} \mathbb{1}_{\mathbb{N}}(l)). \quad (2.8)$$

Then

- (i) it holds for all $d \in \mathbb{N}$ that there exists a unique bounded $(\mathcal{B}(\mathbb{R}^d) \otimes 2^A)/\mathcal{B}(\mathbb{R})$ -measurable $Q_d: \mathbb{R}^d \times A \rightarrow \mathbb{R}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that

$$Q_d(x, a) = g_d(x, a) + \delta_d \mathbb{E} \left[\max_{b \in A} Q_d(X_d^{0, x, a}, b) \right] \quad (2.9)$$

and

- (ii) there exist $\mathfrak{n}: (0, 1] \rightarrow \mathbb{N}$ and $c \in \mathbb{R}$ such that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\mathfrak{C}_{d, \mathfrak{n}(\varepsilon)} \leq c \mathfrak{R}_d \varepsilon^{-c}$ and

$$\sup_{x \in \mathbb{R}^d} \left(\mathbb{E} \left[\max_{a \in A} |Q_d(x, a) - \mathcal{Q}_{d, \mathfrak{n}(\varepsilon)}^0(x, a)|^2 \right] \right)^{1/2} \leq \varepsilon. \quad (2.10)$$

Theorem 2.0.1 is an immediate consequence from Corollary 2.3.2 in Section 2.3 below. Corollary 2.3.2, in turn, follows from Theorem 2.3.1, which is the main result of this chapter. In the following we add some comments on the mathematical objects appearing in Theorem 2.0.1 above.

In Theorem 2.0.1 we introduce in (2.7) a Monte Carlo-type approximation algorithm for a sequence of MDPs indexed by the dimension $d \in \mathbb{N}$ of the state space. To formulate the proposed Monte Carlo-type approximation algorithm in (2.7) we need, roughly speaking, sufficiently many independent random variables which are indexed over a sufficiently large index set. This sufficiently large index set is provided through the set $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$ introduced in Theorem 2.0.1. The triple $(\Omega, \mathcal{F}, \mathbb{P})$ in Theorem 2.0.1 is the probability space on which the random variables are defined. In Theorem 2.0.1 we consider for every $d \in \mathbb{N}$ an MDP

¹Note that for every finite set B it holds that the number of elements of the power set 2^B of B satisfies $|2^B| = 2^{|B|}$. Moreover, note that for all finite sets B and C it holds that the number of elements of the set of all functions B^C from C to B satisfies $|B^C| = |B|^{|C|}$.

with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We assume that all elements of the sequence of MDPs have a common control set A . For every $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ the one-step transition of the controlled Markov chain of the MDP with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ is given by the random variable $X_d^{0,x,a}: \Omega \rightarrow \mathbb{R}^d$. In the language of MDPs for every $d \in \mathbb{N}$, $a \in A$ the transition kernel

$$\kappa_{d,a}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1] \quad (2.11)$$

is thus determined by the distribution of $X_d^{0,\cdot,a}$. For every $d \in \mathbb{N}$ the function $g_d: \mathbb{R}^d \times A \rightarrow \mathbb{R}$ introduced in the first line of Theorem 2.0.1 is the reward function of the MDP with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. In Theorem 2.0.1 we assume that the functions g_d , $d \in \mathbb{N}$, are uniformly bounded in $d \in \mathbb{N}$. For every $d \in \mathbb{N}$ the real number $\delta_d \in [0, 1)$ introduced in Theorem 2.0.1 is the discount factor of the MDP with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$.

Item (i) in Theorem 2.0.1 establishes the essentially well-known result that under the above assumptions for every $d \in \mathbb{N}$ the Bellman equation (2.2) for the Q -function associated to the MDP with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ has a unique solution $Q_d: \mathbb{R}^d \times A \rightarrow \mathbb{R}$.

In (2.7) in Theorem 2.0.1 we specify the MLFP approximation scheme which we propose to approximate the solution of the Bellman equation (2.2). The MLFP approximations $Q_{d,n}^\theta$, $d \in \mathbb{N}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, are indexed by the dimension $d \in \mathbb{N}$, by the number $n \in \mathbb{N}_0$ of fixed-point iterates, and by a parameter $\theta \in \Theta$ which is different for different appearances of MLFP approximations in (2.7). As random input sources the MLFP approximation scheme proposed in (2.7) employs the random variables

$$X_d^{\theta,x,a}: \Omega \rightarrow \mathbb{R}^d \quad (2.12)$$

for $\theta \in \Theta \setminus \{0\}$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$. Note that for every $\theta \in \Theta \setminus \{0\}$, $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ we have that the random variable $X_d^{\theta,x,a}: \Omega \rightarrow \mathbb{R}^d$ which is used as random input source of the MLFP approximation scheme proposed in (2.7) and the random variable $X_d^{0,x,a}: \Omega \rightarrow \mathbb{R}^d$ which is used to formulate the Bellman equation (2.2) are identically distributed. The parameter $\theta \in \Theta$ ensures that different appearances of MLFP approximations in (2.7) are independent and this ensures that (2.7) can be implemented with recursive function calls. The natural number $M \in \mathbb{N} \cap [\sup_{d \in \mathbb{N}} (4|A|^2(1 - \delta_d)^{-2}), \infty]$ in Theorem 2.0.1 determines the number of Monte Carlo samples used in the definition of the MLFP approximation $Q_{d,n}^\theta$, $d \in \mathbb{N}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, in (2.7). The assumption that the natural number M is an element of the set $\mathbb{N} \cap [\sup_{d \in \mathbb{N}} (4|A|^2(1 - \delta_d)^{-2}), \infty]$ ensures that

$$\sup_{d \in \mathbb{N}} (4|A|^2(1 - \delta_d)^{-2}) < \infty. \quad (2.13)$$

This implies that the control set A is finite and that the sequence $(\delta_d)_{d \in \mathbb{N}} \subseteq [0, 1)$ is bounded away from 1 in the sense that $\sup_{d \in \mathbb{N}} \delta_d < 1$. Note that Theorem 2.0.1 also covers the case where for each dimension $d \in \mathbb{N}$ there is an individual control set $\mathcal{A}_d$ with the property that $\sup_{d \in \mathbb{N}} |\mathcal{A}_d| < \infty$ by considering $A = \cup_{d \in \mathbb{N}} \mathcal{A}_d$.

For every $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\theta \in \Theta$ the nonnegative real number $\mathfrak{R}_d \in [0, \infty)$ in Theorem 2.0.1 is understood as an upper bound for the computational cost to compute one realization of the random variable $(X_d^{\theta,x,a})_{a \in A}$. The real numbers $\mathfrak{C}_{d,n} \in [0, \infty)$, $d \in \mathbb{N}$, $n \in \mathbb{N}_0$, in (2.8) in Theorem 2.0.1 model the computational costs of the MLFP approximation scheme in (2.7) (i.e., the sample complexity). More specifically, for every $d \in \mathbb{N}$, $n \in \mathbb{N}_0$ we have that the real number $\mathfrak{C}_{d,n} \in [0, \infty)$ represents an upper bound for the number of realizations of the random variables $(X_d^{\theta,x,a})_{a \in A}$ for $(x, \theta) \in \mathbb{R}^d \times \Theta$ required to compute one realization of the random variable $(Q_{d,n}^0(0, a))_{a \in A}$.

Item (ii) in Theorem 2.0.1 proves that the solutions of the Bellman equations in (2.2) can be approximated in the L^2 -sense by means of the MLFP approximation scheme in (2.7) with a computational cost which grows at most polynomially in the reciprocal ε^{-1} of the prescribed

approximation accuracy $\varepsilon \in (0, 1]$ and linearly in the computational cost $\mathfrak{R}_d$ to compute one realization of any of the random variables $X_d^{\theta, x}: \Omega \rightarrow (\mathbb{R}^d)^A$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, where $d \in \mathbb{N}$ is the dimension of the state space of the associated MDP. In particular, if the computational cost to compute one realization of any of the random variables $X_d^{\theta, x}: \Omega \rightarrow (\mathbb{R}^d)^A$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, grows at most polynomially in the dimension $d \in \mathbb{N}$ of the state space of the associated MDP (as it is often the case in practical applications), then the MLFP approximation scheme in (2.7) overcomes the curse of dimensionality for the approximation of the solutions of the Bellman equations in (2.2). However, we would like to point out that the constant $c \in \mathbb{R}$ appearing in item (ii) in Theorem 2.0.1 may become arbitrary large if $\sup_{d \in \mathbb{N}} \delta_d$ is close to 1. Thus the computational cost of the nonlinear Monte Carlo methods in (2.7) may become impractical even so the methods in (2.7) provably overcome the curse of dimensionality.

In the following we also add some comments on generalizations and variants of Theorem 2.0.1 presented in this chapter. While Theorem 2.0.1 considers a sequence of MDPs indexed by the dimension $d \in \mathbb{N}$ of the Euclidean state spaces $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, Corollary 2.3.2 considers a family of MDPs with general index set $\mathfrak{D}$ and general measurable state spaces $(\mathbb{X}_d, \mathcal{X}_d)$, $d \in \mathfrak{D}$. While Theorem 2.0.1 considers MDPs with bounded reward functions, Corollary 2.3.2 allows for unbounded reward functions. While Theorem 2.0.1 and Corollary 2.3.2 consider Bellman equations of MDPs, Theorem 2.3.1 considers more general functional fixed-point equations (we refer to (2.91) in Theorem 2.3.1 for details). Corollary 2.3.3 proves that a variant of the MLFP approximation scheme (see (2.119) for details) overcomes the curse of dimensionality for the approximation of the solutions of Bellman equations for optimal stopping problems (see (2.121) for details).

The structure of this chapter is as follows. In Section 2.1 below we establish existence, uniqueness, and integrability properties for solutions of functional fixed-point equations. In Section 2.2 below we introduce MLFP approximations for solutions of functional fixed-point equations, we study measurability, distributional, and integrability properties for the introduced MLFP approximations and we establish recursive and subsequently non-recursive upper bounds for the L^2 -distances between the exact solutions of the considered functional fixed-point equations and the proposed MLFP approximations. In Section 2.3 we combine the existence, uniqueness, and regularity properties for solutions of functional fixed-point equations, which we have established in Section 2.1, with the error analysis for MLFP approximations for functional fixed-point equations, which we have established in Section 2.2, to obtain a computational complexity analysis for MLFP approximations for functional fixed-point equations and for Bellman equations of MDPs and optimal stopping problems.

2.1. Existence and Uniqueness of Solutions of Functional Fixed-Point Equations

In this section we provide in Lemma 2.1.2 an elementary existence and uniqueness result for functional fixed-point equations formulated using stochastic kernels. In Corollary 2.1.3 we specialize Lemma 2.1.2 for functional fixed-point equations formulated using random variables. Lemma 2.1.4 establishes a priori estimates for solutions of functional fixed-point equations. In the special case where the functional fixed-point equation takes the form of a Bellman equation associated to an optimal stochastic control problem and the weight function $\mathfrak{w}$ is constant equal to one the results of this section are essentially well-known (see, e.g., [85, Theorem 6.2.5]).

Definition 2.1.1. Let $(\mathbb{X}, \mathcal{X})$ and $(\mathbb{Y}, \mathcal{Y})$ be nonempty measurable spaces and let $\kappa: \mathbb{X} \times \mathcal{Y} \rightarrow [0, 1]$ satisfy for all $M \in \mathcal{Y}$ that $\mathbb{X} \ni x \mapsto \kappa(x, M) \in [0, 1]$ is $\mathcal{X}/\mathcal{B}([0, 1])$ -measurable and for all $x \in \mathbb{X}$ that $\mathcal{Y} \ni M \mapsto \kappa(x, M) \in [0, 1]$ is a probability measure on $(\mathbb{Y}, \mathcal{Y})$. Then we say that κ is a stochastic kernel from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{Y}, \mathcal{Y})$.

Lemma 2.1.2. Let $c, L \in [0, \infty)$ with $cL < 1$, let $(\mathbb{X}, \mathcal{X})$ be a nonempty measurable space, let A be a nonempty at most countable set, for every $a \in A$ let $\kappa_a: \mathbb{X} \times \mathcal{X} \rightarrow [0, 1]$ be a stochastic kernel from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{X}, \mathcal{X})$, let $\tilde{\mathcal{A}} = \bigotimes_{a \in A} \mathcal{B}(\mathbb{R})$, let $f: \mathbb{X} \times \mathbb{R}^A \rightarrow \mathbb{R}$ be $(\mathcal{X} \otimes \tilde{\mathcal{A}})/\mathcal{B}(\mathbb{R})$ -measurable, let $\mathfrak{w}: \mathbb{X} \rightarrow (0, \infty)^A$ be $\mathcal{X}/\bigotimes_{a \in A} \mathcal{B}((0, \infty))$ -measurable, let

$$\mathfrak{W} = \left\{ (u: \mathbb{X} \rightarrow \mathbb{R}^A) : u \text{ is } \mathcal{X}/\tilde{\mathcal{A}}\text{-measurable, } \sup_{(x,a) \in \mathbb{X} \times A} [|(\mathfrak{w}(x))(a)|^{-1} |(u(x))(a)|] < \infty \right\}, \quad (2.14)$$

and assume for all $x \in \mathbb{X}$, $a \in A$, $r, s \in \mathbb{R}^A$ that $|f(x, r) - f(x, s)| \leq L \sup_{b \in A} |r(b) - s(b)|$, $\int_{\mathbb{X}} \sup_{b \in A} (\mathfrak{w}(y))(b) \kappa_a(x, dy) \leq c(\mathfrak{w}(x))(a)$, and $\sup_{(t,b) \in \mathbb{X} \times A} [|(\mathfrak{w}(t))(b)|^{-1} \int_{\mathbb{X}} |f(y, 0)| \kappa_b(t, dy)] < \infty$. Then there exists a unique $v \in \mathfrak{W}$ which satisfies for all $x \in \mathbb{X}$, $a \in A$ that

$$\int_{\mathbb{X}} |f(y, v(y))| \kappa_a(x, dy) < \infty \quad \text{and} \quad (v(x))(a) = \int_{\mathbb{X}} f(y, v(y)) \kappa_a(x, dy). \quad (2.15)$$

Proof of Lemma 2.1.2. Throughout this proof let $\|\cdot\|: \mathfrak{W} \rightarrow [0, \infty)$ satisfy for all $u \in \mathfrak{W}$ that $\|u\| = \sup_{(x,a) \in \mathbb{X} \times A} \frac{|(u(x))(a)|}{|(\mathfrak{w}(x))(a)|}$. Note that $\|\cdot\|$ is a norm on $\mathfrak{W}$ and $(\mathfrak{W}, \|\cdot\|)$ is a Banach space. The assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \sup_{a \in A} |r(a) - s(a)|$ and the assumption that for all $x \in \mathbb{X}$, $a \in A$ it holds that $\int_{\mathbb{X}} \sup_{b \in A} (\mathfrak{w}(y))(b) \kappa_a(x, dy) \leq c(\mathfrak{w}(x))(a)$, yield that for all $u \in \mathfrak{W}$, $x \in \mathbb{X}$, $a \in A$ it holds that

$$\begin{aligned} & \frac{1}{|(\mathfrak{w}(x))(a)|} \left| \int_{\mathbb{X}} f(y, u(y)) \kappa_a(x, dy) \right| \\ & \leq \frac{1}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} |f(y, u(y)) - f(y, 0)| \kappa_a(x, dy) + \frac{1}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy) \\ & \leq \frac{L}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} \sup_{b \in A} |(u(y))(b)| \kappa_a(x, dy) + \frac{1}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy) \\ & = \frac{L}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} \sup_{b \in A} \left\{ \frac{|(u(y))(b)|}{|(\mathfrak{w}(y))(b)|} |(\mathfrak{w}(y))(b)| \right\} \kappa_a(x, dy) + \frac{1}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy) \\ & \leq \frac{L\|u\|}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} \sup_{b \in A} (\mathfrak{w}(y))(b) \kappa_a(x, dy) + \frac{1}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy) \\ & \leq cL\|u\| + \frac{1}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy). \end{aligned} \quad (2.16)$$

The assumption that $\sup_{(x,a) \in \mathbb{X} \times A} [|(\mathfrak{w}(x))(a)|^{-1} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy)] < \infty$ demonstrates that for all $u \in \mathfrak{W}$ it holds that

$$\begin{aligned} & \sup_{(x,a) \in \mathbb{X} \times A} \frac{1}{|(\mathfrak{w}(x))(a)|} \left| \int_{\mathbb{X}} f(y, u(y)) \kappa_a(x, dy) \right| \\ & \leq cL\|u\| + \sup_{(x,a) \in \mathbb{X} \times A} \frac{1}{|(\mathfrak{w}(x))(a)|} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy) < \infty. \end{aligned} \quad (2.17)$$

Next, observe that [65, Lemma 14.20] ensures that for all $u \in \mathfrak{W}$ it holds that the map $\mathbb{X} \ni x \mapsto [A \ni a \mapsto \int_{\mathbb{X}} f(y, u(y)) \kappa_a(x, dy) \in \mathbb{R}] \in \mathbb{R}^A$ is $\mathcal{X}/\tilde{\mathcal{A}}$ -measurable. Let $\Phi: \mathfrak{W} \rightarrow \mathfrak{W}$ be the function which satisfies for all $u \in \mathfrak{W}$, $(x, a) \in \mathbb{X} \times A$ that

$$[\Phi(u)](x)(a) = \int_{\mathbb{X}} f(y, u(y)) \kappa_a(x, dy). \quad (2.18)$$

The assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \sup_{a \in A} |r(a) - s(a)|$, the assumption that $\sup_{(x,a) \in \mathbb{X} \times A} [|(\mathfrak{w}(x))(a)|^{-1} \int_{\mathbb{X}} |f(y, 0)| \kappa_a(x, dy)] < \infty$, and the

assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $\int_{\mathbb{X}} \sup_{b \in A} (\mathfrak{w}(y))(b) \kappa_a(x, dy) \leq c(\mathfrak{w}(x))(a)$, ensure that for all $u, v \in \mathfrak{W}$, $(x, a) \in \mathbb{X} \times A$ it holds that

$$\begin{aligned}
& \frac{1}{(\mathfrak{w}(x))(a)} \left| [\Phi(u)](x)(a) - [\Phi(v)](x)(a) \right| \\
& \leq \frac{1}{(\mathfrak{w}(x))(a)} \int_{\mathbb{X}} |f(y, u(y)) - f(y, v(y))| \kappa_a(x, dy) \\
& \leq \frac{L}{(\mathfrak{w}(x))(a)} \int_{\mathbb{X}} \sup_{b \in A} |(u(y))(b) - (v(y))(b)| \kappa_a(x, dy) \\
& = \frac{L}{(\mathfrak{w}(x))(a)} \int_{\mathbb{X}} \sup_{b \in A} \left\{ \frac{|(u(y))(b) - (v(y))(b)|}{(\mathfrak{w}(y))(b)} (\mathfrak{w}(y))(b) \right\} \kappa_a(x, dy) \\
& \leq \frac{L \|u - v\|}{(\mathfrak{w}(x))(a)} \int_{\mathbb{X}} \sup_{b \in A} (\mathfrak{w}(y))(b) \kappa_a(x, dy) \\
& \leq cL \|u - v\|.
\end{aligned} \tag{2.19}$$

This in turn proves that for all $u, v \in \mathfrak{W}$ it holds that

$$\|\Phi(u) - \Phi(v)\| \leq cL \|u - v\|. \tag{2.20}$$

The assumption that $cL < 1$ shows that Φ is a contraction. Hence Banach's fixed-point theorem proves that there exists a unique function $v \in \mathfrak{W}$ such that $v = \Phi(v)$. The fact that $v \in \mathfrak{W}$ ensures that for all $(x, a) \in \mathbb{X} \times A$ it holds that $\int_{\mathbb{X}} |f(y, v(y))| \kappa_a(x, dy) < \infty$. The proof of Lemma 2.1.2 is thus complete. $\square$

Corollary 2.1.3. *Let $c, L \in [0, \infty)$ with $cL < 1$, let $(\mathbb{X}, \mathcal{X})$ be a nonempty measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let A be a nonempty at most countable set, let $\tilde{\mathcal{A}} = \bigotimes_{a \in A} \mathcal{B}(\mathbb{R})$, let $\tilde{\mathcal{X}} = \bigotimes_{a \in A} \mathcal{X}$, let $X = (X^{x,a}(\omega))_{x \in \mathbb{X}, a \in A, \omega \in \Omega} : \mathbb{X} \times \Omega \rightarrow \mathbb{X}^A$ be $(\mathcal{X} \otimes \mathcal{F})/\tilde{\mathcal{X}}$ -measurable, let $f : \mathbb{X} \times \mathbb{R}^A \rightarrow \mathbb{R}$ be $(\mathcal{X} \otimes \tilde{\mathcal{A}})/\mathcal{B}(\mathbb{R})$ -measurable, let $\mathfrak{w} : \mathbb{X} \rightarrow (0, \infty)^A$ be $\mathcal{X}/\bigotimes_{a \in A} \mathcal{B}((0, \infty))$ -measurable, let*

$$\mathfrak{W} = \left\{ (u : \mathbb{X} \rightarrow \mathbb{R}^A) : u \text{ is } \mathcal{X}/\tilde{\mathcal{A}}\text{-measurable, } \sup_{(x,a) \in \mathbb{X} \times A} [|(\mathfrak{w}(x))(a)|^{-1} |(u(x))(a)|] < \infty \right\}, \tag{2.21}$$

assume that for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \sup_{b \in A} |r(b) - s(b)|$, $\mathbb{E}[\sup_{b \in A} (\mathfrak{w}(X^{x,a}))(b)] \leq c(\mathfrak{w}(x))(a)$, and $\sup_{(y,b) \in \mathbb{X} \times A} [|(\mathfrak{w}(y))(b)|^{-1} \mathbb{E}[|f(X^{y,b}, 0)|]] < \infty$. Then there exists a unique function $v \in \mathfrak{W}$ which satisfies for all $(x, a) \in \mathbb{X} \times A$ that

$$\mathbb{E}[|f(X^{x,a}, v(X^{x,a}))|] < \infty \quad \text{and} \quad (v(x))(a) = \mathbb{E}[f(X^{x,a}, v(X^{x,a}))]. \tag{2.22}$$

Proof of Corollary 2.1.3. Let $\kappa_a : \mathbb{X} \times \mathcal{X} \rightarrow [0, 1]$, $a \in A$, satisfy for all $(x, a) \in \mathbb{X} \times A$, $M \in \mathcal{X}$, that $\kappa_a(x, M) = \mathbb{P}[X^{x,a} \in M]$. The fact that X is $(\mathcal{X} \otimes \mathcal{F})/\tilde{\mathcal{X}}$ -measurable ensures that for all $(x, a) \in \mathbb{X} \times A$ it holds that $X^{x,a} : \Omega \rightarrow \mathbb{X}$ is $\mathcal{F}/\mathcal{X}$ -measurable. This implies that for all $(x, a) \in \mathbb{X} \times A$ it holds that $\mathcal{X} \ni M \mapsto \kappa_a(x, M) = \mathbb{P}[X^{x,a} \in M] \in [0, 1]$ is a probability measure on $(\mathbb{X}, \mathcal{X})$. The fact that X is $(\mathcal{X} \otimes \mathcal{F})/\tilde{\mathcal{X}}$ -measurable and [65, Theorem 14.16] imply that for all $a \in A$, $M \in \mathcal{X}$ it holds that $\mathbb{X} \ni x \mapsto \kappa_a(x, M) = \mathbb{P}[X^{x,a} \in M]$ is $\mathcal{X}/\mathcal{B}([0, 1])$ -measurable. Hence it holds for all $a \in A$ that κ_a is a stochastic kernel from $(\mathbb{X}, \mathcal{X})$ to $(\mathbb{X}, \mathcal{X})$. The assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $\mathbb{E}[\sup_{b \in A} (\mathfrak{w}(X^{x,a}))(b)] \leq c(\mathfrak{w}(x))(a)$ and the assumption that $\sup_{(y,b) \in \mathbb{X} \times A} [|(\mathfrak{w}(y))(b)|^{-1} \mathbb{E}[|f(X^{y,b}, 0)|]] < \infty$ prove that for all $(x, a) \in \mathbb{X} \times A$ it holds that

$$\int_{\mathbb{X}} \sup_{b \in A} (\mathfrak{w}(y))(b) \kappa_a(x, dy) = \int_{\mathbb{X}} \sup_{b \in A} (\mathfrak{w}(y))(b) (X^{x,a}(\mathbb{P}))(dy) = \mathbb{E} \left[\sup_{b \in A} (\mathfrak{w}(X^{x,a}))(b) \right] \leq c(\mathfrak{w}(x))(a), \tag{2.23}$$

and

$$\begin{aligned} \sup_{(y,b) \in \mathbb{X} \times A} \frac{1}{(\mathfrak{w}(y))(b)} \int_{\mathbb{X}} |f(t,0)| \kappa_b(y, dt) &= \sup_{(y,b) \in \mathbb{X} \times A} \frac{1}{(\mathfrak{w}(y))(b)} \int_{\mathbb{X}} |f(t,0)| (X^{y,b}(\mathbb{P}))(dt) \\ &= \sup_{(y,b) \in \mathbb{X} \times A} \frac{1}{(\mathfrak{w}(y))(b)} \mathbb{E}[|f(X^{y,b}, 0)|] < \infty. \end{aligned} \quad (2.24)$$

Lemma 2.1.2 implies that there exists a unique function $v \in \mathfrak{W}$ such that for all $(x, a) \in \mathbb{X} \times A$ it holds that

$$\mathbb{E}[|f(X^{x,a}, v(X^{x,a}))|] = \int_{\mathbb{X}} |f(y, v(y))| (X^{x,a}(\mathbb{P}))(dy) = \int_{\mathbb{X}} |f(y, v(y))| \kappa_a(x, dy) < \infty, \quad (2.25)$$

and

$$(v(x))(a) = \int_{\mathbb{X}} f(y, v(y)) \kappa_a(x, dy) = \int_{\mathbb{X}} f(y, v(y)) (X^{x,a}(\mathbb{P}))(dy) = \mathbb{E}[f(X^{x,a}, v(X^{x,a}))]. \quad (2.26)$$

The proof of Corollary 2.1.3 is thus complete. $\square$

Lemma 2.1.4. *Let $c_f, c_{\mathfrak{w}}, L \in [0, \infty)$ with $c_{\mathfrak{w}}L < 1$, let $(\mathbb{X}, \mathcal{X})$ be a nonempty measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let A be a nonempty at most countable set, let $\tilde{\mathcal{A}} = \bigotimes_{a \in A} \mathcal{B}(\mathbb{R})$, let $\tilde{\mathcal{X}} = \bigotimes_{a \in A} \mathcal{X}$, let $X = (X^{x,a}(\omega))_{x \in \mathbb{X}, a \in A, \omega \in \Omega} : \mathbb{X} \times \Omega \rightarrow \mathbb{X}^A$ be $(\mathcal{X} \otimes \mathcal{F})/\tilde{\mathcal{X}}$ -measurable, let $f : \mathbb{X} \times \mathbb{R}^A \rightarrow \mathbb{R}$ be $(\mathcal{X} \otimes \tilde{\mathcal{A}})/\mathcal{B}(\mathbb{R})$ -measurable, let $\mathfrak{w} : \mathbb{X} \rightarrow (0, \infty)^A$ be $\mathcal{X}/\bigotimes_{a \in A} \mathcal{B}((0, \infty))$ -measurable, let $v : \mathbb{X} \rightarrow \mathbb{R}^A$ be $\mathcal{X}/\tilde{\mathcal{A}}$ -measurable, assume that for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \sup_{b \in A} |r(b) - s(b)|$, $\sup_{(y,b) \in \mathbb{X} \times A} [|(\mathfrak{w}(y))(b)|^{-1} |(v(y))(b)|] < \infty$, $(\mathbb{E}[\sup_{b \in A} |(\mathfrak{w}(X^{x,a}))(b)|^2])^{\frac{1}{2}} \leq c_{\mathfrak{w}}(\mathfrak{w}(x))(a)$, $(\mathbb{E}[|f(X^{x,a}, 0)|^2])^{\frac{1}{2}} \leq c_f(\mathfrak{w}(x))(a)$, and $(v(x))(a) = \mathbb{E}[f(X^{x,a}, v(X^{x,a}))]$. Then*

$$\sup_{(x,a) \in \mathbb{X} \times A} \frac{|(v(x))(a)|}{(\mathfrak{w}(x))(a)} \leq \frac{c_f}{1 - c_{\mathfrak{w}}L} \quad \text{and} \quad \sup_{(x,a) \in \mathbb{X} \times A} \frac{(\mathbb{E}[\sup_{b \in A} |(v(X^{x,a}))(b)|^2])^{\frac{1}{2}}}{(\mathfrak{w}(x))(a)} \leq \frac{c_f c_{\mathfrak{w}}}{1 - c_{\mathfrak{w}}L}. \quad (2.27)$$

Proof of Lemma 2.1.4. Jensen's inequality, the triangle inequality, and the assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \sup_{a \in A} |r(a) - s(a)|$ prove that for all $(x, a) \in \mathbb{X} \times A$ it holds that

$$\begin{aligned} \frac{|(v(x))(a)|}{(\mathfrak{w}(x))(a)} &= \frac{|\mathbb{E}[f(X^{x,a}, v(X^{x,a}))]|}{(\mathfrak{w}(x))(a)} \leq \frac{\mathbb{E}[|f(X^{x,a}, v(X^{x,a}))|^2]^{\frac{1}{2}}}{(\mathfrak{w}(x))(a)} \\ &\leq L \frac{\mathbb{E}[\sup_{b \in A} |(v(X^{x,a}))(b)|^2]^{\frac{1}{2}}}{(\mathfrak{w}(x))(a)} + \frac{\mathbb{E}[|f(X^{x,a}, 0)|^2]^{\frac{1}{2}}}{(\mathfrak{w}(x))(a)} \\ &\leq \frac{L}{(\mathfrak{w}(x))(a)} \mathbb{E} \left[\sup_{b \in A} \left\{ \frac{|(v(X^{x,a}))(b)|}{(\mathfrak{w}(X^{x,a}))(b)} (\mathfrak{w}(X^{x,a}))(b) \right\}^2 \right]^{\frac{1}{2}} + c_f \\ &\leq \frac{L}{(\mathfrak{w}(x))(a)} \sup_{(y,b) \in \mathbb{X} \times A} \left\{ \frac{|(v(y))(b)|}{(\mathfrak{w}(y))(b)} \right\} \mathbb{E}[\sup_{b \in A} \mathfrak{w}(X^{x,a})(b)^2]^{\frac{1}{2}} + c_f \\ &\leq c_{\mathfrak{w}}L \sup_{(y,b) \in \mathbb{X} \times A} \left\{ \frac{|(v(y))(b)|}{(\mathfrak{w}(y))(b)} \right\} + c_f. \end{aligned} \quad (2.28)$$

Combining this, the assumption that $\sup_{(x,a) \in \mathbb{X} \times A} [|\mathfrak{w}(x)(a)|^{-1} |(v(x))(a)|] < \infty$, and the assumption that $c_{\mathfrak{w}}L < 1$ shows that

$$\sup_{(x,a) \in \mathbb{X} \times A} \frac{|(v(x))(a)|}{|\mathfrak{w}(x)(a)|} \leq \frac{c_f}{1 - c_{\mathfrak{w}}L}. \quad (2.29)$$

This and the assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $\mathbb{E}[|\sup_{b \in A} (\mathfrak{w}(X^{x,a}))(b)|^2]^{\frac{1}{2}} \leq c_{\mathfrak{w}}(\mathfrak{w}(x))(a)$ imply that for all $(x, a) \in \mathbb{X} \times A$ it holds that

$$\frac{(\mathbb{E}[\sup_{b \in A} |(v(X^{x,a}))(b)|^2])^{\frac{1}{2}}}{(\mathfrak{w}(x))(a)} \leq \left[\sup_{(y,b) \in \mathbb{X} \times A} \frac{|(v(y))(b)|}{|\mathfrak{w}(y))(b)|} \right] \frac{\mathbb{E}[\sup_{b \in A} (\mathfrak{w}(X^{x,a}))(b)^2]^{\frac{1}{2}}}{(\mathfrak{w}(x))(a)} \leq \frac{c_f c_{\mathfrak{w}}}{1 - c_{\mathfrak{w}}L}. \quad (2.30)$$

Taking the supremum over $\mathbb{X} \times A$ yields

$$\sup_{(x,a) \in \mathbb{X} \times A} \frac{(\mathbb{E}[\sup_{b \in A} |(v(X^{x,a}))(b)|^2])^{\frac{1}{2}}}{(\mathfrak{w}(x))(a)} \leq \frac{c_f c_{\mathfrak{w}}}{1 - c_{\mathfrak{w}}L}. \quad (2.31)$$

The proof of Lemma 2.1.4 is thus complete. $\square$

2.2. Full-History Recursive Multilevel Fixed-Point (MLFP) Approximations

In this section we introduce in Setting 2.2.1 in Subsection 2.2.1 below MLFP approximations for functional fixed-point equations, we establish in Lemma 2.2.2 in Subsection 2.2.2 below basic measurability properties for MLFP approximations, we establish in Lemma 2.2.3 in Subsection 2.2.2 below basic distributional properties for MLFP approximations, and we establish in Lemma 2.2.4 and Lemma 2.2.5 in Subsection 2.2.3 below basic integrability properties for MLFP approximations. In Lemma 2.2.7 in Subsection 2.2.4 below we provide a recursive variance estimate for MLFP approximations for functional fixed-point equations. Our proof of Lemma 2.2.7 in Subsection 2.2.4 below is based on the measurability, distributional, and integrability properties for MLFP approximations established in Subsection 2.2.2 and Subsection 2.2.3. In Proposition 2.2.8 in Subsection 2.2.5 below we provide a full error analysis for MLFP approximations for functional fixed-point equations. Our proof of Proposition 2.2.8 in Subsection 2.2.5 below is based on the idea to combine the recursive variance estimate in Lemma 2.2.7 in Subsection 2.2.4 below with bias-variance decompositions for the approximation errors of the proposed MLFP approximation schemes, and the discrete Gronwall-type two step recursion in [55, Lemma 2.1].

2.2.1. Mathematical Description of MLFP Approximations

Setting 2.2.1. Let $M \in \mathbb{N}$, let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, let $(\mathbb{X}, \mathcal{X})$ be a nonempty measurable space, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let A be a nonempty set, let $\tilde{\mathcal{A}} = \bigotimes_{a \in A} \mathcal{B}(\mathbb{R})$, let $\|\cdot\|_{\infty} : \mathbb{R}^A \rightarrow [0, \infty]$ satisfy for all $r \in \mathbb{R}^A$ that $\|r\|_{\infty} = \sup_{a \in A} |r(a)|$, let $\tilde{\mathcal{X}} = \bigotimes_{a \in A} \mathcal{X}$, let $f : \mathbb{X} \times \mathbb{R}^A \rightarrow \mathbb{R}$ be $(\mathcal{X} \otimes \tilde{\mathcal{A}})/\mathcal{B}(\mathbb{R})$ -measurable, let $\mathbb{F}^{\theta} \subseteq \mathcal{F}$, $\theta \in \Theta$, be independent sub-sigma-algebras of $\mathcal{F}$, let $X^{\theta} = (X^{\theta,x,a}(\omega))_{x \in \mathbb{X}, a \in A, \omega \in \Omega} : \mathbb{X} \times \Omega \rightarrow \mathbb{R}^A$, $\theta \in \Theta$, be i.i.d. random fields, assume for all $\theta \in \Theta$ that X^{θ} is $(\mathcal{X} \otimes \mathbb{F}^{\theta})/\tilde{\mathcal{X}}$ -measurable, let $V_n^{\theta} : \mathbb{X} \times \Omega \rightarrow \mathbb{R}^A$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, satisfy for all $n \in \mathbb{N}_0$, $\theta \in \Theta$, $x \in \mathbb{X}$, $a \in A$ that

$$(V_n^{\theta}(x))(a) = \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \left[f(X^{(\theta,l,i),x,a}, V_l^{(\theta,l,i)}(X^{(\theta,l,i),x,a})) - \mathbb{1}_{\mathbb{N}}(l) f(X^{(\theta,l,i),x,a}, V_{\max\{l-1,0\}}^{(\theta,-l,i)}(X^{(\theta,l,i),x,a})) \right]. \quad (2.32)$$

2.2.2. Measurability and Distributional Properties for MLFP Approximations

Lemma 2.2.2. *Assume Setting 2.2.1. Then*

- (i) *it holds for all $n \in \mathbb{N}_0$, $\theta \in \Theta$ that V_n^θ is $(\mathcal{X} \otimes \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})) / \tilde{\mathcal{A}}$ -measurable and*
- (ii) *it holds for all $n \in \mathbb{N}_0$, $\theta, \vartheta \in \Theta$ that*

$$\mathbb{X} \times \Omega \ni (x, \omega) \mapsto [A \ni a \mapsto f(X^{\theta, x, a}(\omega), V_n^\vartheta(X^{\theta, x, a}(\omega), \omega)) \in \mathbb{R}] \in \mathbb{R}^A \quad (2.33)$$

is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\vartheta, \eta)}) \cup \mathbb{F}^\theta)) / \tilde{\mathcal{A}}$ -measurable.

Proof of Lemma 2.2.2. Note that (2.32) implies that for all $\theta \in \Theta$, $x \in \mathbb{X}$, $a \in A$ it holds that $(V_0^\theta(x))(a) = 0$. Hence for all $\theta \in \Theta$ it holds that V_0^θ is $(\mathcal{X} \otimes \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})) / \tilde{\mathcal{A}}$ -measurable. Fix $n \in \mathbb{N}$ and assume for all $l \in \{0, 1, \dots, n-1\}$, $\theta \in \Theta$ that V_l^θ is $(\mathcal{X} \otimes \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})) / \tilde{\mathcal{A}}$ -measurable. Observe that for all $l \in \{0, 1, \dots, n-1\}$, $\theta \in \Theta$ it holds that

$$\mathbb{X} \times \Omega \ni (x, \omega) \mapsto (x, V_l^\theta(x, \omega)) \in \mathbb{X} \times \mathbb{R}^A \quad (2.34)$$

is $(\mathcal{X} \otimes \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})) / (\mathcal{X} \otimes \tilde{\mathcal{A}})$ -measurable. Since for all $\theta \in \Theta$ it holds that X^θ is $(\mathcal{X} \otimes \mathbb{F}^\theta) / \tilde{\mathcal{X}}$ -measurable it follows that for all $\theta \in \Theta$, $a \in A$ it holds that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto X^{\theta, x, a}(\omega) \in \mathbb{X}$ is $(\mathcal{X} \otimes \mathbb{F}^\theta) / \mathcal{X}$ -measurable. This implies that for all $\theta, \vartheta \in \Theta$, $a \in A$ it holds that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto (X^{\theta, x, a}(\omega), \omega) \in \mathbb{X} \times \Omega$ is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\vartheta, \eta)}) \cup \mathbb{F}^\theta)) / (\mathcal{X} \otimes \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\vartheta, \eta)}))$ -measurable. This and (2.34) ensure that for all $l \in \{0, 1, \dots, n-1\}$, $\theta, \vartheta \in \Theta$, $a \in A$ it holds that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto (X^{\theta, x, a}(\omega), V_l^\vartheta(X^{\theta, x, a}(\omega), \omega)) \in \mathbb{X} \times \mathbb{R}^A$ is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\vartheta, \eta)}) \cup \mathbb{F}^\theta)) / (\mathcal{X} \otimes \tilde{\mathcal{A}})$ -measurable. Therefore for all $l \in \{0, 1, \dots, n-1\}$, $\theta, \vartheta \in \Theta$, $a \in A$ it holds that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto f(X^{\theta, x, a}(\omega), V_l^\vartheta(X^{\theta, x, a}(\omega), \omega)) \in \mathbb{R}$ is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\vartheta, \eta)}) \cup \mathbb{F}^\theta)) / \mathcal{B}(\mathbb{R})$ -measurable. This implies that for all $l \in \{0, 1, \dots, n-1\}$, $\theta, \vartheta \in \Theta$ it holds that

$$\mathbb{X} \times \Omega \ni (x, \omega) \mapsto [A \ni a \mapsto f(X^{\theta, x, a}(\omega), V_l^\vartheta(X^{\theta, x, a}(\omega), \omega)) \in \mathbb{R}] \in \mathbb{R}^A \quad (2.35)$$

is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\vartheta, \eta)}) \cup \mathbb{F}^\theta)) / \tilde{\mathcal{A}}$ -measurable. Note that for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$ it holds that $\sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, l, i, \eta)}) \cup \mathbb{F}^{(\theta, l, i)}) \subseteq \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})$ and $\sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, -l, i, \eta)}) \cup \mathbb{F}^{(\theta, -l, i)}) \subseteq \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})$. This and (2.35) ensure that for all $\theta \in \Theta$ it holds that $V_n^\theta: \mathbb{X} \times \Omega \rightarrow \mathbb{R}^A$ is $(\mathcal{X} \otimes \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})) / \tilde{\mathcal{A}}$ -measurable. Induction hence proves that for all $n \in \mathbb{N}_0$, $\theta \in \Theta$ it holds that V_n^θ is $(\mathcal{X} \otimes \sigma(\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, \eta)})) / \tilde{\mathcal{A}}$ -measurable. This establishes item (i). Note that item (i) and (2.35) imply that for all $n \in \mathbb{N}_0$, $\theta, \vartheta \in \Theta$ it holds that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto [A \ni a \mapsto f(X^{\theta, x, a}(\omega), V_n^\vartheta(X^{\theta, x, a}(\omega), \omega)) \in \mathbb{R}] \in \mathbb{R}^A$ is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\vartheta, \eta)}) \cup \mathbb{F}^\theta)) / \tilde{\mathcal{A}}$ -measurable. This demonstrates item (ii). The proof of Lemma 2.2.2 is thus complete. $\square$

Lemma 2.2.3. *Assume Setting 2.2.1. Then for all $n \in \mathbb{N}_0$ it holds that V_n^θ , $\theta \in \Theta$, are identically distributed random fields.*

Proof of Lemma 2.2.3. For all $\theta \in \Theta$, $(x, a) \in \mathbb{X} \times A$ it holds that $(V_0^\theta(x))(a) = 0$. Therefore V_0^θ , $\theta \in \Theta$, are identically distributed random fields. Fix $n \in \mathbb{N}$ and assume that for all $l \in \{0, 1, \dots, n-1\}$ it holds that V_l^θ , $\theta \in \Theta$, are identically distributed random fields. Lemma 2.2.2 and [10, Lemma 2.6] ensure for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$ that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto (V_l^{(\theta, l, i)}(x, \omega), V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(x, \omega)) \in \mathbb{R}^A \times \mathbb{R}^A$ and $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto (V_l^{(0, l, i)}(x, \omega), V_{\max\{l-1, 0\}}^{(0, -l, i)}(x, \omega)) \in \mathbb{R}^A \times \mathbb{R}^A$ are identically distributed random fields. This ensures that for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$ it holds that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto f(x, V_l^{(\theta, l, i)}(x, \omega)) - \mathbb{1}_{\mathbb{N}}(l)f(x, V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(x, \omega)) \in \mathbb{R}$ and $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto f(x, V_l^{(0, l, i)}(x, \omega)) - \mathbb{1}_{\mathbb{N}}(l)f(x, V_{\max\{l-1, 0\}}^{(0, -l, i)}(x, \omega)) \in \mathbb{R}$ are identically distributed random fields. Combining this, the

assumption that for all $\theta \in \Theta$ it holds that X^θ is $(\mathcal{X} \otimes \mathbb{F}^\theta)/\tilde{\mathcal{X}}$ -measurable, and [10, Lemma 2.5] establishes that for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$ it holds that

$$\begin{aligned} \mathbb{X} \times A \times \Omega \ni (x, a, \omega) \mapsto & f(X^{(\theta, l, i), x, a}(\omega), V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \\ & - \mathbf{1}_{\mathbb{N}}(l) f(X^{(\theta, l, i), x, a}(\omega), V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \in \mathbb{R} \end{aligned} \quad (2.36)$$

and

$$\begin{aligned} \mathbb{X} \times A \times \Omega \ni (x, a, \omega) \mapsto & f(X^{(0, l, i), x, a}(\omega), V_l^{(0, l, i)}(X^{(0, l, i), x, a}(\omega), \omega)) \\ & - \mathbf{1}_{\mathbb{N}}(l) f(X^{(0, l, i), x, a}(\omega), V_{\max\{l-1, 0\}}^{(0, -l, i)}(X^{(0, l, i), x, a}(\omega), \omega)) \in \mathbb{R} \end{aligned} \quad (2.37)$$

are identically distributed random fields. This implies that for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$ it holds that

$$\begin{aligned} \mathbb{X} \times \Omega \ni (x, \omega) \mapsto & \left[A \ni a \mapsto f(X^{(\theta, l, i), x, a}(\omega), V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \right. \\ & \left. - \mathbf{1}_{\mathbb{N}}(l) f(X^{(\theta, l, i), x, a}(\omega), V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \in \mathbb{R} \right] \in \mathbb{R}^A \end{aligned} \quad (2.38)$$

and

$$\begin{aligned} \mathbb{X} \times \Omega \ni (x, \omega) \mapsto & \left[A \ni a \mapsto f(X^{(0, l, i), x, a}(\omega), V_l^{(0, l, i)}(X^{(0, l, i), x, a}(\omega), \omega)) \right. \\ & \left. - \mathbf{1}_{\mathbb{N}}(l) f(X^{(0, l, i), x, a}(\omega), V_{\max\{l-1, 0\}}^{(0, -l, i)}(X^{(0, l, i), x, a}(\omega), \omega)) \in \mathbb{R} \right] \in \mathbb{R}^A \end{aligned} \quad (2.39)$$

are identically distributed random fields. Let $g_k: (\mathbb{R}^A)^k \rightarrow \mathbb{R}^A$, $k \in \mathbb{N}$, satisfy for all $k \in \mathbb{N}$, $r_1, \dots, r_k \in \mathbb{R}^A$ that $g_k(r_1, \dots, r_k) = \sum_{j=1}^k r_j$. Note that for all $k \in \mathbb{N}$ it holds that g_k is $\tilde{\mathcal{A}}^{\otimes k}/\tilde{\mathcal{A}}$ -measurable. Let $U_{l,i}^\theta: \mathbb{X} \times \Omega \rightarrow \mathbb{R}^A$, $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$, satisfy for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$, $(x, \omega) \in \mathbb{X} \times \Omega$ that

$$\begin{aligned} U_{l,i}^\theta(x, \omega) = & \left[A \ni a \mapsto f(X^{(\theta, l, i), x, a}(\omega), V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \right. \\ & \left. - \mathbf{1}_{\mathbb{N}}(l) f(X^{(\theta, l, i), x, a}(\omega), V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \in \mathbb{R} \right]. \end{aligned} \quad (2.40)$$

Lemma 2.2.2 demonstrates that for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$ it holds that $U_{l,i}^\theta$ is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, l, i, \eta)} \cup \mathbb{F}^{(\theta, -l, i, \eta)}) \cup \mathbb{F}^{(\theta, l, i)}))/\tilde{\mathcal{A}}$ -measurable. Note that (2.32) ensures for all $(x, \omega) \in \mathbb{X} \times \Omega$, $\theta \in \Theta$ that $V_n^\theta(x, \omega) = \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \sum_{i=1}^{M^{n-l}} U_{l,i}^\theta(x, \omega)$. Let $\tilde{M} = \sum_{j=1}^n M^j \in \mathbb{N}$, let $Y_n^\theta: \mathbb{X} \times \Omega \rightarrow (\mathbb{R}^A)^{\tilde{M}}$, $\theta \in \Theta$, satisfy for all $\theta \in \Theta$ that

$$\begin{aligned} Y_n^\theta = & \left(\frac{1}{M^n} U_{0,1}^\theta, \frac{1}{M^n} U_{0,2}^\theta, \dots, \frac{1}{M^n} U_{0,M^n}^\theta, \frac{1}{M^{n-1}} U_{1,1}^\theta, \frac{1}{M^{n-1}} U_{1,2}^\theta, \dots, \frac{1}{M^{n-1}} U_{1,M^{n-1}}^\theta, \dots \right. \\ & \left. \dots, \frac{1}{M} U_{n-1,1}^\theta, \frac{1}{M} U_{n-1,2}^\theta, \dots, \frac{1}{M} U_{n-1,M}^\theta \right). \end{aligned} \quad (2.41)$$

Observe that for all $(x, \omega) \in \mathbb{X} \times \Omega$, $\theta \in \Theta$ it holds that

$$g_{\tilde{M}}(Y_n^\theta(x, \omega)) = \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \sum_{i=1}^{M^{n-l}} U_{l,i}^\theta(x, \omega) = V_n^\theta(x, \omega). \quad (2.42)$$

The fact that for all $l \in \{0, 1, \dots, n-1\}$, $i \in \mathbb{N}$, $\theta \in \Theta$ it holds that $U_{l,i}^\theta$ is $(\mathcal{X} \otimes \sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, l, i, \eta)} \cup \mathbb{F}^{(\theta, -l, i, \eta)}) \cup \mathbb{F}^{(\theta, l, i)}))/\tilde{\mathcal{A}}$ -measurable, the assumption that $(\mathbb{F}^\theta)_{\theta \in \Theta}$ are independent sigma-algebras, and [10, Lemma 2.6] prove for all $\theta \in \Theta$ that Y_n^θ and Y_n^0 are identically distributed random fields. This implies for all $\theta \in \Theta$ that V_n^θ and V_n^0 are identically distributed random fields. Induction and the fact that V_0^θ , $\theta \in \Theta$, are identically distributed random fields prove that for all $n \in \mathbb{N}_0$ it holds that V_n^θ , $\theta \in \Theta$, are identically distributed random fields. The proof of Lemma 2.2.3 is thus complete. $\square$

2.2.3. Integrability Properties for MLFP Approximations

Lemma 2.2.4. Assume Setting 2.2.1, assume A is finite, let $L \in [0, \infty)$, let $\mathfrak{w}: \mathbb{X} \rightarrow (0, \infty)$ be $\mathcal{X}/\mathcal{B}((0, \infty))$ -measurable, assume that for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{b \in A} |r(b) - s(b)|$, and $(\mathbb{E}[|f(X^{0,x,a}, 0)|^2 + |\mathfrak{w}(X^{0,x,a})|^2])^{\frac{1}{2}} \leq L\mathfrak{w}(x)$. Then it holds for all $n \in \mathbb{N}_0$ that $\sup_{x \in \mathbb{X}} (\mathbb{E}[\max_{a \in A} |(V_n^0(x))(a)|^2] |\mathfrak{w}(x)|^{-2}) < \infty$.

Proof of Lemma 2.2.4. For all $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that $(V_0^\theta(x))(a) = 0$. Hence it holds that $\sup_{x \in \mathbb{X}} (\mathbb{E}[\|V_0^0(x)\|_\infty^2] |\mathfrak{w}(x)|^{-2}) < \infty$. Fix $n \in \mathbb{N}$ and assume for all $l \in \{0, 1, \dots, n-1\}$ that $\sup_{x \in \mathbb{X}} (\mathbb{E}[\|V_l^0(x)\|_\infty^2] |\mathfrak{w}(x)|^{-2}) < \infty$. Note that for all $m \in \mathbb{N}$ it holds that

$$\sup_{x \in \mathbb{X}} \frac{\mathbb{E}[\|V_m^0(x)\|_\infty^2]}{|\mathfrak{w}(x)|^2} \leq \sup_{x \in \mathbb{X}} \sum_{a \in A} \frac{\mathbb{E}[|(V_m^0(x))(a)|^2]}{|\mathfrak{w}(x)|^2} \leq \sum_{a \in A} \sup_{x \in \mathbb{X}} \frac{\mathbb{E}[|(V_m^0(x))(a)|^2]}{|\mathfrak{w}(x)|^2}. \quad (2.43)$$

Lemma 2.2.3 and Jensen's inequality prove that for all $m \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that

$$\begin{aligned} \mathbb{E}[|(V_m^0(x))(a)|^2] &= \mathbb{E}[|(V_m^\theta(x))(a)|^2] \\ &\leq \sum_{l=0}^{m-1} \frac{m}{M^{m-l}} \sum_{i=1}^{M^{m-l}} \mathbb{E} \left[|f(X^{(\theta,l,i),x,a}, V_l^{(\theta,l,i)}(X^{(\theta,l,i),x,a})) \right. \\ &\quad \left. - \mathbb{1}_{\mathbb{N}}(l) f(X^{(\theta,l,i),x,a}, V_{\max\{l-1,0\}}^{(\theta,-l,i)}(X^{(\theta,l,i),x,a}))|^2 \right] \\ &\leq \sum_{l=0}^{m-1} \frac{2m}{M^{m-l}} \sum_{i=1}^{M^{m-l}} \mathbb{E} [|f(X^{(\theta,l,i),x,a}, V_l^{(\theta,l,i)}(X^{(\theta,l,i),x,a}))|^2] \\ &\quad + \mathbb{E} [|f(X^{(\theta,l,i),x,a}, V_{\max\{l-1,0\}}^{(\theta,-l,i)}(X^{(\theta,l,i),x,a}))|^2]. \end{aligned} \quad (2.44)$$

The triangle inequality, Jensen's inequality, the assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, and the assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $(\mathbb{E}[|f(X^{0,x,a}, 0)|^2])^{\frac{1}{2}} \leq L\mathfrak{w}(x)$ ensure that for all $m \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$, $\theta, \vartheta \in \Theta$ it holds that

$$\begin{aligned} \mathbb{E}[|f(X^{\theta,x,a}, V_m^\vartheta(X^{\theta,x,a}))|^2] &\leq 2\mathbb{E}[|f(X^{\theta,x,a}, V_m^\vartheta(X^{\theta,x,a})) - f(X^{\theta,x,a}, 0)|^2] + 2\mathbb{E}[|f(X^{\theta,x,a}, 0)|^2] \\ &\leq 2L^2 \mathbb{E}[\max_{b \in A} |(V_m^\vartheta(X^{\theta,x,a}))(b)|^2] + 2L^2 |\mathfrak{w}(x)|^2. \end{aligned} \quad (2.45)$$

Lemma 2.2.3 ensures for all $m \in \mathbb{N}$, $\vartheta \in \Theta$ that V_m^ϑ and V_m^0 are identically distributed. Combining this and the assumption that A is finite demonstrates that for all $m \in \mathbb{N}$, $\vartheta \in \Theta$ it holds that $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto \|V_m^\vartheta(x, \omega)\|_\infty^2 \in [0, \infty)$ and $\mathbb{X} \times \Omega \ni (x, \omega) \mapsto \|V_m^0(x, \omega)\|_\infty^2 \in [0, \infty)$ are identically distributed. This, Lemma 2.2.2, the assumption that for all $\theta \in \Theta$ it holds that X^θ is $(\mathcal{X} \otimes \mathbb{F}^\theta)/\tilde{\mathcal{X}}$ -measurable, the assumption that the sigma-algebras $\mathbb{F}^\theta$, $\theta \in \Theta$, are independent, and [51, Lemma 2.2] imply that for all $m \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$, $\theta, \vartheta \in \Theta$ with $\theta \notin \bigcup_{\eta \in \Theta} \{(\vartheta, \eta)\}$ it holds that

$$\begin{aligned} \mathbb{E}[\|V_m^\vartheta(X^{\theta,x,a})\|_\infty^2] &= \int_{\mathbb{X}} \mathbb{E}[\|V_m^\vartheta(y)\|_\infty^2] (X^{\theta,x,a}(\mathbb{P}))(dy) \\ &= \int_{\mathbb{X}} \mathbb{E}[\|V_m^0(y)\|_\infty^2] (X^{0,x,a}(\mathbb{P}))(dy). \end{aligned} \quad (2.46)$$

This, (2.45), the assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $(\mathbb{E}[|\mathfrak{w}(X^{0,x,a})|^2])^{\frac{1}{2}} \leq L\mathfrak{w}(x)$, and the assumption that for all $l \in \{0, 1, \dots, n-1\}$ it holds that $\sup_{x \in \mathbb{X}} (\mathbb{E}[\|V_l^0(x)\|_\infty^2] |\mathfrak{w}(x)|^{-2}) < \infty$

∞ prove for all $l \in \{0, 1, \dots, n-1\}$, $(x, a) \in \mathbb{X} \times A$, $\theta, \vartheta \in \Theta$ with $\theta \notin \bigcup_{\eta \in \Theta} \{(\vartheta, \eta)\}$ that

$$\begin{aligned} \frac{\mathbb{E}[|f(X^{\theta, x, a}, V_l^{\vartheta}(X^{\theta, x, a}))|^2]}{|\mathfrak{w}(x)|^2} &\leq \frac{2L^2}{|\mathfrak{w}(x)|^2} \int_{\mathbb{X}} \mathbb{E}[\|V_l^0(y)\|_{\infty}^2] (X^{0, x, a}(\mathbb{P}))(dy) + 2L^2 \\ &\leq 2L^2 \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|V_l^0(y)\|_{\infty}^2]}{|\mathfrak{w}(y)|^2} \right) \left(\frac{\mathbb{E}[|\mathfrak{w}(X^{0, x, a})|^2]}{|\mathfrak{w}(x)|^2} \right) + 2L^2 \\ &\leq 2L^4 \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|V_l^0(y)\|_{\infty}^2]}{|\mathfrak{w}(y)|^2} \right) + 2L^2. \end{aligned} \quad (2.47)$$

This ensures that for all $l \in \{0, 1, \dots, n-1\}$, $\theta, \vartheta \in \Theta$ with $\theta \notin \bigcup_{\eta \in \Theta} \{(\vartheta, \eta)\}$ that

$$\sup_{(x, a) \in \mathbb{X} \times A} \frac{\mathbb{E}[|f(X^{\theta, x, a}, V_l^{\vartheta}(X^{\theta, x, a}))|^2]}{|\mathfrak{w}(x)|^2} \leq 2L^4 \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|V_l^0(y)\|_{\infty}^2]}{|\mathfrak{w}(y)|^2} \right) + 2L^2 < \infty. \quad (2.48)$$

This and (2.44) demonstrate for all $a \in A$ that $\sup_{x \in \mathbb{X}} (\mathbb{E}[|(V_n^0(x))(a)|^2] |\mathfrak{w}(x)|^{-2}) < \infty$. This, the assumption that A is finite, and (2.43) show that $\sup_{x \in \mathbb{X}} (\mathbb{E}[\|V_n^0(x)\|_{\infty}^2] |\mathfrak{w}(x)|^{-2}) < \infty$. Induction hence proves that for all $n \in \mathbb{N}_0$ it holds that $\sup_{x \in \mathbb{X}} (\mathbb{E}[\max_{a \in A} |(V_n^0(x))(a)|^2] |\mathfrak{w}(x)|^{-2}) < \infty$. The proof of Lemma 2.2.4 is thus complete. $\square$

Lemma 2.2.5. *Assume Setting 2.2.1, assume A is finite, let $L \in [0, \infty)$, let $\mathfrak{w}: \mathbb{X} \rightarrow (0, \infty)$ be $\mathcal{X}/\mathcal{B}((0, \infty))$ -measurable, assume that for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{b \in A} |r(b) - s(b)|$, $(\mathbb{E}[|f(X^{0, x, a}, 0)|^2] + |\mathfrak{w}(X^{0, x, a})|^2)^{\frac{1}{2}} \leq L\mathfrak{w}(x)$. Then it holds for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ that $\mathbb{E}[|f(X^{0, x, a}, V_{n-1}^0(X^{0, x, a}))| + |(V_n^{\theta}(x))(a)|] < \infty$ and $\mathbb{E}[(V_n^{\theta}(x))(a)] = \mathbb{E}[f(X^{0, x, a}, V_{n-1}^0(X^{0, x, a}))]$.*

Proof of Lemma 2.2.5. The assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, Lemma 2.2.3, Lemma 2.2.4, the fact that $0 \notin \bigcup_{\eta \in \Theta} \{(0, \eta)\}$, the assumption that $(\mathbb{E}[|f(X^{0, x, a}, 0)|^2])^{\frac{1}{2}} \leq L\mathfrak{w}(x)$, and [51, Lemma 2.2] prove that for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that $\mathbb{E}[|f(X^{0, x, a}, V_{n-1}^0(X^{0, x, a}))|^2 + |(V_n^{\theta}(x))(a)|^2] < \infty$. This ensures for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ that $\mathbb{E}[|f(X^{0, x, a}, V_{n-1}^0(X^{0, x, a}))| + |(V_n^{\theta}(x))(a)|] < \infty$. This, Lemma 2.2.2, the assumption that X^{θ} , $\theta \in \Theta$, are i.i.d. random fields, Lemma 2.2.3, and [51, Lemma 2.2] demonstrate that for all $n \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$, $\theta, \vartheta \in \Theta$ with $\theta \notin \bigcup_{\eta \in \Theta} \{(\vartheta, \eta)\}$ it holds that

$$\mathbb{E}[f(X^{\theta, x, a}, V_n^{\vartheta}(X^{\theta, x, a}))] = \mathbb{E}[f(X^{0, x, a}, V_n^0(X^{0, x, a}))]. \quad (2.49)$$

This establishes for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ that

$$\begin{aligned} \mathbb{E}[(V_n^{\theta}(x))(a)] &= \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \mathbb{E}[f(X^{(\theta, l, i), x, a}, V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a}))] \\ &\quad - \mathbf{1}_{\mathbb{N}}(l) \mathbb{E}[f(X^{(\theta, l, i), x, a}, V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a}))] \\ &= \sum_{l=0}^{n-1} \mathbb{E}[f(X^{0, x, a}, V_l^0(X^{0, x, a}))] - \mathbf{1}_{\mathbb{N}}(l) \mathbb{E}[f(X^{0, x, a}, V_{\max\{l-1, 0\}}^0(X^{0, x, a}))] \\ &= \mathbb{E}[f(X^{0, x, a}, V_{n-1}^0(X^{0, x, a}))]. \end{aligned} \quad (2.50)$$

The proof of Lemma 2.2.5 is thus complete. $\square$

2.2.4. Recursive Error Bounds for MLFP Approximations

Lemma 2.2.6. *Assume Setting 2.2.1, assume A is finite, let $c, L \in [0, \infty)$, let $\mathbf{w}: \mathbb{X} \rightarrow (0, \infty)$ be $\mathcal{X}/\mathcal{B}((0, \infty))$ -measurable, let $v: \mathbb{X} \rightarrow \mathbb{R}^A$ be $\mathcal{X}/\tilde{\mathcal{A}}$ -measurable, assume that for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{b \in A} |r(b) - s(b)|$, $(\mathbb{E}[|f(X^{0,x,a}, 0)|^2 + \|v(X^{0,x,a})\|_\infty^2 + |\mathbf{w}(X^{0,x,a})|^2])^{\frac{1}{2}} \leq c\mathbf{w}(x)$, and $(v(x))(a) = \mathbb{E}[f(X^{0,x,a}, v(X^{0,x,a}))]$. Then it holds for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ that*

$$|(v(x))(a) - \mathbb{E}[(V_n^\theta(x))(a)]| \leq L \left(\mathbb{E}[\|v(X^{0,x,a}) - V_{n-1}^0(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}}. \quad (2.51)$$

Proof of Lemma 2.2.6. The assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $(v(x))(a) = \mathbb{E}[f(X^{0,x,a}, v(X^{0,x,a}))]$, Lemma 2.2.5, the assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, Lemma 2.2.4, and the Cauchy-Schwarz inequality imply that for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that

$$\begin{aligned} |v(x)(a) - \mathbb{E}[V_n^\theta(x)(a)]| &\leq \mathbb{E}[|f(X^{0,x,a}, v(X^{0,x,a})) - f(X^{0,x,a}, V_{n-1}^0(X^{0,x,a}))|] \\ &\leq L \mathbb{E}[\|v(X^{0,x,a}) - V_{n-1}^0(X^{0,x,a})\|_\infty] \\ &\leq L \left(\mathbb{E}[\|v(X^{0,x,a}) - V_{n-1}^0(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}}. \end{aligned} \quad (2.52)$$

The proof of Lemma 2.2.6 is thus complete. $\square$

Lemma 2.2.7. *Assume Setting 2.2.1, assume A is finite, let $c, L \in [0, \infty)$, let $\mathbf{w}: \mathbb{X} \rightarrow (0, \infty)$ be $\mathcal{X}/\mathcal{B}((0, \infty))$ -measurable, let $v: \mathbb{X} \rightarrow \mathbb{R}^A$ be $\mathcal{X}/\tilde{\mathcal{A}}$ -measurable, assume that for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{b \in A} |r(b) - s(b)|$, $(\mathbb{E}[|f(X^{0,x,a}, 0)|^2 + \|v(X^{0,x,a})\|_\infty^2 + |\mathbf{w}(X^{0,x,a})|^2])^{\frac{1}{2}} \leq c\mathbf{w}(x)$, and $(v(x))(a) = \mathbb{E}[f(X^{0,x,a}, v(X^{0,x,a}))]$. Then it holds for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ that*

$$\begin{aligned} &\left(\text{Var} \left[(V_n^\theta(x))(a) \right] \right)^{\frac{1}{2}} \\ &\leq \frac{1}{\sqrt{M^n}} \left(\mathbb{E}[|f(X^{0,x,a}, 0)|^2] \right)^{\frac{1}{2}} + \mathbb{1}_{[2, \infty)}(n) \frac{L}{\sqrt{M}} \left(\mathbb{E}[\|V_{n-1}^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} \\ &\quad + \mathbb{1}_{[2, \infty)}(n) \frac{L\sqrt{M}}{\sqrt{M^n}} \left(\mathbb{E}[\|v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} + \sum_{l=1}^{n-2} \frac{L(1 + \sqrt{M})}{\sqrt{M^{n-l}}} \left(\mathbb{E}[\|V_l^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}}. \end{aligned} \quad (2.53)$$

Proof of Lemma 2.2.7. Lemma 2.2.4 demonstrates for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ that $\text{Var}[(V_n^\theta(x))(a)] < \infty$. Lemma 2.2.2 yields that for all $l, i \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that

$$\begin{aligned} \Omega \ni \omega \mapsto &f(X^{(\theta, l, i), x, a}(\omega), V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \\ &- \mathbb{1}_{\mathbb{N}}(l) f(X^{(\theta, l, i), x, a}(\omega), V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \in \mathbb{R} \end{aligned} \quad (2.54)$$

is $\sigma((\bigcup_{\eta \in \Theta} \mathbb{F}^{(\theta, l, i, \eta)} \cup \mathbb{F}^{(\theta, -l, i, \eta)}) \cup \mathbb{F}^{(\theta, l, i)})/\mathcal{B}(\mathbb{R})$ -measurable. This demonstrates that for all $n \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$ it holds that

$$\begin{aligned} \text{Var} \left[(V_n^\theta(x))(a) \right] &= \sum_{l=0}^{n-1} \frac{1}{(M^{n-l})^2} \sum_{i=1}^{M^{n-l}} \text{Var} \left[f(X^{(\theta, l, i), x, a}, V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a})) \right. \\ &\quad \left. - \mathbb{1}_{\mathbb{N}}(l) f(X^{(\theta, l, i), x, a}, V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a})) \right]. \end{aligned} \quad (2.55)$$

Lemma 2.2.2, Lemma 2.2.3, and [10, Lemma 2.6] ensure that for all $l, i \in \mathbb{N}_0$, $\theta \in \Theta$ it holds that

$$\begin{aligned}\mathbb{X} \times \Omega \ni (x, \omega) &\mapsto (V_l^{(\theta, l, i)}(x, \omega), V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(x, \omega)) \in \mathbb{R}^A \times \mathbb{R}^A \quad \text{and} \\ \mathbb{X} \times \Omega \ni (x, \omega) &\mapsto (V_l^0(x, \omega), V_{\max\{l-1, 0\}}^1(x, \omega)) \in \mathbb{R}^A \times \mathbb{R}^A.\end{aligned}\tag{2.56}$$

are identically distributed random fields. This and the fact that f is $(\mathcal{X} \otimes \tilde{\mathcal{A}})/\mathcal{B}(\mathbb{R})$ -measurable show that for all $l, i \in \mathbb{N}_0$, $\theta \in \Theta$ it holds that

$$\begin{aligned}\mathbb{X} \times \Omega \ni (x, \omega) &\mapsto f(x, V_l^{(\theta, l, i)}(x, \omega)) - \mathbb{1}_{\mathbb{N}}(l)f(x, V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(x, \omega)) \in \mathbb{R} \quad \text{and} \\ \mathbb{X} \times \Omega \ni (x, \omega) &\mapsto f(x, V_l^0(x, \omega)) - \mathbb{1}_{\mathbb{N}}(l)f(x, V_{\max\{l-1, 0\}}^1(x, \omega)) \in \mathbb{R}\end{aligned}\tag{2.57}$$

are identically distributed random fields. This and [10, Lemma 2.5] prove that for all $l, i \in \mathbb{N}_0$, $\theta \in \Theta$ it holds that

$$\begin{aligned}\mathbb{X} \times A \times \Omega \ni (x, a, \omega) &\mapsto f(X^{(\theta, l, i), x, a}(\omega), V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \\ &\quad - \mathbb{1}_{\mathbb{N}}(l)f(X^{(\theta, l, i), x, a}(\omega), V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a}(\omega), \omega)) \in \mathbb{R} \quad \text{and} \\ \mathbb{X} \times A \times \Omega \ni (x, a, \omega) &\mapsto f(X^{0, x, a}(\omega), V_l^0(X^{0, x, a}(\omega), \omega)) \\ &\quad - \mathbb{1}_{\mathbb{N}}(l)f(X^{0, x, a}(\omega), V_{\max\{l-1, 0\}}^1(X^{0, x, a}(\omega), \omega)) \in \mathbb{R}\end{aligned}\tag{2.58}$$

are identically distributed random fields. This implies for all $l, i \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ that

$$\begin{aligned}\text{Var} \left[f(X^{(\theta, l, i), x, a}, V_l^{(\theta, l, i)}(X^{(\theta, l, i), x, a})) - \mathbb{1}_{\mathbb{N}}(l)f(X^{(\theta, l, i), x, a}, V_{\max\{l-1, 0\}}^{(\theta, -l, i)}(X^{(\theta, l, i), x, a})) \right] \\ = \text{Var} \left[f(X^{0, x, a}, V_l^0(X^{0, x, a})) - \mathbb{1}_{\mathbb{N}}(l)f(X^{0, x, a}, V_{\max\{l-1, 0\}}^1(X^{0, x, a})) \right].\end{aligned}\tag{2.59}$$

This and (2.55) ensure that for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that

$$\begin{aligned}\text{Var} [(V_n^\theta(x))(a)] &= \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \text{Var} \left[f(X^{0, x, a}, V_l^0(X^{0, x, a})) - \mathbb{1}_{\mathbb{N}}(l)f(X^{0, x, a}, V_{\max\{l-1, 0\}}^1(X^{0, x, a})) \right] \\ &\leq \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \mathbb{E} \left[\left| f(X^{0, x, a}, V_l^0(X^{0, x, a})) - \mathbb{1}_{\mathbb{N}}(l)f(X^{0, x, a}, V_{\max\{l-1, 0\}}^1(X^{0, x, a})) \right|^2 \right].\end{aligned}\tag{2.60}$$

Combining this, the fact that for all $(x, a) \in \mathbb{X} \times A$ it holds that $(V_0^0(x))(a) = 0$, and the fact that for all $n \in \mathbb{N}$, $r_1, r_2, \dots, r_n \in [0, \infty)$ it holds that $\sqrt{r_1 + r_2 + \dots + r_n} \leq \sqrt{r_1} + \sqrt{r_2} + \dots + \sqrt{r_n}$ prove that for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that

$$\begin{aligned}(\text{Var} [(V_n^\theta(x))(a)])^{\frac{1}{2}} &\leq \sum_{l=0}^{n-1} \frac{1}{\sqrt{M^{n-l}}} \left(\mathbb{E} \left[\left| f(X^{0, x, a}, V_l^0(X^{0, x, a})) - \mathbb{1}_{\mathbb{N}}(l)f(X^{0, x, a}, V_{\max\{l-1, 0\}}^1(X^{0, x, a})) \right|^2 \right] \right)^{\frac{1}{2}} \\ &= \sum_{l=1}^{n-1} \frac{1}{\sqrt{M^{n-l}}} \left(\mathbb{E} \left[\left| f(X^{0, x, a}, V_l^0(X^{0, x, a})) - \mathbb{1}_{\mathbb{N}}(l)f(X^{0, x, a}, V_{l-1}^1(X^{0, x, a})) \right|^2 \right] \right)^{\frac{1}{2}} \\ &\quad + \frac{1}{\sqrt{M^n}} \left(\mathbb{E} [|f(X^{0, x, a}, 0)|^2] \right)^{\frac{1}{2}}.\end{aligned}\tag{2.61}$$

The assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$ and the triangle inequality imply that for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that

$$\begin{aligned}
& (\text{Var} [(V_n^\theta(x))(a)])^{\frac{1}{2}} \\
& \leq \frac{1}{\sqrt{M^n}} \left(\mathbb{E}[|f(X^{0,x,a}, 0)|^2] \right)^{\frac{1}{2}} + \sum_{l=1}^{n-1} \frac{L}{\sqrt{M^{n-l}}} \left(\mathbb{E}[\|V_l^0(X^{0,x,a}) - V_{l-1}^1(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} \\
& \leq \sum_{l=1}^{n-1} \frac{L}{\sqrt{M^{n-l}}} \left[\left(\mathbb{E}[\|V_l^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} + \left(\mathbb{E}[\|v(X^{0,x,a}) - V_{l-1}^1(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} \right] \\
& \quad + \frac{1}{\sqrt{M^n}} \left(\mathbb{E}[|f(X^{0,x,a}, 0)|^2] \right)^{\frac{1}{2}}. \tag{2.62}
\end{aligned}$$

The fact that for all $l \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$ it holds that $\Omega \ni \omega \mapsto V_l^0(X^{0,x,a}(\omega), \omega) \in \mathbb{R}^A$ and $\Omega \ni \omega \mapsto V_l^1(X^{0,x,a}(\omega), \omega) \in \mathbb{R}^A$ are identically distributed ensures for all $l \in \mathbb{N}_0$, $(x, a) \in \mathbb{X} \times A$ that $\Omega \ni \omega \mapsto \|v(X^{0,x,a}(\omega)) - V_l^0(X^{0,x,a}(\omega), \omega)\|_\infty^2 \in [0, \infty)$ and $\Omega \ni \omega \mapsto \|v(X^{0,x,a}(\omega)) - V_l^1(X^{0,x,a}(\omega), \omega)\|_\infty^2 \in [0, \infty)$ are identically distributed. This and (2.62) establish that for all $n \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$, $\theta \in \Theta$ it holds that

$$\begin{aligned}
& (\text{Var} [(V_n^\theta(x))(a)])^{\frac{1}{2}} \\
& \leq \sum_{l=1}^{n-1} \frac{L}{\sqrt{M^{n-l}}} \left[\left(\mathbb{E}[\|V_l^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} + \left(\mathbb{E}[\|v(X^{0,x,a}) - V_{l-1}^0(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} \right] \\
& \quad + \frac{1}{\sqrt{M^n}} \left(\mathbb{E}[|f(X^{0,x,a}, 0)|^2] \right)^{\frac{1}{2}} \\
& = \frac{1}{\sqrt{M^n}} \left(\mathbb{E}[|f(X^{0,x,a}, 0)|^2] \right)^{\frac{1}{2}} + \sum_{l=1}^{n-1} \frac{L}{\sqrt{M^{n-l}}} \left(\mathbb{E}[\|V_l^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} \\
& \quad + \sum_{l=0}^{n-2} \frac{L\sqrt{M}}{\sqrt{M^{n-l}}} \left(\mathbb{E}[\|V_l^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} \\
& = \frac{1}{\sqrt{M^n}} \left(\mathbb{E}[|f(X^{0,x,a}, 0)|^2] \right)^{\frac{1}{2}} + \mathbb{1}_{[2, \infty)}(n) \frac{L}{\sqrt{M}} \left(\mathbb{E}[\|V_{n-1}^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} \\
& \quad + \mathbb{1}_{[2, \infty)}(n) \frac{L\sqrt{M}}{\sqrt{M^n}} \left(\mathbb{E}[\|v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}} + \sum_{l=1}^{n-2} \frac{L(1 + \sqrt{M})}{\sqrt{M^{n-l}}} \left(\mathbb{E}[\|V_l^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] \right)^{\frac{1}{2}}. \tag{2.63}
\end{aligned}$$

The proof of Lemma 2.2.7 is thus complete. $\square$

2.2.5. Non-Recursive Error Bounds for MLFP Approximations

Proposition 2.2.8. *Assume Setting 2.2.1, assume A is finite, let $c_f, c_v, c_{\mathfrak{w}}, L \in [0, \infty)$, let $\mathfrak{w}: \mathbb{X} \rightarrow (0, \infty)$ be $\mathcal{X}/\mathcal{B}((0, \infty))$ -measurable, let $v: \mathbb{X} \rightarrow \mathbb{R}^A$ be $\mathcal{X}/\tilde{\mathcal{A}}$ -measurable, let $c = \frac{3}{2} \max \left\{ \frac{c_v}{c_{\mathfrak{w}}}, c_v L + c_f, \frac{|A|c_f}{c_{\mathfrak{w}}|A|L+1} \right\}$, assume that for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{b \in A} |r(b) - s(b)|$, $(\mathbb{E}[|f(X^{0,x,a}, 0)|^2])^{\frac{1}{2}} \leq c_f \mathfrak{w}(x)$, $(\mathbb{E}[\|v(X^{0,x,a})\|_\infty^2])^{\frac{1}{2}} \leq c_v \mathfrak{w}(x)$, $(\mathbb{E}[|\mathfrak{w}(X^{0,x,a})|^2])^{\frac{1}{2}} \leq c_{\mathfrak{w}} \mathfrak{w}(x)$, and $(v(x))(a) = \mathbb{E}[f(X^{0,x,a}, v(X^{0,x,a}))]$. Then it holds for all $n \in \mathbb{N}$, $x \in \mathbb{X}$, $\theta \in \Theta$ that*

$$\left(\frac{\mathbb{E}[\|v(x) - V_n^\theta(x)\|_\infty^2]}{|\mathfrak{w}(x)|^2} \right)^{1/2} \leq c \left(\frac{c_{\mathfrak{w}} L (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(c_{\mathfrak{w}} L (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4 c_{\mathfrak{w}} L M^{-\frac{1}{2}} (|A| - 1)}}{2} \right)^n. \tag{2.64}$$

Proof of Proposition 2.2.8. The triangle inequality, Lemma 2.2.4, and the assumption that A is finite ensure that for all $n \in \mathbb{N}$, $x \in \mathbb{X}$, $\theta \in \Theta$ it holds that

$$\begin{aligned}
\left(\mathbb{E}[\|v(x) - V_n^\theta(x)\|_\infty^2]\right)^{\frac{1}{2}} &= \left(\mathbb{E}[\|v(x) - \mathbb{E}[V_n^\theta(x)] + \mathbb{E}[V_n^\theta(x)] - V_n^\theta(x)\|_\infty^2]\right)^{\frac{1}{2}} \\
&\leq \left(\mathbb{E}[\|v(x) - \mathbb{E}[V_n^\theta(x)]\|_\infty^2]\right)^{\frac{1}{2}} + \left(\mathbb{E}[\|V_n^\theta(x) - \mathbb{E}[V_n^\theta(x)]\|_\infty^2]\right)^{\frac{1}{2}} \\
&\leq \|v(x) - \mathbb{E}[V_n^\theta(x)]\|_\infty + \sum_{a \in A} \left(\mathbb{E}[|V_n^\theta(x)(a) - \mathbb{E}[V_n^\theta(x)(a)]|^2]\right)^{\frac{1}{2}} \\
&= \|v(x) - \mathbb{E}[V_n^\theta(x)]\|_\infty + \sum_{a \in A} (\text{Var}[V_n^\theta(x)(a)])^{\frac{1}{2}}. \tag{2.65}
\end{aligned}$$

The assumption that $X^0: \mathbb{X} \times \Omega \rightarrow \mathbb{X}^A$ is $(\mathcal{X} \otimes \mathbb{F}^0)/\tilde{\mathcal{X}}$ -measurable, the assumption that the sigma-algebras $\mathbb{F}^\theta$, $\theta \in \Theta$, are independent, the assumption that A is finite, Lemma 2.2.2, and the disintegration theorem [51, Lemma 2.2] prove that for all $l \in \mathbb{N}$, $(x, a) \in \mathbb{X} \times A$ it holds that

$$\begin{aligned}
\mathbb{E}[\|V_l^0(X^{0,x,a}) - v(X^{0,x,a})\|_\infty^2] &= \int_{\mathbb{X}} \mathbb{E}[\|V_l^0(y) - v(y)\|_\infty^2](X^{0,x,a}(\mathbb{P}))(dy) \\
&\leq \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|V_l^0(y) - v(y)\|_\infty^2]}{|\mathfrak{w}(y)|^2}\right) c_{\mathfrak{w}}^2 |\mathfrak{w}(x)|^2. \tag{2.66}
\end{aligned}$$

This, Lemma 2.2.6, Lemma 2.2.7, and (2.65) imply that for all $n \in \mathbb{N}$, $x \in \mathbb{X}$ it holds that

$$\begin{aligned}
&\left(\mathbb{E}[\|v(x) - V_n^0(x)\|_\infty^2]\right)^{\frac{1}{2}} \\
&\leq c_{\mathfrak{w}} \mathfrak{w}(x) L \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_{n-1}^0(y)\|_\infty^2]}{|\mathfrak{w}(y)|^2}\right)^{\frac{1}{2}} + \sum_{a \in A} \left[\frac{1}{\sqrt{M^n}} (\mathbb{E}[|f(X^{0,x,a}, 0)|^2])^{\frac{1}{2}}\right. \\
&\quad \left.+ \mathbb{1}_{[2,\infty)}(n) \frac{L\sqrt{M}}{\sqrt{M^n}} (\mathbb{E}[\|v(X^{0,x,a})\|_\infty^2])^{\frac{1}{2}} + \mathbb{1}_{[2,\infty)}(n) \frac{c_{\mathfrak{w}} \mathfrak{w}(x) L}{\sqrt{M}} \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_{n-1}^0(y)\|_\infty^2]}{|\mathfrak{w}(y)|^2}\right)^{\frac{1}{2}}\right. \\
&\quad \left.+ \sum_{l=1}^{n-2} \frac{c_{\mathfrak{w}} \mathfrak{w}(x) L(1 + \sqrt{M})}{\sqrt{M^{n-l}}} \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_l^0(y)\|_\infty^2]}{|\mathfrak{w}(y)|^2}\right)^{\frac{1}{2}}\right] \\
&\leq c_{\mathfrak{w}} \mathfrak{w}(x) L \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_{n-1}^0(y)\|_\infty^2]}{|\mathfrak{w}(y)|^2}\right)^{\frac{1}{2}} + \frac{|A| \mathfrak{w}(x)}{\sqrt{M^n}} (c_f + \mathbb{1}_{[2,\infty)}(n) c_v L \sqrt{M}) \\
&\quad + \mathbb{1}_{[2,\infty)}(n) \frac{c_{\mathfrak{w}} |A| L \mathfrak{w}(x)}{\sqrt{M}} \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_{n-1}^0(y)\|_\infty^2]}{|\mathfrak{w}(y)|^2}\right)^{\frac{1}{2}} \\
&\quad + c_{\mathfrak{w}} \mathfrak{w}(x) |A| L (1 + \sqrt{M}) \sum_{l=1}^{n-2} \frac{1}{\sqrt{M^{n-l}}} \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_l^0(y)\|_\infty^2]}{|\mathfrak{w}(y)|^2}\right)^{\frac{1}{2}}. \tag{2.67}
\end{aligned}$$

For all $n \in \mathbb{N}_0$ let

$$F_n = \left(\sup_{x \in \mathbb{X}} \frac{\mathbb{E}[\|v(x) - V_n^0(x)\|_\infty^2]}{|\mathfrak{w}(x)|^2}\right)^{\frac{1}{2}} \quad \text{and} \quad a_n = M^{\frac{n}{2}} F_n. \tag{2.68}$$

Lemma 2.2.4 ensures that for all $n \in \mathbb{N}_0$ it holds that $a_n, F_n \in [0, \infty)$. Note that (2.67) shows

for all $n \in \mathbb{N} \cap [2, \infty)$ that

$$\begin{aligned}
F_n &\leq |A|M^{-\frac{n}{2}}(c_f + c_v LM^{\frac{1}{2}}) + c_{\mathfrak{w}}L(1 + |A|M^{-\frac{1}{2}}) \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_{n-1}^0(y)\|_{\infty}^2]}{|\mathfrak{w}(y)|^2} \right)^{\frac{1}{2}} \\
&\quad + c_{\mathfrak{w}}|A|L(1 + M^{\frac{1}{2}}) \sum_{l=1}^{n-2} M^{-\frac{n-l}{2}} \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y) - V_l^0(y)\|_{\infty}^2]}{|\mathfrak{w}(y)|^2} \right)^{\frac{1}{2}} \\
&= |A|M^{-\frac{n}{2}}(c_f + c_v LM^{\frac{1}{2}}) + c_{\mathfrak{w}}L(1 + |A|M^{-\frac{1}{2}})F_{n-1} + c_{\mathfrak{w}}|A|L(1 + M^{\frac{1}{2}}) \sum_{l=1}^{n-2} M^{-\frac{n-l}{2}} F_l.
\end{aligned} \tag{2.69}$$

This implies for all $n \in \mathbb{N} \cap [2, \infty)$ that

$$\begin{aligned}
a_n &\leq |A|(c_f + c_v LM^{\frac{1}{2}}) + c_{\mathfrak{w}}M^{\frac{1}{2}}L(1 + |A|M^{-\frac{1}{2}})M^{\frac{n-1}{2}}F_{n-1} + c_{\mathfrak{w}}|A|L(1 + M^{\frac{1}{2}}) \sum_{l=1}^{n-2} M^{\frac{l}{2}}F_l \\
&= |A|(c_f + c_v LM^{\frac{1}{2}}) + c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|)a_{n-1} + c_{\mathfrak{w}}|A|L(1 + M^{\frac{1}{2}}) \sum_{l=1}^{n-2} a_l.
\end{aligned} \tag{2.70}$$

Moreover it holds that

$$a_0 = F_0 = \left(\sup_{y \in \mathbb{X}} \frac{\mathbb{E}[\|v(y)\|_{\infty}^2]}{|\mathfrak{w}(y)|^2} \right)^{\frac{1}{2}} = \sup_{(y,a) \in \mathbb{X} \times A} \frac{|(v(y))(a)|}{|\mathfrak{w}(y)|}. \tag{2.71}$$

The assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $(v(x))(a) = \mathbb{E}[f(X^{0,x,a}, v(X^{0,x,a}))]$, Jensen's inequality, the triangle inequality, the assumption that for all $x \in \mathbb{X}$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, and the assumption that for all $(x, a) \in \mathbb{X} \times A$ it holds that $(\mathbb{E}[\|f(X^{0,x,a}, 0)\|^2])^{\frac{1}{2}} \leq c_f \mathfrak{w}(x)$ and $(\mathbb{E}[\|v(X^{0,x,a})\|_{\infty}^2])^{\frac{1}{2}} \leq c_v \mathfrak{w}(x)$ prove for all $(x, a) \in \mathbb{X} \times A$ that

$$\begin{aligned}
\frac{|(v(x))(a)|}{|\mathfrak{w}(x)|} &= \frac{1}{\mathfrak{w}(x)} (\mathbb{E}[f(X^{0,x,a}, v(X^{0,x,a}))])^{\frac{1}{2}} \leq \frac{1}{\mathfrak{w}(x)} (\mathbb{E}[\|f(X^{0,x,a}, v(X^{0,x,a}))\|^2])^{\frac{1}{2}} \\
&\leq \frac{1}{\mathfrak{w}(x)} \left[L(\mathbb{E}[\sup_{b \in A} |v(X^{0,x,a})(b)|^2])^{\frac{1}{2}} + (\mathbb{E}[\|f(X^{0,x,a}, 0)\|^2])^{\frac{1}{2}} \right] \leq c_v L + c_f.
\end{aligned} \tag{2.72}$$

This and (2.71) ensure that $a_0 \leq c_v L + c_f$. Note that (2.67) implies that $a_1 \leq M^{\frac{1}{2}}(c_{\mathfrak{w}}La_0 + |A|M^{-\frac{1}{2}}c_f) \leq c_{\mathfrak{w}}LM^{\frac{1}{2}}(c_v L + c_f) + |A|c_f$. Moreover (2.70) ensures $a_2 \leq |A|(c_f + c_v LM^{\frac{1}{2}}) + c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|)a_1$. Let $(\xi_n)_{n \in \mathbb{N}_0}$, $(b_n)_{n \in \mathbb{N}_0} \subseteq \mathbb{R}$ satisfy for all $n \in \mathbb{N}_0$ that

$$\begin{aligned}
b_0 &= \xi_0 = \max \left\{ \frac{c_v}{c_{\mathfrak{w}}}, c_v L + c_f, \frac{|A|c_f}{c_{\mathfrak{w}}|A|L + 1} \right\}, \quad b_1 = |A|c_f - (c_{\mathfrak{w}}|A|L + 1)b_0, \\
b_2 &= |A|LM^{\frac{1}{2}}(c_v - c_{\mathfrak{w}}b_0), \quad b_{n+3} = 0, \quad \xi_1 = c_{\mathfrak{w}}LM^{\frac{1}{2}}b_0 + |A|c_f, \\
\xi_{n+2} &= |A|(c_f + c_v LM^{\frac{1}{2}}) + c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|)\xi_{n+1} + c_{\mathfrak{w}}|A|L(1 + M^{\frac{1}{2}}) \sum_{l=1}^n \xi_l.
\end{aligned} \tag{2.73}$$

Combining (2.67), (2.70), (2.73), and induction establishes for all $n \in \mathbb{N}_0$ that $a_n \leq \xi_n$. Observe that (2.73) ensures that for all $n \in \mathbb{N}_0$ it holds that

$$\begin{aligned}
\xi_0 &= b_0, \quad \xi_1 = b_1 + (c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1)\xi_0, \\
\xi_{n+2} &= b_{n+2} + (c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1)\xi_{n+1} + c_{\mathfrak{w}}LM^{\frac{1}{2}}(|A| - 1)\xi_n.
\end{aligned} \tag{2.74}$$

Let $x_1, x_2 \in \mathbb{R}$ satisfy

$$\begin{aligned} x_1 &= \frac{c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1 - \sqrt{(c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1)^2 + 4c_{\mathfrak{w}}LM^{\frac{1}{2}}(|A| - 1)}}{2}, \\ x_2 &= \frac{c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1 + \sqrt{(c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1)^2 + 4c_{\mathfrak{w}}LM^{\frac{1}{2}}(|A| - 1)}}{2}. \end{aligned} \quad (2.75)$$

Note that for all $i \in \{1, 2\}$ it holds that $x_i^2 = (c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1)x_i + c_{\mathfrak{w}}LM^{\frac{1}{2}}(|A| - 1)$. Moreover note that the assumption that A is nonempty yields that $c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1 \geq 1$ and $c_{\mathfrak{w}}LM^{\frac{1}{2}}(|A| - 1) \geq 0$. This ensures that $x_2 > 0 \geq x_1$. This and (2.75) imply that $|x_2| \geq |x_1|$. Hence it holds that

$$\frac{|x_2|}{|x_2 - x_1|} \leq 1 \quad \text{and} \quad \frac{|x_1|}{|x_2 - x_1|} \leq \frac{1}{2}. \quad (2.76)$$

This, the fact that $|x_1| \leq |x_2|$, and the triangle inequality prove that for all $n \in \mathbb{N}_0$ it holds that

$$\frac{x_2^{n+1} - x_1^{n+1}}{x_2 - x_1} \leq \frac{|x_2|^{n+1}}{x_2 - x_1} + \frac{|x_1|^{n+1}}{x_2 - x_1} \leq \frac{x_2}{x_2 - x_1}x_2^n + \frac{|x_1|}{x_2 - x_1}x_2^n \leq \frac{3}{2}x_2^n. \quad (2.77)$$

The discrete Gronwall-type two-step recursion in [55, Lemma 2.1] (applied with $\kappa \curvearrowright c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1$, $\lambda \curvearrowright c_{\mathfrak{w}}LM^{\frac{1}{2}}(|A| - 1)$, $(a_k)_{k \in \mathbb{N}_0} \curvearrowright (\xi_k)_{k \in \mathbb{N}_0}$, $(b_k)_{k \in \mathbb{N}_0} \curvearrowright (b_k)_{k \in \mathbb{N}_0}$, $x_{1/2} \curvearrowright x_{1/2}$ in the notation of [55, Lemma 2.1]) demonstrates that for all $n \in \mathbb{N}_0$ it holds that

$$\xi_n = \frac{1}{x_2 - x_1} \left(b_0(x_2^{n+1} - x_1^{n+1}) + b_1(x_2^n - x_1^n) + b_2(x_2^{\max\{n-1, 0\}} - x_1^{\max\{n-1, 0\}}) \right). \quad (2.78)$$

Note that (2.73) ensures that $b_1 \leq 0$ and $b_2 \leq 0$. This, the fact that $x_2 > x_1$, and the fact that for all $n \in \mathbb{N}_0$ it holds that $a_n \leq \xi_n$ proves that for all $n \in \mathbb{N}_0$ it holds that

$$a_n \leq b_0 \frac{x_2^{n+1} - x_1^{n+1}}{x_2 - x_1}. \quad (2.79)$$

Combining this and (2.77) implies that for all $n \in \mathbb{N}_0$ it holds that

$$F_n \leq \frac{3}{2}b_0(M^{-\frac{1}{2}}x_2)^n. \quad (2.80)$$

Furthermore, observe that

$$\begin{aligned} M^{-\frac{1}{2}}x_2 &= M^{-\frac{1}{2}} \left(\frac{c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1 + \sqrt{(c_{\mathfrak{w}}L(M^{\frac{1}{2}} + |A|) + 1)^2 + 4c_{\mathfrak{w}}LM^{\frac{1}{2}}(|A| - 1)}}{2} \right) \\ &= \frac{c_{\mathfrak{w}}L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(c_{\mathfrak{w}}L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4c_{\mathfrak{w}}LM^{-\frac{1}{2}}(|A| - 1)}}{2}. \end{aligned} \quad (2.81)$$

This, (2.68), and (2.80) prove that for all $x \in \mathbb{X}$, $n \in \mathbb{N}$ it holds that

$$\begin{aligned} &\left(\frac{\mathbb{E}[\|v(x) - V_n^\theta(x)\|_\infty^2]}{|\mathfrak{w}(x)|^2} \right)^{\frac{1}{2}} \\ &\leq c \left(\frac{c_{\mathfrak{w}}L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(c_{\mathfrak{w}}L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4c_{\mathfrak{w}}LM^{-\frac{1}{2}}(|A| - 1)}}{2} \right)^n. \end{aligned} \quad (2.82)$$

The proof of Proposition 2.2.8 is thus complete. $\square$

Corollary 2.2.9. Assume Setting 2.2.1, assume A is finite, let $c_f, c_v, c_w, L \in [0, \infty)$ with $c_w L < 1$, let $M > (1 + c_w L(2|A| - 1))^2(1 - c_w L)^{-2}$, let $w: \mathbb{X} \rightarrow (0, \infty)$ be $\mathcal{X}/\mathcal{B}((0, \infty))$ -measurable, let $v: \mathbb{X} \rightarrow \mathbb{R}^A$ be $\mathcal{X}/\mathcal{A}$ -measurable, assume for all $(x, a) \in \mathbb{X} \times A$, $r, s \in \mathbb{R}^A$ that $|f(x, r) - f(x, s)| \leq L \max_{b \in A} |r(b) - s(b)|$, $(\mathbb{E}[\|f(X^{0,x,a}, 0)\|^2])^{\frac{1}{2}} \leq c_f w(x)$, $(\mathbb{E}[\|v(X^{0,x,a})\|_\infty^2])^{\frac{1}{2}} \leq c_v w(x)$, $(\mathbb{E}[|w(X^{0,x,a})|^2])^{\frac{1}{2}} \leq c_w w(x)$, and $(v(x))(a) = \mathbb{E}[f(X^{0,x,a}, v(X^{0,x,a}))]$. Then it holds that

$$\limsup_{n \rightarrow \infty} \left[\sup_{x \in \mathbb{X}} \left(\frac{\mathbb{E}[\|v(x) - V_n^0(x)\|_\infty^2]}{|w(x)|^2} \right)^{\frac{1}{2}} \right] = 0. \quad (2.83)$$

Proof of Corollary 2.2.9. Let $c \in [0, \infty)$ satisfy $c = \frac{3}{2} \max \left\{ \frac{c_v}{c_w}, c_v L + c_f, \frac{|A|c_f}{c_w|A|L+1} \right\}$. Proposition 2.2.8 establishes that for all $n \in \mathbb{N}$, $x \in \mathbb{X}$ it holds that

$$\begin{aligned} & \left(\frac{\mathbb{E}[\|v(x) - V_n^0(x)\|_\infty^2]}{|w(x)|^2} \right)^{\frac{1}{2}} \\ & \leq c \left(\frac{c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4c_w L M^{-\frac{1}{2}}(|A| - 1)}}{2} \right)^n. \end{aligned} \quad (2.84)$$

Observe that the assumption that $M > (1 + c_w L(2|A| - 1))^2(1 - c_w L)^{-2}$ implies $M^{-\frac{1}{2}} < (1 - c_w L)(1 + c_w L(2|A| - 1))^{-1}$. This proves that

$$\begin{aligned} c_w L(|A| - 1)M^{-\frac{1}{2}} & < \frac{c_w L(|A| - 1)(1 - c_w L)}{1 + c_w L(2|A| - 1)} \\ & = 1 - \left(c_w L + (1 - c_w L) - \frac{c_w L(|A| - 1)(1 - c_w L)}{1 + c_w L(2|A| - 1)} \right) \\ & = 1 - \left(c_w L + (1 + c_w L(2|A| - 1) - c_w L(|A| - 1)) \frac{1 - c_w L}{1 + c_w L(2|A| - 1)} \right) \\ & = 1 - \left(c_w L + (1 + c_w |A|L) \frac{1 - c_w L}{1 + c_w L(2|A| - 1)} \right) \\ & < 1 - (c_w L + (1 + c_w |A|L)M^{-\frac{1}{2}}) = 1 - (c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}}). \end{aligned} \quad (2.85)$$

The assumption that A is nonempty ensures that $|A| \leq 2|A| - 1$. This and the assumption that $M > (1 + c_w L(2|A| - 1))^2(1 - c_w L)^{-2}$ yield $M > (1 + c_w L|A|)^2(1 - c_w L)^{-2}$. Therefore it holds that $c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} = c_w L + (1 + c_w |A|L)M^{-\frac{1}{2}} < 1 < 2$. This and (2.85) establish that

$$\begin{aligned} & \frac{c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4c_w L M^{-\frac{1}{2}}(|A| - 1)}}{2} \\ & < \frac{c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4 - 4(c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})}}{2} \\ & = \frac{c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(2 - c_w L(1 + |A|M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2}}{2} = 1. \end{aligned} \quad (2.86)$$

Combining (2.84) and (2.86) implies that

$$\limsup_{n \rightarrow \infty} \left[\sup_{x \in \mathbb{X}} \left(\frac{\mathbb{E}[\|v(x) - V_n^0(x)\|_\infty^2]}{|w(x)|^2} \right)^{\frac{1}{2}} \right] = 0. \quad (2.87)$$

The proof of Corollary 2.2.9 is thus complete. $\square$

2.3. Computational Complexity Analysis for MLFP Approximations

In this section we combine the error analysis for the proposed MLFP approximations in Proposition 2.2.8 in Subsection 2.2.5 with a computational cost analysis for the proposed MLFP approximation schemes to obtain in Theorem 2.3.1 in Subsection 2.3.1 below as the main result of this chapter a complexity analysis for the proposed MLFP approximation schemes. Corollary 2.3.2 in Subsection 2.3.2 below specializes Theorem 2.3.1 to the specific situation of a functional fixed-point equation taking the form of a Bellman equation associated to optimal stochastic control problems. Corollary 2.3.3 in Subsection 2.3.3 below specializes Corollary 2.3.2 to the specific situation of a functional fixed-point equation taking the form of a Bellman equation associated to optimal stopping problems. Theorem 2.0.1 in the introduction, in turn, is an immediate consequence of Corollary 2.3.2 in Subsection 2.3.2.

2.3.1. MLFP Approximations for Functional Fixed-Point Equations

Theorem 2.3.1. *Let $M \in \mathbb{N}$, let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, let A be a finite nonempty set, let $\kappa \in [0, \infty)$, let $\mathfrak{D}$ be a nonempty set, let $\lambda_d, L_d, \mathfrak{R}_d \in [0, \infty)$, $d \in \mathfrak{D}$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $(\mathbb{X}_d, \mathcal{X}_d)$, $d \in \mathfrak{D}$, be nonempty measurable spaces, let $\|\cdot\|_\infty : \mathbb{R}^A \rightarrow [0, \infty)$ satisfy for all $r \in \mathbb{R}^A$ that $\|r\|_\infty = \max_{a \in A} |r(a)|$, for every $d \in \mathfrak{D}$ let $\mathfrak{w}_d : \mathbb{X}_d \rightarrow (0, \infty)$ be $\mathcal{X}_d/\mathcal{B}((0, \infty))$ -measurable, for every $d \in \mathfrak{D}$ let $f_d : \mathbb{X}_d \times \mathbb{R}^A \rightarrow \mathbb{R}$ be $(\mathcal{X}_d \otimes (\bigotimes_{a \in A} \mathcal{B}(\mathbb{R})))/\mathcal{B}(\mathbb{R})$ -measurable, for every $d \in \mathfrak{D}$ let $\mathbb{F}_d^\theta \subseteq \mathcal{F}$, $\theta \in \Theta$, be independent sub-sigma-algebras of $\mathcal{F}$, for every $d \in \mathfrak{D}$ let $X_d^\theta : \mathbb{X}_d \times \Omega \rightarrow \mathbb{X}_d$, $\theta \in \Theta$, be i.i.d. random fields which satisfy for all $d \in \mathfrak{D}$, $\theta \in \Theta$ that X_d^θ is $(\mathcal{X}_d \otimes \mathbb{F}_d^\theta)/(\bigotimes_{a \in A} \mathcal{X}_d)$ -measurable, assume for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$, $r, s \in \mathbb{R}^A$ that $|f_d(x, r) - f_d(x, s)| \leq L_d \max_{b \in A} |r(b) - s(b)|$, $\left(\mathbb{E}[|f_d(X_d^{0,x,a}, 0)|^2]\right)^{\frac{1}{2}} \leq \kappa \mathfrak{w}_d(x)$, $\left(\mathbb{E}[|\mathfrak{w}_d(X_d^{0,x,a})|^2]\right)^{\frac{1}{2}} \leq \lambda_d \mathfrak{w}_d(x)$, and $\sup_{u \in \mathfrak{D}} \lambda_u L_u < 1$, assume*

$$M > \frac{(1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1))^2}{(1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d))^2}, \quad (2.88)$$

for every $d \in \mathfrak{D}$ let $V_{n,d}^\theta : \mathbb{X}_d \times \Omega \rightarrow \mathbb{R}^A$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, satisfy for all $n \in \mathbb{N}_0$, $(x, a) \in \mathbb{X}_d \times A$, $\theta \in \Theta$ that

$$\begin{aligned} (V_{n,d}^\theta(x))(a) &= \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \sum_{i=1}^{M^{n-l}} f_d(X_d^{(\theta,l,i),x,a}, V_{l,d}^{(\theta,l,i)}(X_d^{(\theta,l,i),x,a})) \\ &\quad - \mathbb{1}_{\mathbb{N}}(l) f_d(X_d^{(\theta,l,i),x,a}, V_{\max\{l-1,0\},d}^{(\theta,-l,i)}(X_d^{(\theta,l,i),x,a})), \end{aligned} \quad (2.89)$$

and let $\mathfrak{C}_{n,d} \in [0, \infty)$, $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$, satisfy for all $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$ that

$$\mathfrak{C}_{n,d} \leq \sum_{l=0}^{n-1} M^{n-l} (\mathfrak{R}_d + |A| \mathfrak{C}_{l,d} + \mathbb{1}_{\mathbb{N}}(l) |A| \mathfrak{C}_{\max\{l-1,0\},d}). \quad (2.90)$$

Then the following holds:

- (i) For every $d \in \mathfrak{D}$ there exists a unique function $v_d : \mathbb{X}_d \rightarrow \mathbb{R}^A$ which is $\mathcal{X}_d/(\bigotimes_{a \in A} \mathcal{B}(\mathbb{R}))$ -measurable and satisfies for all $(x, a) \in \mathbb{X}_d \times A$ that $\sup_{y \in \mathbb{X}_d} \frac{\|v_d(y)\|_\infty}{\mathfrak{w}_d(y)} < \infty$, $\mathbb{E}[|f_d(X_d^{0,x,a}, v_d(X_d^{0,x,a}))|] < \infty$, and

$$(v_d(x))(a) = \mathbb{E}[f_d(X_d^{0,x,a}, v_d(X_d^{0,x,a}))]. \quad (2.91)$$

(ii) There exist $\mathbf{n}: (0, 1] \rightarrow \mathbb{N}$ and $c \in \mathbb{R}$ such that for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ it holds that $\mathfrak{C}_{\mathbf{n}(\varepsilon), d} \leq c\mathfrak{R}_d\varepsilon^{-c}$ and

$$\sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\|v_d(x) - V_{\mathbf{n}(\varepsilon), d}^0(x)\|_\infty^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \leq \varepsilon. \quad (2.92)$$

Proof of Theorem 2.3.1. Throughout this proof let $\alpha, \beta, \gamma \in \mathbb{R} \cup \{\infty\}$ satisfy

$$\begin{aligned} \alpha &= \frac{1}{2} \sup_{d \in \mathfrak{D}} \left[\lambda_d L_d (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(\lambda_d L_d (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4M^{-\frac{1}{2}} \lambda_d L_d (|A| - 1)} \right], \\ \beta &= \frac{\ln(3|A|M)}{\ln(\alpha^{-1})}, \quad \text{and} \quad \gamma = \frac{3}{2} \sup_{d \in \mathfrak{D}} \left(\max \left\{ \frac{\kappa}{1 - \lambda_d L_d}, \frac{|A|\kappa}{\lambda_d |A| L_d + 1} \right\} \right), \end{aligned} \quad (2.93)$$

for every $d \in \mathfrak{D}$ let $\mathfrak{B}_{n,d} \in [0, \infty)$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$ that $\mathfrak{B}_{n,d} = |A|^{-n} \mathfrak{C}_{n,d}$, and let $\mathbf{n}: (0, 1] \rightarrow \mathbb{N} \cup \{\infty\}$ satisfy for all $\varepsilon \in (0, 1]$ that $\mathbf{n}(\varepsilon) = \min(\{n \in \mathbb{N} : \gamma \alpha^n \leq \varepsilon\} \cup \{\infty\})$. The assumption that $\sup_{d \in \mathfrak{D}} \lambda_d L_d < 1$ ensures that $\gamma \in [0, \infty)$. The triangle inequality yields

$$\begin{aligned} \alpha &= \frac{1}{2} \sup_{d \in \mathfrak{D}} \left[\lambda_d L_d (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right. \\ &\quad \left. + \sqrt{(\lambda_d L_d (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4M^{-\frac{1}{2}} \lambda_d L_d (|A| - 1)} \right], \\ &\leq \frac{1}{2} \left[(\sup_{d \in \mathfrak{D}} \lambda_d L_d) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right. \\ &\quad \left. + \sqrt{((\sup_{d \in \mathfrak{D}} \lambda_d L_d) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4M^{-\frac{1}{2}} (\sup_{d \in \mathfrak{D}} \lambda_d L_d) (|A| - 1)} \right]. \end{aligned} \quad (2.94)$$

The assumption that $M^{-\frac{1}{2}} < \frac{1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d)}{1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1)}$ ensures that

$$\begin{aligned} 1 - ((\sup_{d \in \mathfrak{D}} \lambda_d L_d) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}}) &= 1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d) - (1 + |A| (\sup_{d \in \mathfrak{D}} \lambda_d L_d)) M^{-\frac{1}{2}} \\ &> (1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d)) - (1 + |A| (\sup_{d \in \mathfrak{D}} \lambda_d L_d)) \frac{1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d)}{1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1)} \\ &= \left(1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1) - (|A| (\sup_{d \in \mathfrak{D}} \lambda_d L_d) + 1) \right) \frac{1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d)}{1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1)} \\ &= \left((\sup_{d \in \mathfrak{D}} \lambda_d L_d) (|A| - 1) \right) \frac{1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d)}{1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1)} \\ &> (\sup_{d \in \mathfrak{D}} \lambda_d L_d) (|A| - 1) M^{-\frac{1}{2}}. \end{aligned} \quad (2.95)$$

The assumption that A is nonempty implies that $|A| \leq 2|A| - 1$. Hence it holds that $M > \frac{(1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1))^2}{(1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d))^2} \geq \frac{(1 + |A| (\sup_{t \in T} \lambda_d L_d))^2}{(1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d))^2}$. This establishes that

$$(\sup_{d \in \mathfrak{D}} \lambda_d L_d) + (|A| (\sup_{d \in \mathfrak{D}} \lambda_d L_d) + 1) M^{-\frac{1}{2}} < 1. \quad (2.96)$$

Combining this, (2.94), and (2.95) demonstrates that

$$\begin{aligned}
\alpha &\leq \frac{1}{2} \left[\left(\sup_{d \in \mathfrak{D}} \lambda_d L_d \right) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right. \\
&\quad \left. + \sqrt{\left(\left(\sup_{d \in \mathfrak{D}} \lambda_d L_d \right) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right)^2 + 4 M^{-\frac{1}{2}} \left(\sup_{d \in \mathfrak{D}} \lambda_d L_d \right) (|A| - 1)} \right] \\
&< \frac{1}{2} \left[\left(\sup_{d \in \mathfrak{D}} \lambda_d L_d \right) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right. \\
&\quad \left. + \sqrt{\left(\left(\sup_{d \in \mathfrak{D}} \lambda_d L_d \right) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right)^2 + 4 \left(1 - \left(\sup_{d \in \mathfrak{D}} \left(\lambda_d L_d \right) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right) \right)} \right] \\
&= \frac{1}{2} \left[\left(\sup_{d \in \mathfrak{D}} \lambda_d L_d \right) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{\left(2 - \left(\sup_{d \in \mathfrak{D}} \left(\lambda_d L_d \right) (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} \right) \right)^2} \right] \\
&= 1.
\end{aligned} \tag{2.97}$$

This and (2.93) ensure that $\alpha \in [M^{-\frac{1}{2}}, 1)$. The assumption that $M > \frac{(1 + (\sup_{d \in \mathfrak{D}} \lambda_d L_d)(2|A| - 1))^2}{(1 - (\sup_{d \in \mathfrak{D}} \lambda_d L_d))^2}$ implies that $M > 1$. Hence it holds that $[M^{-\frac{1}{2}}, 1) \neq \emptyset$. The fact that $\alpha \in (0, 1)$ demonstrates that for all $\varepsilon \in (0, 1]$ it holds that $\mathfrak{n}(\varepsilon) \in \mathbb{N}$. The fact that $\alpha \in [M^{-\frac{1}{2}}, 1)$ and the fact that $M > 1$ ensure that $\beta \in (2, \infty)$. Corollary 2.1.3 yields that for all $d \in \mathfrak{D}$ there exists a unique function $v_d: \mathbb{X}_d \rightarrow \mathbb{R}^A$ which is $\mathcal{X}_d / (\bigotimes_{a \in A} \mathcal{B}(\mathbb{R}))$ -measurable and satisfies for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ that $\sup_{y \in \mathbb{X}_d} \frac{\|v_d(y)\|_\infty}{\mathfrak{w}_d(y)} < \infty$, $\mathbb{E}[|f_d(X_d^{0,x,a}, v_d(X_d^{0,x,a}))|] < \infty$, and $v_d(x, a) = \mathbb{E}[f_d(X_d^{0,x,a}, v_d(X_d^{0,x,a}))]$. This proves item (i). Lemma 2.1.4 implies for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ that $(\mathbb{E}[\|v_d(X_d^{0,x,a})\|_\infty^2])^{\frac{1}{2}} \leq \frac{\kappa \lambda_d}{1 - \lambda_d L_d} \mathfrak{w}_d(x)$. Proposition 2.2.8 establishes for all $n \in \mathbb{N}$, $d \in \mathfrak{D}$, $x \in \mathbb{X}_d$ that

$$\begin{aligned}
&\left(\frac{\mathbb{E}[\|v_d(x) - V_{n,d}^0(x)\|_\infty^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \\
&\leq \gamma \left(\frac{\lambda_d L_d (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}} + \sqrt{(\lambda_d L_d (1 + |A| M^{-\frac{1}{2}}) + M^{-\frac{1}{2}})^2 + 4 \lambda_d L_d M^{-\frac{1}{2}} (|A| - 1)}}{2} \right)^n.
\end{aligned} \tag{2.98}$$

This demonstrates for all $n \in \mathbb{N}$, $d \in \mathfrak{D}$ that

$$\sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\|v_d(x) - V_{n,d}^0(x)\|_\infty^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \leq \gamma \alpha^n. \tag{2.99}$$

Next, note that the fact that for all $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$ it holds that $\mathfrak{B}_{n,d} = |A|^{-n} \mathfrak{C}_{n,d}$ and the assumption that for all $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$ it holds that

$$\mathfrak{C}_{n,d} \leq \sum_{l=0}^{n-1} M^{n-l} (\mathfrak{R}_d + |A| \mathfrak{C}_{l,d} + \mathbf{1}_{\mathbb{N}}(l) |A| \mathfrak{C}_{\max\{l-1, 0\}, d}) \tag{2.100}$$

imply for all $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$ that

$$\begin{aligned}
\mathfrak{B}_{n,d} &\leq |A|^{-n} \sum_{l=0}^{n-1} M^{n-l} (\mathfrak{R}_d + |A| \mathfrak{C}_{l,d} + \mathbf{1}_{\mathbb{N}}(l) |A| \mathfrak{C}_{\max\{l-1, 0\}, d}) \\
&= \sum_{l=0}^{n-1} M^{n-l} (|A|^{-n} \mathfrak{R}_d + |A|^{1-n+l} \mathfrak{B}_{l,d} + \mathbf{1}_{\mathbb{N}}(l) |A|^{1-n+\max\{l-1, 0\}} \mathfrak{B}_{\max\{l-1, 0\}, d}).
\end{aligned} \tag{2.101}$$

Moreover, observe that the assumption that A is nonempty implies that for all $n \in \mathbb{N}$, $l \in \mathbb{N}_0 \cap [0, n-1]$ it holds that $\max\{|A|^{-n}, |A|^{1-n+l}, |A|^{1-n+\max\{l-1, 0\}}\} \leq 1$. Combining this, the

fact that for all $d \in \mathfrak{D}$ it holds that $\mathfrak{B}_{0,d} = |A|^0 \mathfrak{C}_{0,d} = 0$, and (2.101) assures for all $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$ that

$$\mathfrak{B}_{n,d} \leq \sum_{l=0}^{n-1} M^{n-l} (\mathfrak{R}_d + \mathfrak{B}_{l,d} + \mathbf{1}_{\mathbb{N}}(l) \mathfrak{B}_{\max\{l-1,0\},d}). \quad (2.102)$$

This and [6, Lemma 3.14] (applied for every $d \in \mathfrak{D}$ with $\alpha \curvearrowright \mathfrak{R}_d + \mathfrak{B}_{0,d}$, $\beta \curvearrowright \mathfrak{R}_d$, $M \curvearrowright M$, $(C_n)_{n \in \mathbb{N}_0} \curvearrowright (\mathfrak{B}_{n,d})_{n \in \mathbb{N}_0}$ in the notation of [6, Lemma 3.14]) prove that for all $n \in \mathbb{N}$, $d \in \mathfrak{D}$ it holds that

$$|A|^{-n} \mathfrak{C}_{n,d} = \mathfrak{B}_{n,d} \leq \left(\frac{\mathfrak{R}_d + \mathfrak{B}_{0,d} + \mathfrak{R}_d + \mathfrak{B}_{0,d}}{2} \right) (3M)^n = \mathfrak{R}_d (3M)^n. \quad (2.103)$$

Hence, it holds for all $n \in \mathbb{N}$, $d \in \mathfrak{D}$ that $\mathfrak{C}_{n,d} = |A|^n \mathfrak{B}_{n,d} \leq \mathfrak{R}_d (3|A|M)^n$. This, (2.99), and [6, Lemma 3.15] (applied for every $d \in \mathfrak{D}$ with $m \curvearrowright 1$, $\alpha \curvearrowright \alpha$, $\beta \curvearrowright 3|A|M$, $\kappa_1 \curvearrowright \gamma$, $\kappa_2 \curvearrowright \mathfrak{R}_d$,

$$(c_n)_{n \in \mathbb{N}} \curvearrowright (\mathfrak{C}_{n,d})_{n \in \mathbb{N}}, \quad (e_n)_{n \in \mathbb{N}} \curvearrowright \left(\sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\|v_d(x) - V_{n,d}^0(x)\|_\infty^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \right)_{n \in \mathbb{N}}, \quad (2.104)$$

$N \curvearrowright n$ in the notation of [6, Lemma 3.15]) implies for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ that

$$\sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\|v_d(x) - V_{n(\varepsilon),d}^0(x)\|_\infty^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \leq \varepsilon \quad \text{and} \quad \mathfrak{C}_{n(\varepsilon),d} \leq 3|A|M\mathfrak{R}_d \max\{1, \gamma\}^\beta \varepsilon^{-\beta}. \quad (2.105)$$

Let $c \in \mathbb{R}$ satisfy

$$c = \max\{\beta, 3|A|M \max\{1, \gamma\}^\beta\}. \quad (2.106)$$

Note that (2.105) and (2.106) ensure for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ that $\mathfrak{C}_{n(\varepsilon),d} \leq c\mathfrak{R}_d \varepsilon^{-c}$. This establishes item (ii). The proof of Theorem 2.3.1 is thus complete. $\square$

2.3.2. MLFP Approximations for Bellman Equations of Optimal Control Problems

Corollary 2.3.2. *Let $M \in \mathbb{N}$, $\kappa \in [0, \infty)$, $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, let A be a finite nonempty set, let $\mathfrak{D}$ be a nonempty set, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\lambda_d, \delta_d, \mathfrak{R}_d \in [0, \infty)$, $d \in \mathfrak{D}$, let $(\mathbb{X}_d, \mathcal{X}_d)$, $d \in \mathfrak{D}$, be nonempty measurable spaces, for every $d \in \mathfrak{D}$ let $\mathfrak{w}_d: \mathbb{X}_d \rightarrow (0, \infty)$ be $\mathcal{X}_d/\mathcal{B}((0, \infty))$ -measurable, for every $d \in \mathfrak{D}$ let $g_d: \mathbb{X}_d \times A \rightarrow \mathbb{R}$ be $(\mathcal{X}_d \otimes 2^A)/\mathcal{B}(\mathbb{R})$ -measurable, for every $d \in \mathfrak{D}$ let $\mathbb{F}_d^\theta \subseteq \mathcal{F}$, $\theta \in \Theta$, be independent sub-sigma-algebras of $\mathcal{F}$, for every $d \in \mathfrak{D}$ let $X_d^\theta = (X_d^{\theta,x,a}(\omega))_{x \in \mathbb{X}_d, a \in A, \omega \in \Omega}: \mathbb{X}_d \times A \times \Omega \rightarrow \mathbb{X}_d$, $\theta \in \Theta$, be i.i.d. random fields which satisfy for all $d \in \mathfrak{D}$, $\theta \in \Theta$ that X_d^θ is $(\mathcal{X}_d \otimes 2^A \otimes \mathbb{F}_d^\theta)/\mathcal{X}_d$ -measurable, assume $M > \frac{1 + (\sup_{d \in \mathfrak{D}} \lambda_d \delta_d)(2|A|-1)^2}{(1 - (\sup_{d \in \mathfrak{D}} \lambda_d \delta_d))^2}$, assume for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ that $\max_{b \in A} |g_d(x, b)| \leq \kappa \mathfrak{w}_d(x)$, $(\mathbb{E}[\|\mathfrak{w}_d(X_d^{0,x,a})\|^2])^{\frac{1}{2}} \leq \lambda_d \mathfrak{w}_d(x)$, $\sup_{u \in \mathfrak{D}} \lambda_u \delta_u < 1$, for every $d \in \mathfrak{D}$ let $\mathcal{Q}_{n,d}^\theta: \mathbb{X}_d \times A \times \Omega \rightarrow \mathbb{R}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $(x, a) \in \mathbb{X}_d \times A$, $\theta \in \Theta$ that*

$$\begin{aligned} \mathcal{Q}_{n,d}^\theta(x, a) &= g_d(x, a) + \sum_{l=0}^{n-1} \frac{\delta_d}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \max_{b \in A} \{ \mathcal{Q}_{l,d}^{(\theta,l,i)}(X_d^{(\theta,l,i),x,a}, b) \} \\ &\quad - \mathbf{1}_{\mathbb{N}}(l) \max_{b \in A} \{ \mathcal{Q}_{\max\{l-1,0\},d}^{(\theta,-l,i)}(X_d^{(\theta,l,i),x,a}, b) \}, \end{aligned} \quad (2.107)$$

and let $\mathfrak{C}_{n,d} \in [0, \infty)$, $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$, satisfy for all $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$ that

$$\mathfrak{C}_{n,d} \leq \sum_{l=0}^{n-1} M^{n-l} (\mathfrak{R}_d + |A| \mathfrak{C}_{l,d} + \mathbf{1}_{\mathbb{N}}(l) |A| \mathfrak{C}_{\max\{l-1,0\},d}). \quad (2.108)$$

Then the following holds:

(i) For every $d \in \mathfrak{D}$ there exists a unique function $Q_d: \mathbb{X}_d \times A \rightarrow \mathbb{R}$ which is $(\mathcal{X}_d \otimes 2^A)/\mathcal{B}(\mathbb{R})$ -measurable and satisfies for all $(x, a) \in \mathbb{X}_d \times A$ that $\sup_{y \in \mathbb{X}_d} \frac{\max_{b \in A} |Q_d(y, b)|}{\mathfrak{w}_d(y)} < \infty$, $\mathbb{E}[\max_{b \in A} Q_d(X_d^{0,x,a}, b)] < \infty$, and

$$Q_d(x, a) = g_d(x, a) + \delta_d \mathbb{E}[\max_{b \in A} Q_d(X_d^{0,x,a}, b)]. \quad (2.109)$$

(ii) There exist $n: (0, 1] \rightarrow \mathbb{N}$ and $c \in \mathbb{R}$ such that for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ it holds that $\mathfrak{C}_{n(\varepsilon), d} \leq c \mathfrak{R}_d \varepsilon^{-c}$ and

$$\sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\max_{a \in A} |Q_d(x, a) - \mathcal{Q}_{n(\varepsilon), d}^0(x, a)|^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \leq \varepsilon. \quad (2.110)$$

Proof of Corollary 2.3.2. Note that for every $d \in \mathfrak{D}$ it holds that the function $\mathbb{X}_d \times \mathbb{R}^A \ni (x, r) \mapsto \delta_d \max_{a \in A} \{g_d(x, a) + r(a)\} \in \mathbb{R}$ is $(\mathcal{X}_d \otimes (\bigotimes_{a \in A} \mathcal{B}(\mathbb{R}))) / \mathcal{B}(\mathbb{R})$ -measurable. Moreover, for all $d \in \mathfrak{D}$, $x \in \mathbb{X}_d$, $r, s \in \mathbb{R}^A$ it holds that

$$|\delta_d \max_{a \in A} \{g_d(x, a) + r(a)\} - \delta_d \max_{a \in A} \{g_d(x, a) + s(a)\}| \leq \delta_d \max_{a \in A} |r(a) - s(a)|. \quad (2.111)$$

The assumption that for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ it holds that $\max_{b \in A} |g_d(x, b)| \leq \kappa \mathfrak{w}_d(x)$, $(\mathbb{E}[|\mathfrak{w}_d(X_d^{0,x,a})|^2])^{\frac{1}{2}} \leq \lambda_d \mathfrak{w}_d(x)$, and $\sup_{u \in \mathfrak{D}} \lambda_u \delta_u < 1$ yields for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ that

$$\left(\mathbb{E} \left[\left| \delta_d \max_{b \in A} \{g_d(X_d^{0,x,a}, b)\} \right|^2 \right] \right)^{\frac{1}{2}} \leq \delta_d \kappa \left(\mathbb{E} \left[|\mathfrak{w}_d(X_d^{0,x,a})|^2 \right] \right)^{\frac{1}{2}} \leq \delta_d \kappa \lambda_d \mathfrak{w}_d(x) \leq \kappa \mathfrak{w}_d(x). \quad (2.112)$$

For every $d \in \mathfrak{D}$ let $R_{n,d}^\theta: \mathbb{X}_d \times A \times \Omega \rightarrow \mathbb{R}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $(x, a) \in \mathbb{X}_d \times A$, $\theta \in \Theta$ that

$$\begin{aligned} R_{n,d}^\theta(x, a) = & \sum_{l=0}^{n-1} \frac{\delta_d}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \max_{b \in A} \left\{ g_d(X_d^{(\theta, l, i), x, a}, b) + R_{l,d}^{(\theta, l, i)}(X_d^{(\theta, l, i), x, a}, b) \right\} \\ & - \mathbb{1}_{\mathbb{N}}(l) \max_{b \in A} \left\{ g_d(X_d^{(\theta, l, i), x, a}, b) + R_{\max\{l-1, 0\}, d}^{(\theta, -l, i)}(X_d^{(\theta, l, i), x, a}, b) \right\}. \end{aligned} \quad (2.113)$$

This and Theorem 2.3.1 (applied with $M \curvearrowright M$, $\Theta \curvearrowright \Theta$, $A \curvearrowright A$, $\kappa \curvearrowright \kappa$, $(\Omega, \mathcal{F}, \mathbb{P}) \curvearrowright (\Omega, \mathcal{F}, \mathbb{P})$, $\mathfrak{D} \curvearrowright \mathfrak{D}$, $(\lambda_d)_{d \in \mathfrak{D}} \curvearrowright (\lambda_d)_{d \in \mathfrak{D}}$, $(L_d)_{d \in \mathfrak{D}} \curvearrowright (\delta_d)_{d \in \mathfrak{D}}$, $(\mathfrak{R}_d)_{d \in \mathfrak{D}} \curvearrowright (\mathfrak{R}_d)_{d \in \mathfrak{D}}$, $(\mathbb{X}_d, \mathcal{X}_d)_{d \in \mathfrak{D}} \curvearrowright (\mathbb{X}_d, \mathcal{X}_d)_{d \in \mathfrak{D}}$, $(\mathfrak{w}_d)_{d \in \mathfrak{D}} \curvearrowright (\mathfrak{w}_d)_{d \in \mathfrak{D}}$, $(f_d)_{d \in \mathfrak{D}} \curvearrowright (\mathbb{X}_d \times \mathbb{R}^A \ni (x, r) \mapsto \delta_d \max_{a \in A} \{g_d(x, a) + r(a)\} \in \mathbb{R})_{d \in \mathfrak{D}}$, $(\mathbb{F}_d^\theta)_{d \in \mathfrak{D}, \theta \in \Theta} \curvearrowright (\mathbb{F}_d^\theta)_{d \in \mathfrak{D}, \theta \in \Theta}$, $(X_d^\theta)_{d \in \mathfrak{D}, \theta \in \Theta} \curvearrowright (X_d^\theta)_{d \in \mathfrak{D}, \theta \in \Theta}$, $(\mathfrak{C}_{n,d})_{n \in \mathbb{N}_0, d \in \mathfrak{D}} \curvearrowright (\mathfrak{C}_{n,d})_{n \in \mathbb{N}_0, d \in \mathfrak{D}}$, $(V_{n,d}^\theta)_{n \in \mathbb{N}_0, d \in \mathfrak{D}, \theta \in \Theta} \curvearrowright (R_{n,d}^\theta)_{n \in \mathbb{N}_0, d \in \mathfrak{D}, \theta \in \Theta}$ in the notation of Theorem 2.3.1) imply that

(I) for every $d \in \mathfrak{D}$ there exists a unique function $R_d: \mathbb{X}_d \times A \rightarrow \mathbb{R}$ which is $(\mathcal{X}_d \otimes 2^A)/\mathcal{B}(\mathbb{R})$ -measurable and satisfies for all $(x, a) \in \mathbb{X}_d \times A$ that $\sup_{y \in \mathbb{X}_d} \frac{\max_{b \in A} |R_d(y, b)|}{\mathfrak{w}_d(y)} < \infty$, $\mathbb{E}[\max_{b \in A} \{g_d(X_d^{0,x,a}, b) + R_d(X_d^{0,x,a}, b)\}] < \infty$,

$$R_d(x, a) = \delta_d \mathbb{E}[\max_{b \in A} \{g_d(X_d^{0,x,a}, b) + R_d(X_d^{0,x,a}, b)\}], \quad (2.114)$$

and

(II) there exist $n: (0, 1] \rightarrow \mathbb{N}$ and $c \in \mathbb{R}$ such that for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ it holds that $\mathfrak{C}_{n(\varepsilon), d} \leq c\mathfrak{R}_d\varepsilon^{-c}$ and

$$\sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\max_{a \in A} |R_d(x, a) - R_{n(\varepsilon), d}^0(x, a)|^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \leq \varepsilon. \quad (2.115)$$

For every $d \in \mathfrak{D}$ let $Q_d: \mathbb{X}_d \times A \rightarrow \mathbb{R}$ satisfy for all $(x, a) \in \mathbb{X}_d \times A$ that $Q_d(x, a) = g_d(x, a) + R_d(x, a)$. This and item (I) ensure for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ that

$$\begin{aligned} Q_d(x, a) &= g_d(x, a) + R_d(x, a) = g_d(x, a) + \delta_d \mathbb{E} \left[\max_{b \in A} \left\{ g_d(X_d^{0, x, a}, b) + R_d(X_d^{0, x, a}, b) \right\} \right] \\ &= g_d(x, a) + \delta_d \mathbb{E} \left[\max_{b \in A} \left\{ Q_d(X_d^{0, x, a}, b) \right\} \right]. \end{aligned} \quad (2.116)$$

Moreover, note that the assumption that for all $d \in \mathfrak{D}$, $x \in \mathbb{X}_d$ it holds that $\max_{a \in A} |g_d(x, a)| \leq \kappa \mathfrak{w}_d(x)$ and item (I) demonstrate that for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ it holds that $\sup_{y \in \mathbb{X}_d} \frac{\max_{b \in A} |Q_d(y, b)|}{\mathfrak{w}_d(y)} < \infty$ and $\mathbb{E} \left[\left| \max_{b \in A} \left\{ Q_d(X_d^{0, x, a}, b) \right\} \right| \right] < \infty$. Furthermore, for every $d \in \mathfrak{D}$ let $S_d: \mathbb{X}_d \times A \rightarrow \mathbb{R}$ be $(\mathcal{X}_d \otimes 2^A)/\mathcal{B}(\mathbb{R})$ -measurable and satisfy for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ that $\sup_{y \in \mathbb{X}_d} \frac{\max_{b \in A} |S_d(y, b)|}{\mathfrak{w}_d(y)} < \infty$, $\mathbb{E} \left[\left| \max_{b \in A} \left\{ S_d(X_d^{0, x, a}, b) \right\} \right| \right] < \infty$, and $S_d(x, a) = g_d(x, a) + \delta_d \mathbb{E} \left[\max_{b \in A} \left\{ S_d(X_d^{0, x, a}, b) \right\} \right]$. It holds for all $d \in \mathfrak{D}$, $(x, a) \in \mathbb{X}_d \times A$ that

$$\begin{aligned} S_d(x, a) - g_d(x, a) &= \delta_d \mathbb{E} \left[\max_{b \in A} \left\{ S_d(X_d^{0, x, a}, b) \right\} \right] \\ &= \delta_d \mathbb{E} \left[\max_{b \in A} \left\{ g_d(X_d^{0, x, a}, b) + S_d(X_d^{0, x, a}, b) - g_d(X_d^{0, x, a}, b) \right\} \right]. \end{aligned} \quad (2.117)$$

Item (I) implies that for all $d \in \mathfrak{D}$ it holds that $S_d - g_d = R_d$. This establishes for all $d \in \mathfrak{D}$ that $S_d = Q_d$. This proves item (i). Note that induction and (2.113) demonstrate that for all $d \in \mathfrak{D}$, $n \in \mathbb{N}_0$, $(x, a) \in \mathbb{X}_d \times A$, $\theta \in \Theta$ it holds that $\mathcal{Q}_{n, d}^\theta(x, a) = g_d(x, a) + R_{n, d}^\theta(x, a)$. Combining this and item (II) implies for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ that

$$\sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\max_{a \in A} |Q_d(x, a) - \mathcal{Q}_{n(\varepsilon), d}^0(x, a)|^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} = \sup_{x \in \mathbb{X}_d} \left(\frac{\mathbb{E}[\max_{a \in A} |R_d(x, a) - R_{n(\varepsilon), d}^0(x, a)|^2]}{|\mathfrak{w}_d(x)|^2} \right)^{\frac{1}{2}} \leq \varepsilon. \quad (2.118)$$

This proves item (ii). The proof of Corollary 2.3.2 is thus complete. $\square$

2.3.3. MLFP Approximations for Bellman Equations of Optimal Stopping Problems

Corollary 2.3.3. *Let $M \in \mathbb{N}$, let $\Theta = \bigcup_{n \in \mathbb{N}} \mathbb{Z}^n$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathfrak{D}$ be a nonempty set, let $\delta_d, \mathfrak{R}_d \in [0, \infty)$, $d \in \mathfrak{D}$, let $(\mathbb{X}_d, \mathcal{X}_d)$, $d \in \mathfrak{D}$, be nonempty Borel spaces, for every $d \in \mathfrak{D}$ let $g_d: \mathbb{X}_d \rightarrow \mathbb{R}$ and $G_d: \mathbb{X}_d \rightarrow \mathbb{R}$ be $\mathcal{X}_d/\mathcal{B}(\mathbb{R})$ -measurable, for every $d \in \mathfrak{D}$ let $\mathbb{F}_d^\theta \subseteq \mathcal{F}$, $\theta \in \Theta$, be independent sub-sigma-algebras of $\mathcal{F}$, for every $d \in \mathfrak{D}$ let $X_d^\theta = (X_d^{\theta, x}(\omega))_{x \in \mathbb{X}_d, \omega \in \Omega}: \mathbb{X}_d \times \Omega \rightarrow \mathbb{X}_d$, $\theta \in \Theta$, be i.i.d. random fields which satisfy for all $d \in \mathfrak{D}$, $\theta \in \Theta$ that X_d^θ is $(\mathcal{X}_d \otimes \mathbb{F}_d^\theta)/\mathcal{X}_d$ -measurable, assume $\sup_{d \in \mathfrak{D}} \delta_d < 1$, assume $M > \frac{(1+3(\sup_{d \in \mathfrak{D}} \delta_d))^2}{(1-(\sup_{d \in \mathfrak{D}} \delta_d))^2}$, assume $\sup_{d \in \mathfrak{D}} (\sup_{x \in \mathbb{X}_d} |g_d(x)| + |G_d(x)|) < \infty$, for every $d \in \mathfrak{D}$ let $\mathcal{Q}_{n, d}^\theta: \mathbb{X}_d \times \Omega \rightarrow \mathbb{R}$, $n \in \mathbb{N}_0$, $\theta \in \Theta$, satisfy for all $n \in \mathbb{N}_0$, $x \in \mathbb{X}_d$, $\theta \in \Theta$ that*

$$\begin{aligned} \mathcal{Q}_{n, d}^\theta(x) &= g_d(x) + \sum_{l=0}^{n-1} \frac{\delta_d}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \max \left\{ G_d(X_d^{(\theta, l, i), x}), \mathcal{Q}_{l, d}^{(\theta, l, i)}(X_d^{(\theta, l, i), x}) \right\} \\ &\quad - \mathbb{1}_{\mathbb{N}}(l) \max \left\{ G_d(X_d^{(\theta, l, i), x}), \mathcal{Q}_{\max\{l-1, 0\}, d}^{(\theta, -l, i)}(X_d^{(\theta, l, i), x}) \right\} \end{aligned} \quad (2.119)$$

and let $\mathfrak{C}_{n,d} \in [0, \infty)$, $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$, satisfy for all $n \in \mathbb{N}_0$, $d \in \mathfrak{D}$ that

$$\mathfrak{C}_{n,d} \leq \sum_{l=0}^{n-1} M^{n-l} (\mathfrak{R}_d + 2\mathfrak{C}_{l,d} + \mathbb{1}_{\mathbb{N}}(l) 2\mathfrak{C}_{\max\{l-1,0\},d}). \quad (2.120)$$

Then the following holds:

- (i) For every $d \in \mathfrak{D}$ there exists a unique function $Q_d: \mathbb{X}_d \rightarrow \mathbb{R}$ which is $\mathcal{X}_d/\mathcal{B}(\mathbb{R})$ -measurable and satisfies for all $x \in \mathbb{X}_d$ that $\sup_{y \in \mathbb{X}_d} |Q_d(y)| < \infty$, and

$$Q_d(x) = g_d(x) + \delta_d \mathbb{E} \left[\max \left\{ G_d(X_d^{0,x}), Q_d(X_d^{0,x}) \right\} \right]. \quad (2.121)$$

- (ii) There exist $n: (0,1] \rightarrow \mathbb{N}$ and $c \in \mathbb{R}$ such that for all $d \in \mathfrak{D}$, $\varepsilon \in (0,1]$ it holds that $\mathfrak{C}_{n(\varepsilon),d} \leq c\mathfrak{R}_d \varepsilon^{-c}$ and

$$\sup_{x \in \mathbb{X}_d} \left(\mathbb{E} \left[|Q_d(x) - \mathcal{Q}_{n(\varepsilon),d}^0(x)|^2 \right] \right)^{\frac{1}{2}} \leq \varepsilon. \quad (2.122)$$

Proof of Corollary 2.3.3. Let $\mathbb{Y}_d$, $d \in \mathfrak{D}$, be nonempty sets which satisfy that there exists $\Upsilon \in \bigcap_{d \in \mathfrak{D}} \mathbb{Y}_d$, and that for all $d \in \mathfrak{D}$ it holds that $\mathbb{Y}_d \setminus \{\Upsilon\} = \mathbb{X}_d$, for every $d \in \mathfrak{D}$ let $\mathcal{Y}_d = \sigma_{\mathbb{Y}_d}(\mathcal{X}_d)$, for every $d \in \mathfrak{D}$, $\theta \in \Theta$ let $Y_d^\theta = (Y_d^{\theta,y,a}(\omega))_{y \in \mathbb{Y}_d, a \in \{0,1\}, \omega \in \Omega}: \mathbb{Y}_d \times \{0,1\} \times \Omega \rightarrow \mathbb{Y}_d$ satisfy for all $d \in \mathfrak{D}$, $\theta \in \Theta$, $(y,a) \in \mathbb{Y}_d \times \{0,1\}$, $\omega \in \Omega$ that

$$Y_d^{\theta,y,a}(\omega) = \begin{cases} \Upsilon & : (y = \Upsilon) \vee (a = 0), \\ X_d^{\theta,y}(\omega) & : (y \neq \Upsilon) \wedge (a = 1). \end{cases} \quad (2.123)$$

Note that for all $d \in \mathfrak{D}$ it holds that Y_d^θ , $\theta \in \Theta$, are i.i.d. random fields and that for every $d \in \mathfrak{D}$, $\theta \in \Theta$ it holds that Y_d^θ is $(\mathcal{Y}_d \otimes 2^{\{0,1\}} \otimes \mathbb{F}_d^\theta)/\mathcal{Y}_d$ -measurable. For every $d \in \mathfrak{D}$ let $h_d: \mathbb{Y}_d \times \{0,1\} \rightarrow \mathbb{R}$ satisfy for all $d \in \mathfrak{D}$, $(y,a) \in \mathbb{Y}_d \times \{0,1\}$ that

$$h_d(y,a) = \begin{cases} 0 & : y = \Upsilon, \\ g_d(y) & : y \neq \Upsilon, a = 1, \\ G_d(y) & : y \neq \Upsilon, a = 0. \end{cases} \quad (2.124)$$

Note that for all $d \in \mathfrak{D}$ it holds that h_d is $(\mathcal{Y}_d \otimes 2^{\{0,1\}})/\mathcal{B}(\mathbb{R})$ -measurable. Moreover, for all $d \in \mathfrak{D}$, $(y,a) \in \mathbb{X}_d \times \{0,1\}$ it holds that

$$|h_d(y,a)| \leq \max\{|g_d(y)|, |G_d(y)|\} \leq \sup_{u \in \mathfrak{D}} \left(\sup_{x \in \mathbb{X}_d} |g_u(x)| + |G_u(x)| \right). \quad (2.125)$$

For every $d \in \mathfrak{D}$ let $q_{n,d}^\theta: \mathbb{Y}_d \times \{0,1\} \times \Omega \rightarrow \mathbb{Y}_d$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}_0$, $\theta \in \Theta$, $(y,a) \in \mathbb{Y}_d \times \{0,1\}$ that

$$\begin{aligned} q_{n,d}^\theta(y,a) = h_d(y,a) + \sum_{l=0}^{n-1} \frac{\delta_d}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \max_{b \in \{0,1\}} \{ q_{l,d}^{(\theta,l,i)}(Y_d^{(\theta,l,i),y,a}, b) \} \\ - \mathbb{1}_{\mathbb{N}}(l) \max_{b \in \{0,1\}} \{ q_{\max\{l-1,0\},d}^{(\theta,-l,i)}(Y_d^{(\theta,l,i),y,a}, b) \}. \end{aligned} \quad (2.126)$$

Corollary 2.3.2 (applied with $A \curvearrowright \{0,1\}$, $\kappa \curvearrowright \sup_{d \in \mathfrak{D}} (\sup_{x \in \mathbb{X}_d} |g_d(x)| + |G_d(x)|)$, $g_d \curvearrowright h_d$, $(\mathbb{X}_d, \mathcal{X}_d) \curvearrowright (\mathbb{Y}_d, \mathcal{Y}_d)$, $\mathcal{Q}_{n,d}^\theta \curvearrowright q_{n,d}^\theta$ for $d \in \mathfrak{D}$ in the notation of Corollary 2.3.2) establishes the following:

- (I) For every $d \in \mathfrak{D}$ there exists a unique function $q_d: \mathbb{Y}_d \times \{0, 1\} \rightarrow \mathbb{R}$ which is $(\mathcal{Y}_d \otimes 2^{\{0,1\}})/\mathcal{B}(\mathbb{R})$ -measurable and satisfies for all $(y, a) \in \mathbb{Y}_d \times \{0, 1\}$ that $\sup_{(z,b) \in \mathbb{Y}_d \times \{0,1\}} |q_d(z, b)| < \infty$ and

$$q_d(y, a) = h_d(y, a) + \delta_d \mathbb{E} \left[\max_{b \in \{0,1\}} q_d(Y_d^{0,y,a}, b) \right]. \quad (2.127)$$

- (II) There exist $n: (0, 1] \rightarrow \mathbb{N}$ and $c \in \mathbb{R}$ such that for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ it holds that $\mathfrak{C}_{n(\varepsilon),d} \leq c \mathfrak{R}_d \varepsilon^c$ and

$$\sup_{y \in \mathbb{Y}_d} \left(\mathbb{E} \left[\max_{a \in \{0,1\}} |q_d(y, a) - q_{n(\varepsilon),d}^0(y, a)|^2 \right] \right)^{\frac{1}{2}} \leq \varepsilon. \quad (2.128)$$

Note that for all $d \in \mathfrak{D}$, $(y, a) \in \mathbb{Y}_d \setminus \{\Upsilon\} \times \{0, 1\}$ it holds that $q_d(\Upsilon, a) = 0$ and $q_d(y, 0) = G_d(y)$. Moreover, for all $d \in \mathfrak{D}$, $y \in \mathbb{Y}_d \setminus \{\Upsilon\}$ it holds that

$$\begin{aligned} q_d(y, 1) &= h_d(y, 1) + \delta_d \mathbb{E} \left[\max \{ q_d(Y_d^{0,y,1}, 0), q_d(Y_d^{0,y,1}, 1) \} \right] \\ &= g_d(y) + \delta_d \mathbb{E} \left[\max \{ G_d(X_d^{0,y}), q_d(X_d^{0,y}, 1) \} \right]. \end{aligned} \quad (2.129)$$

This and item (I) demonstrate for all $d \in \mathfrak{D}$ that there exists a unique function $Q_d: \mathbb{X}_d \rightarrow \mathbb{R}$ which is $\mathcal{X}_d/\mathcal{B}(\mathbb{R})$ -measurable and satisfies for all $x \in \mathbb{X}_d$ that $\sup_{y \in \mathbb{X}_d} |Q_d(y)| < \infty$ and

$$Q_d(x) = g_d(x) + \delta_d \mathbb{E} \left[\max \{ G_d(X_d^{0,x}), Q_d(X_d^{0,x}) \} \right]. \quad (2.130)$$

This proves item (i). Furthermore, note that for all $d \in \mathfrak{D}$, $\theta \in \Theta$, $(y, a) \in \mathbb{Y}_d \setminus \{\Upsilon\} \times \{0, 1\}$ it holds $\mathbb{P}$ -a.s. that $q_{0,d}^\theta(\Upsilon, a) = 0$, $q_{0,d}^\theta(y, 0) = G_d(y)$, and $q_{0,d}^\theta(y, 1) = g_d(y)$. This and induction yield that for all $d \in \mathfrak{D}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, $(y, a) \in \mathbb{Y}_d \setminus \{\Upsilon\} \times \{0, 1\}$ it holds $\mathbb{P}$ -a.s. that $q_{n,d}^\theta(\Upsilon, a) = 0$ and $q_{n,d}^\theta(y, 0) = G_d(y)$. Combining this, (2.123), (2.124), and (2.126) demonstrates that for all $d \in \mathfrak{D}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, $y \in \mathbb{Y}_d \setminus \{\Upsilon\} = \mathbb{X}_d$ it holds $\mathbb{P}$ -a.s. that

$$\begin{aligned} q_{n,d}^\theta(y, 1) &= h_d(y, 1) + \sum_{l=0}^{n-1} \frac{\delta_d}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \max_{b \in \{0,1\}} \left\{ q_{l,d}^{(\theta,l,i)}(Y_d^{(\theta,l,i),y,1}, b) \right\} \\ &\quad - \mathbb{1}_{\mathbb{N}}(l) \max_{b \in \{0,1\}} \left\{ q_{\max\{l-1,0\},d}^{(\theta,-l,i)}(Y_d^{(\theta,l,i),y,1}, b) \right\} \\ &= g_d(y) + \sum_{l=0}^{n-1} \frac{\delta_d}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \max \left\{ G_d(X_d^{(\theta,l,i),y}), q_{l,d}^{(\theta,l,i)}(X_d^{(\theta,l,i),y}, 1) \right\} \\ &\quad - \mathbb{1}_{\mathbb{N}}(l) \max \left\{ G_d(X_d^{(\theta,l,i),y}), q_{\max\{l-1,0\},d}^{(\theta,-l,i)}(X_d^{(\theta,l,i),y}, 1) \right\}. \end{aligned} \quad (2.131)$$

This and (2.119) yield that for all $d \in \mathfrak{D}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, $y \in \mathbb{Y}_d \setminus \{\Upsilon\} = \mathbb{X}_d$ it holds $\mathbb{P}$ -a.s. that

$$q_{n,d}^\theta(y, 1) = Q_{n,d}^\theta(y). \quad (2.132)$$

Combining this and (2.128) implies that for all $d \in \mathfrak{D}$, $\varepsilon \in (0, 1]$ it holds that

$$\sup_{x \in \mathbb{X}_d} \left(\mathbb{E} \left[|Q_d(x) - Q_{n(\varepsilon),d}^0(x)|^2 \right] \right)^{\frac{1}{2}} \leq \sup_{y \in \mathbb{Y}_d} \left(\mathbb{E} \left[\max_{a \in \{0,1\}} |q_d(y, a) - q_{n(\varepsilon),d}^0(y, a)|^2 \right] \right)^{\frac{1}{2}} \leq \varepsilon. \quad (2.133)$$

This establishes item (ii). The proof of Corollary 2.3.3 is thus complete. $\square$

The results of this chapter establish Monte Carlo approximations in the form of the MLFP approximation scheme for the considered functional fixed-point equations and, as a specialization, provide approximations for the Q -functions of the considered classes of MDPs. In particular, under suitable assumptions on high-dimensional instances of the MDPs, the MLFP approximation method achieves a prescribed approximation accuracy with a computational cost that grows at most polynomially in the reciprocal of the approximation accuracy and in the dimension of the underlying state space.

Following the general approximation strategy outlined in Section 1.3, the next step concerns the transfer of these algorithmic results to approximations by means of ANNs. This requires additional tools. First, the components of the MLFP scheme have to be represented by ANNs. Second, since the problem inputs of the associated stochastic control problem, i.e., the nonlinearity and the one-step transitions, may only be approximated by ANNs, the effect of these perturbations on the solutions of the corresponding fixed-point equations has to be controlled. The following chapter introduces the ANN framework and calculus employed for this purpose. Afterwards, Chapter 4 combines stability results for functional fixed-point equations with an ANN representation of the MLFP approximation method to establish ANN approximation results.

Chapter 3.

Artificial Neural Network (ANN) Calculus

In this chapter, we collect the notation, definitions, and auxiliary results on ANNs that are used throughout the remainder of this thesis. In particular, we introduce the ANN framework considered in this thesis and present constructions for the composition, extension, parallelization, and linear combination of ANNs. Throughout this thesis, we consider fully connected feedforward ANNs and follow the framework and notation used, e.g., in Jentzen et al. [57] (see also, e.g., Grohs et al. [41, Section 2], Grohs et al. [42, Section 3], Ackermann et al. [1, Section 2], and the references therein).

The precise formulations of the definitions and statements presented in this chapter are largely based on Grohs et al. [41, Section 2], Grohs et al. [42, Section 3], and Ackermann et al. [1, Section 2]. Several definitions and auxiliary results are stated in an adapted form to cover variants required in the subsequent chapters. Whenever a statement is extended beyond the cited formulation, a proof of the extension is included. For transparency, we specify below the correspondence between the definitions and results in this chapter and their respective sources. Definition 3.1.1 is a slightly extended version of [41, Definition 2.1]. Definitions 3.1.4 and 3.1.5 correspond to [41, Definition 2.2] and [41, Definition 2.3], respectively. Definition 3.2.1 is [41, Definition 2.5], Lemma 3.2.2 is a version of [41, Proposition 2.6], and Lemma 3.2.3 corresponds to [41, Lemma 2.8]. Definitions 3.2.5 and 3.2.6 are [41, Definition 2.9] and [41, Definition 2.11], respectively. Lemma 3.2.7 is a slightly extended version of [41, Lemma 2.13], for which a proof of the extension is provided. Lemma 3.2.8 is [41, Lemma 2.14], Definition 3.3.1 is a slightly modified version of [41, Definition 2.17], and Lemma 3.3.2 combines [41, Lemma 2.18, Proposition 2.19, and Proposition 2.20]. Definition 3.3.3 is [41, Definition 2.22] and Corollary 3.3.4 is a slightly extended version of [41, Corollary 2.23], where a proof of the extension is provided. Definition 3.4.1 is based on [1, Definitions 2.13 and 2.15], while Definitions 3.4.2 to 3.4.4 correspond to [1, Definitions 2.16 to 2.18], respectively. Lemma 3.4.5 is [42, Lemma 3.14]. Lemmas 3.4.6 and 3.4.7 are slightly modified versions of [1, Lemma 2.19] and [1, Lemma 2.20], respectively, where proofs of the modified statements are included.

The framework presented in this chapter will be used both for the ANN approximation of functional fixed-point equations of the form (1.3) in Chapter 4 of this part and for the ANN approximation of expectations of the form (1.5) in Chapter 6 in Part II.

3.1. Structured Descriptions of ANNs

Definition 3.1.1 (ANNs). We denote by $\mathbf{N}$ the set given by

$$\mathbf{N} = \bigcup_{L \in \mathbb{N}} \bigcup_{l_0, l_1, \dots, l_L \in \mathbb{N}} (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})) \quad (3.1)$$

and we denote by $\mathcal{P}: \mathbf{N} \rightarrow \mathbb{N}$, $\mathcal{L}: \mathbf{N} \rightarrow \mathbb{N}$, $\mathcal{I}: \mathbf{N} \rightarrow \mathbb{N}$, $\mathcal{O}: \mathbf{N} \rightarrow \mathbb{N}$, $\mathcal{H}: \mathbf{N} \rightarrow \mathbb{N}_0 = \mathbb{N} \cup \{0\}$, $\mathcal{D}: \mathbf{N} \rightarrow \bigcup_{l=2}^{\infty} \mathbb{N}^l$, and $\mathbb{D}_n: \mathbf{N} \rightarrow \mathbb{N}_0$, $n \in \mathbb{N}_0$, the functions which satisfy for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi \in (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, $n \in \mathbb{N}_0$ that $\mathcal{P}(\Phi) = \sum_{k=1}^L l_k(l_{k-1} + 1)$, $\mathcal{L}(\Phi) = L$, $\mathcal{I}(\Phi) = l_0$, $\mathcal{O}(\Phi) = l_L$, $\mathcal{H}(\Phi) = L - 1$, $\mathcal{D}(\Phi) = (l_0, l_1, \dots, l_L)$, and

$$\mathbb{D}_n(\Phi) = \begin{cases} l_n & : n \leq L \\ 0 & : n > L. \end{cases} \quad (3.2)$$

Definition 3.1.2 (ANN). We say that Φ is an ANN if and only if it holds that $\Phi \in \mathbf{N}$.

Lemma 3.1.3. For every $d \in \mathbb{N}$, $x \in (x_1, \dots, x_d) \in \mathbb{R}^d$ let $\lfloor x \rfloor \in \mathbb{R}$ satisfy $\lfloor x \rfloor = \max_{i \in \{1, \dots, d\}} |x_i|$. Then it holds for all $\Phi \in \mathbf{N}$ that $\max\{\mathcal{L}(\Phi), \lfloor \mathcal{D}(\Phi) \rfloor\} \leq \mathcal{P}(\Phi)$ and $\mathcal{P}(\Phi) \leq 2\mathcal{L}(\Phi) \lfloor \mathcal{D}(\Phi) \rfloor^2$.

Proof of Lemma 3.1.3. First, note that [1, Lemma 2.4] demonstrates for all $\Phi \in \mathbf{N}$ that $\mathcal{L}(\Phi) + \lfloor \mathcal{D}(\Phi) \rfloor \leq \mathcal{P}(\Phi)$. Hence, it holds for all $\Phi \in \mathbf{N}$ that $\max\{\mathcal{L}(\Phi), \lfloor \mathcal{D}(\Phi) \rfloor\} \leq \mathcal{P}(\Phi)$. Moreover, observe that for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi \in (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$ it holds that

$$\mathcal{P}(\Phi) = \sum_{k=1}^L l_k(l_{k-1} + 1) \leq \sum_{k=1}^L \lfloor \mathcal{D}(\Phi) \rfloor (\lfloor \mathcal{D}(\Phi) \rfloor + 1) \leq 2L \lfloor \mathcal{D}(\Phi) \rfloor^2 = 2\mathcal{L}(\Phi) \lfloor \mathcal{D}(\Phi) \rfloor^2. \quad (3.3)$$

The proof of Lemma 3.1.3 is thus complete. $\square$

Definition 3.1.4 (Multidimensional version). Let $d \in \mathbb{N}$, $\phi \in C(\mathbb{R}, \mathbb{R})$. Then we denote by $\mathfrak{M}_{\phi, d}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ the function which satisfies for all $x = (x_1, x_2, \dots, x_d) \in \mathbb{R}^d$ that

$$\mathfrak{M}_{\phi, d}(x) = (\phi(x_1), \phi(x_2), \dots, \phi(x_d)). \quad (3.4)$$

Definition 3.1.5 (Realization associated to an ANN). Let $\phi \in C(\mathbb{R}, \mathbb{R})$. Then we denote by $\mathcal{R}_{\phi}: \mathbf{N} \rightarrow \cup_{k, l \in \mathbb{N}} C(\mathbb{R}^k, \mathbb{R}^l)$ the function which satisfies for all $L \in \mathbb{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi = ((W_1, B_1), (W_2, B_2), \dots, (W_L, B_L)) \in (\times_{k=1}^L (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k}))$, $x_0 \in \mathbb{R}^{l_0}$, $x_1 \in \mathbb{R}^{l_1}$, $\dots$, $x_{L-1} \in \mathbb{R}^{l_{L-1}}$ with $x_k = \mathfrak{M}_{\phi, l_k}(W_k x_{k-1} + B_k)$, $k \in \mathbb{N}_0 \cap [1, L-1]$, that

$$\mathcal{R}_{\phi}(\Phi) \in C(\mathbb{R}^{l_0}, \mathbb{R}^{l_L}), \quad \text{and} \quad (\mathcal{R}_{\phi}(\Phi))(x_0) = W_L x_{L-1} + B_L. \quad (3.5)$$

Before we continue with the presentation of ANN constructions for, e.g., composition, parallelization, and linear combination of ANNs, we briefly explain in words the mathematical objects introduced in the present section. The set $\mathbf{N}$ denotes the set of all ANNs we consider as approximators in the remainder of this thesis. For an ANN $\Phi \in \mathbf{N}$, the tuple of natural numbers $\mathcal{D}(\Phi) \in \cup_{l=2}^{\infty} \mathbb{N}^l$ denotes the layer structure of the ANN Φ . The natural numbers $\mathcal{I}(\Phi), \mathcal{O}(\Phi) \in \mathbb{N}$ denote the input and output dimensions of the ANN $\Phi \in \mathbf{N}$. The natural number $\mathcal{L}(\Phi)$ corresponds to the depth of the ANN $\Phi \in \mathbf{N}$ while the number $\lfloor \mathcal{D}(\Phi) \rfloor$, where $\lfloor \cdot \rfloor$ is defined as in Lemma 3.1.3, denotes the maximum width of the ANN $\Phi \in \mathbf{N}$. The natural number $\mathcal{P}(\Phi) \in \mathbb{N}$ denotes the number of parameters that describe the ANN $\Phi \in \mathbf{N}$. Roughly speaking, this number corresponds to the size of the ANN Φ . For every $\phi \in C(\mathbb{R}, \mathbb{R})$ and $d \in \mathbb{N}$ the function $\mathfrak{M}_{\phi, d}: \mathbb{R}^d \rightarrow \mathbb{R}^d$ corresponds to the multidimensional version of the one dimensional function ϕ . For every ANN $\Phi \in \mathbf{N}$ and $\phi \in C(\mathbb{R}, \mathbb{R})$ the function $\mathcal{R}_{\phi}(\Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$ denotes the realization function of the ANN Φ with activation function ϕ .

3.2. Compositions, Affine Linear Transformations, Powers, and Extensions of ANNs

Definition 3.2.1 (Compositions of ANNs). We denote by $\bullet: \{(\Phi_1, \Phi_2) \in \mathbf{N} \times \mathbf{N} : \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)\} \rightarrow \mathbf{N}$ the function which satisfies for all $K, L \in \mathbb{N}$, $k_0, k_1, \dots, k_K, l_0, l_1, \dots, l_L \in \mathbb{N}$, $\Phi_1 = ((W_1, B_1), \dots, (W_K, B_K)) \in (\times_{j=1}^K (\mathbb{R}^{k_j \times k_{j-1}} \times \mathbb{R}^{k_j}))$, $\Phi_2 = ((\mathcal{W}_1, \mathcal{B}_1), \dots, (\mathcal{W}_L, \mathcal{B}_L)) \in$

$(\times_{j=1}^L (\mathbb{R}^{l_j \times l_{j-1}} \times \mathbb{R}^{l_j}))$ with $k_0 = \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2) = l_L$ that

$$\Phi_1 \bullet \Phi_2 = \begin{cases} ((W_1 \mathcal{W}_1, W_1 \mathcal{B}_1 + B_1)) & : K = 1 = L \\ ((W_1 \mathcal{W}_1, W_1 \mathcal{B}_1 + B_1), (W_2, B_2), \dots, (W_K, B_K)) & : K > 1 = L \\ ((\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2), \dots, (\mathcal{W}_{L-1}, \mathcal{B}_{L-1}), (W_1 \mathcal{W}_L, W_1 \mathcal{B}_L + B_1)) & : K = 1 < L \\ ((\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2), \dots, (\mathcal{W}_{L-1}, \mathcal{B}_{L-1}), (W_1 \mathcal{W}_L, W_1 \mathcal{B}_L + B_1), (W_2, B_2), \dots, (W_K, B_K)) & : K > 1 < L. \end{cases} \quad (3.6)$$

Lemma 3.2.2 (Compositions of ANNs). *Let $K, L \in \mathbb{N}$, $k_0, k_1, \dots, k_K, l_0, l_1, \dots, l_L \in \mathbb{N}$, let $\Phi_1 = ((W_1, B_1), \dots, (W_K, B_K)) \in (\times_{j=1}^K (\mathbb{R}^{k_j \times k_{j-1}} \times \mathbb{R}^{k_j}))$, $\Phi_2 = ((\mathcal{W}_1, \mathcal{B}_1), \dots, (\mathcal{W}_L, \mathcal{B}_L)) \in (\times_{j=1}^L (\mathbb{R}^{l_j \times l_{j-1}} \times \mathbb{R}^{l_j}))$ satisfy $k_0 = \mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2) = l_L$. Then*

- (i) *it holds that $\mathcal{D}(\Phi_1 \bullet \Phi_2) = (l_0, \dots, l_{L-1}, k_1, \dots, k_K) \in \mathbb{N}^{L+K}$,*
- (ii) *it holds that $\mathcal{L}(\Phi_1 \bullet \Phi_2) = \mathcal{L}(\Phi_1) + \mathcal{L}(\Phi_2) - 1$,*
- (iii) *it holds that $\mathcal{H}(\Phi_1 \bullet \Phi_2) = \mathcal{H}(\Phi_1) + \mathcal{H}(\Phi_2)$,*
- (iv) *it holds that*

$$\begin{aligned} \mathcal{P}(\Phi_1 \bullet \Phi_2) &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + k_1(l_{L-1} + 1) - k_1(k_0 + 1) - l_L(l_{L-1} + 1) \\ &= \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + k_1 l_{L-1} - k_1 k_0 - l_L l_{L-1} - l_L \\ &\leq \mathcal{P}(\Phi_1) + \mathcal{P}(\Phi_2) + k_1 l_{L-1}, \end{aligned} \quad (3.7)$$

and

- (v) *it holds for all $\phi \in C(\mathbb{R}, \mathbb{R})$ that $\mathcal{R}_\phi(\Phi_1 \bullet \Phi_2) \in C(\mathbb{R}^{\mathcal{I}(\Phi_2)}, \mathbb{R}^{\mathcal{O}(\Phi_1)})$ and $\mathcal{R}_\phi(\Phi_1 \bullet \Phi_2) = \mathcal{R}_\phi(\Phi_1) \circ \mathcal{R}_\phi(\Phi_2)$.*

Lemma 3.2.3 (Associativity of compositions of ANNs). *Let $\Phi_1, \Phi_2, \Phi_3 \in \mathbf{N}$ satisfy that $\mathcal{I}(\Phi_1) = \mathcal{O}(\Phi_2)$ and $\mathcal{I}(\Phi_2) = \mathcal{O}(\Phi_3)$. Then it holds that*

$$(\Phi_1 \bullet \Phi_2) \bullet \Phi_3 = \Phi_1 \bullet (\Phi_2 \bullet \Phi_3). \quad (3.8)$$

Definition 3.2.4 (Identity matrix). Let $d \in \mathbb{N}$. Then we denote by $\mathbf{I}_d \in \mathbb{R}^{d \times d}$ the identity matrix in $\mathbb{R}^{d \times d}$.

Definition 3.2.5 (Powers of ANNs). We denote by $(\cdot)^{\bullet n} : \{\Phi \in \mathbf{N} : \mathcal{I}(\Phi) = \mathcal{O}(\Phi)\} \rightarrow \mathbf{N}$, $n \in \mathbb{N}_0$, the functions which satisfy for all $n \in \mathbb{N}_0$, $\Phi \in \mathbf{N}$ with $\mathcal{I}(\Phi) = \mathcal{O}(\Phi)$ that

$$\Phi^{\bullet n} = \begin{cases} (\mathbf{I}_{\mathcal{O}(\Phi)}, (0, 0, \dots, 0)) \in (\mathbb{R}^{\mathcal{O}(\Phi) \times \mathcal{O}(\Phi)} \times \mathbb{R}^{\mathcal{O}(\Phi)}) & : n = 0 \\ \Phi \bullet (\Phi^{\bullet(n-1)}) & : n \in \mathbb{N}. \end{cases} \quad (3.9)$$

Definition 3.2.6 (Extensions of ANNs). Let $L \in \mathbb{N}$, let $\Psi \in \mathbf{N}$ with $\mathcal{I}(\Psi) = \mathcal{O}(\Psi)$. Then we denote by $\mathcal{E}_{L, \Psi} : \{\Phi \in \mathbf{N} : (\mathcal{L}(\Phi) \leq L \text{ and } \mathcal{O}(\Phi) = \mathcal{I}(\Psi))\} \rightarrow \mathbf{N}$ satisfy for all $\Phi \in \mathbf{N}$ with $\mathcal{L}(\Phi) \leq L$ and $\mathcal{I}(\Psi) = \mathcal{O}(\Phi)$ that $\mathcal{E}_{L, \Psi}(\Phi) = (\Psi^{\bullet(L-\mathcal{L}(\Phi))}) \bullet \Phi$.

Lemma 3.2.7 (Certain powers and extensions of ANNs). *Let $d, l \in \mathbb{N}$, let $\Psi \in \mathbf{N}$ satisfy $\mathcal{D}(\Psi) = (d, l, d) \in \mathbb{N}^3$. Then*

- (i) *it holds for all $n \in \mathbb{N}_0$ that $\mathcal{L}(\Psi^{\bullet n}) = n + 1$, $\mathcal{D}(\Psi^{\bullet n}) \in \mathbb{N}^{n+2}$, and*

$$\mathcal{D}(\Psi^{\bullet n}) = \begin{cases} (d, d) & : n = 0 \\ (d, l, l, \dots, l, d) & : n \in \mathbb{N}, \end{cases} \quad (3.10)$$

(ii) it holds for all $\Phi \in \mathbf{N}$, $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$ with $\mathcal{O}(\Phi) = d$ that $\mathcal{L}(\mathcal{E}_{L,\Psi}(\Phi)) = L$ and

$$\mathcal{P}(\mathcal{E}_{L,\Psi}(\Phi)) \leq \begin{cases} \mathcal{P}(\Phi) & : \mathcal{L}(\Phi) = L \\ \left[\left(\max\{1, \frac{l}{d}\} \right) \mathcal{P}(\Phi) + ((L - \mathcal{L}(\Phi) - 1)l + d)(l + 1) \right] & : \mathcal{L}(\Phi) < L, \end{cases} \quad (3.11)$$

and

(iii) it holds for all $\Phi \in \mathbf{N}$ with $\mathcal{O}(\Phi) = d$ that $\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi) = \Phi$.

Proof of Lemma 3.2.7. First, note that [41, Lemma 2.13] proves items (i) and (ii). Next, observe that it holds that $\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi) = (\Psi^{\bullet(\mathcal{L}(\Psi)-\mathcal{L}(\Phi))}) \bullet \Phi = \Psi^{\bullet 0} \bullet \Phi$. This, the assumption that $\mathcal{O}(\Phi) = \mathcal{I}(\Psi) = d$, the fact that $\Psi^{\bullet 0} = (\text{Id}, (0, 0, \dots, 0)) \in \mathbb{R}^{d \times d} \times \mathbb{R}^d$, and Lemma 3.2.2 demonstrate that $\mathcal{E}_{\mathcal{L}(\Phi),\Psi}(\Phi) = \Psi^{\bullet 0} \bullet \Phi = \Phi$. This proves item (iii). The proof of Lemma 3.2.7 is thus complete. $\square$

Lemma 3.2.8 (Extensions of ANNs involving identities). *Let $\phi \in C(\mathbb{R}, \mathbb{R})$, let $\mathbb{I} \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}^{\mathcal{I}(\mathbb{I})}$ that $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\mathbb{I})$ and $(\mathcal{R}_\phi(\mathbb{I}))(x) = x$. Then*

(i) it holds for all $n \in \mathbb{N}_0$, $x \in \mathbb{R}^{\mathcal{I}(\mathbb{I})}$ that $\mathcal{R}_\phi(\mathbb{I}^{\bullet n}) \in C(\mathbb{R}^{\mathcal{I}(\mathbb{I})}, \mathbb{R}^{\mathcal{I}(\mathbb{I})})$ and $(\mathcal{R}_\phi(\mathbb{I}^{\bullet n}))(x) = x$ and

(ii) it holds for all $\Phi \in \mathbf{N}$, $L \in \mathbb{N} \cap [\mathcal{L}(\Phi), \infty)$, $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ with $\mathcal{I}(\mathbb{I}) = \mathcal{O}(\Phi)$ that $\mathcal{R}_\phi(\mathcal{E}_{L,\mathbb{I}}(\Phi)) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$ and $(\mathcal{R}_\phi(\mathcal{E}_{L,\mathbb{I}}(\Phi)))(x) = (\mathcal{R}_\phi(\Phi))(x)$.

3.3. Parallelizations of ANNs

Definition 3.3.1 (Parallelizations of ANNs with the same depth). Let $n \in \mathbb{N}$. Then we denote by $\mathbf{P}_n: \{(\Phi_1, \dots, \Phi_n) \in \mathbf{N}^n : \mathcal{L}(\Phi_1) = \mathcal{L}(\Phi_2) = \dots = \mathcal{L}(\Phi_n)\} \rightarrow \mathbf{N}$ the function which satisfies for all $L \in \mathbb{N}$, $(l_{1,0}, l_{1,1}, \dots, l_{1,L}), \dots, (l_{n,0}, l_{n,1}, \dots, l_{n,L}) \in \mathbb{N}^{L+1}$,

$$\begin{aligned} \Phi_1 &= ((W_{1,1}, B_{1,1}), (W_{1,2}, B_{1,2}), \dots, (W_{1,L}, B_{1,L})) \in (\times_{k=1}^L (\mathbb{R}^{l_{1,k} \times l_{1,k-1}} \times \mathbb{R}^{l_{1,k}})), \\ \Phi_2 &= ((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \dots, (W_{2,L}, B_{2,L})) \in (\times_{k=1}^L (\mathbb{R}^{l_{2,k} \times l_{2,k-1}} \times \mathbb{R}^{l_{2,k}})), \\ &\vdots \end{aligned} \quad (3.12)$$

$$\Phi_n = ((W_{n,1}, B_{n,1}), (W_{n,2}, B_{n,2}), \dots, (W_{n,L}, B_{n,L})) \in (\times_{k=1}^L (\mathbb{R}^{l_{n,k} \times l_{n,k-1}} \times \mathbb{R}^{l_{n,k}})),$$

$((\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2), \dots, (\mathcal{W}_L, \mathcal{B}_L)) \in (\times_{j=1}^L (\mathbb{R}^{(\sum_{k=1}^n l_{k,j}) \times (\sum_{k=1}^n l_{k,j-1})} \times \mathbb{R}^{(\sum_{k=1}^n l_{k,j})}))$ with

$$\forall i \in \mathbb{N} \cap [1, L] : (\mathcal{W}_i, \mathcal{B}_i) = \left(\begin{pmatrix} W_{1,i} & 0 & 0 & \dots & 0 \\ 0 & W_{2,i} & 0 & \dots & 0 \\ 0 & 0 & W_{3,i} & \dots & 0 \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & 0 & \dots & W_{n,i} \end{pmatrix}, \begin{pmatrix} B_{1,i} \\ B_{2,i} \\ B_{3,i} \\ \vdots \\ B_{n,i} \end{pmatrix} \right), \quad (3.13)$$

that

$$\mathbf{P}_n(\Phi_1, \Phi_2, \dots, \Phi_n) = ((\mathcal{W}_1, \mathcal{B}_1), (\mathcal{W}_2, \mathcal{B}_2), \dots, (\mathcal{W}_L, \mathcal{B}_L)). \quad (3.14)$$

Lemma 3.3.2. *Let $n, L \in \mathbb{N}$, let $(l_{1,0}, l_{1,1}, \dots, l_{1,L}), \dots, (l_{n,0}, l_{n,1}, \dots, l_{n,L}) \in \mathbb{N}^{L+1}$, and let*

$$\begin{aligned} \Phi_1 &= ((W_{1,1}, B_{1,1}), (W_{1,2}, B_{1,2}), \dots, (W_{1,L}, B_{1,L})) \in (\times_{k=1}^L (\mathbb{R}^{l_{1,k} \times l_{1,k-1}} \times \mathbb{R}^{l_{1,k}})), \\ \Phi_2 &= ((W_{2,1}, B_{2,1}), (W_{2,2}, B_{2,2}), \dots, (W_{2,L}, B_{2,L})) \in (\times_{k=1}^L (\mathbb{R}^{l_{2,k} \times l_{2,k-1}} \times \mathbb{R}^{l_{2,k}})), \\ &\vdots \\ \Phi_n &= ((W_{n,1}, B_{n,1}), (W_{n,2}, B_{n,2}), \dots, (W_{n,L}, B_{n,L})) \in (\times_{k=1}^L (\mathbb{R}^{l_{n,k} \times l_{n,k-1}} \times \mathbb{R}^{l_{n,k}})). \end{aligned} \quad (3.15)$$

Then

(i) it holds that

$$\mathbf{P}_n(\Phi_1, \dots, \Phi_n) \in \left(\times_{j=1}^L (\mathbb{R}(\sum_{k=1}^n l_{k,j}) \times (\sum_{k=1}^n l_{k,j-1}) \times \mathbb{R}(\sum_{k=1}^n l_{k,j})) \right) \subseteq \mathbf{N}, \quad (3.16)$$

(ii) it holds that

$$\mathcal{R}_\phi(\mathbf{P}_n(\Phi_1, \dots, \Phi_n)) \in C(\mathbb{R}(\sum_{k=1}^n \mathcal{I}(\Phi_k)), \mathbb{R}(\sum_{k=1}^n \mathcal{O}(\Phi_k))), \quad (3.17)$$

(iii) it holds for all $x_1 \in \mathbb{R}^{\mathcal{I}(\Phi_1)}$, $x_2 \in \mathbb{R}^{\mathcal{I}(\Phi_2)}$, $\dots$, $x_n \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ that

$$\begin{aligned} & (\mathcal{R}_\phi(\mathbf{P}_n(\Phi_1, \dots, \Phi_n)))(x_1, x_2, \dots, x_n) \\ &= ((\mathcal{R}_\phi(\Phi_1))(x_1), (\mathcal{R}_\phi(\Phi_2))(x_2), \dots, (\mathcal{R}_\phi(\Phi_n))(x_n)) \in \mathbb{R}(\sum_{k=1}^n \mathcal{O}(\Phi_k)), \end{aligned} \quad (3.18)$$

and

(iv) it holds that

$$\mathcal{D}(\mathbf{P}_n(\Phi_1, \dots, \Phi_n)) = \left(\sum_{k=1}^n l_{k,0}, \sum_{k=1}^n l_{k,1}, \dots, \sum_{k=1}^n l_{k,L} \right). \quad (3.19)$$

Definition 3.3.3 (Parallelizations of ANNs with different layer structure). Let $n \in \mathbb{N}$, $\Psi = (\Psi_1, \Psi_2, \dots, \Psi_n) \in \mathbf{N}^n$ satisfy for all $i \in \mathbb{N} \cap [1, n]$ that $\mathcal{L}(\Psi_i) = 2$, $\mathcal{I}(\Psi_i) = \mathcal{O}(\Psi_i)$. Then we denote by $\mathbf{P}_{n,\Psi}: \{(\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n : \mathcal{I}(\Psi_i) = \mathcal{O}(\Phi_i), i \in \mathbb{N} \cap [1, n]\} \rightarrow \mathbf{N}$ the function which satisfies for all $\Phi = (\Phi_1, \dots, \Phi_n) \in \mathbf{N}^n$ with $\mathcal{I}(\Psi_i) = \mathcal{O}(\Phi_i)$, $i \in \mathbb{N} \cap [1, n]$, that

$$\mathbf{P}_{n,\Psi}(\Phi) = \mathbf{P}_n(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_1}(\Phi_1), \dots, \mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_n}(\Phi_n)). \quad (3.20)$$

Corollary 3.3.4. Let $n \in \mathbb{N}$, $\phi \in C(\mathbb{R}, \mathbb{R})$, $\mathbb{I} = (\mathbb{I}_1, \mathbb{I}_2, \dots, \mathbb{I}_n)$, $\Phi = (\Phi_1, \Phi_2, \dots, \Phi_n) \in \mathbf{N}^n$ satisfy for all $i \in \mathbb{N} \cap [1, n]$, $x \in \mathbb{R}^{\mathcal{O}(\Phi_i)}$ that $\mathcal{L}(\mathbb{I}_i) = 2$, $\mathcal{I}(\mathbb{I}_i) = \mathcal{O}(\mathbb{I}_i) = \mathcal{O}(\Phi_i)$, and $(\mathcal{R}_\phi(\mathbb{I}_i))(x) = x$. Then

(i) it holds that $\mathcal{R}_\phi(\mathbf{P}_{n,\mathbb{I}}(\Phi)) \in C(\mathbb{R}(\sum_{i=1}^n \mathcal{I}(\Phi_i)), \mathbb{R}(\sum_{i=1}^n \mathcal{O}(\Phi_i)))$ and

(ii) it holds for all $x_1 \in \mathbb{R}^{\mathcal{I}(\Phi_1)}$, $\dots$, $x_n \in \mathbb{R}^{\mathcal{I}(\Phi_n)}$ that

$$(\mathcal{R}_\phi(\mathbf{P}_{n,\mathbb{I}}(\Phi)))(x_1, x_2, \dots, x_n) = \begin{pmatrix} (\mathcal{R}_\phi(\Phi_1))(x_1) \\ (\mathcal{R}_\phi(\Phi_2))(x_2) \\ \vdots \\ (\mathcal{R}_\phi(\Phi_n))(x_n) \end{pmatrix} \in \mathbb{R}(\sum_{i=1}^n \mathcal{O}(\Phi_i)), \quad (3.21)$$

and

(iii) it holds that

$$\begin{aligned} \mathcal{D}(\mathbf{P}_{n,\mathbb{I}}(\Phi)) &= \left(\sum_{k=1}^n \mathbb{D}_0(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_k}(\Phi_k)), \sum_{k=1}^n \mathbb{D}_1(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_k}(\Phi_k)), \dots \right. \\ &\quad \left. \dots, \sum_{k=1}^n \mathbb{D}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i)}(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_k}(\Phi_k)) \right). \end{aligned} \quad (3.22)$$

Proof of Corollary 3.3.4. First, note that [41, Corollary 2.23] proves items (i) and (ii). Moreover, observe that the assumption that for all $i \in \mathbb{N} \cap [1, n]$ it holds that $\mathcal{L}(\mathbb{I}_i) = 2$, the assumption that for all $i \in \mathbb{N} \cap [1, n]$ it holds that $\mathcal{I}(\mathbb{I}_i) = \mathcal{O}(\mathbb{I}_i) = \mathcal{O}(\Phi_i)$, and item (ii) in Lemma 3.2.7

ensure for all $k \in \mathbb{N} \cap [1, n]$ it holds that $\mathcal{L}(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_k}(\Phi_k)) = \max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i)$. Combining this, (3.2), (3.20), and item (iv) in Lemma 3.3.2 yields that

$$\begin{aligned} \mathcal{D}(\mathbb{P}_{n, \mathbb{I}}(\Phi)) &= \left(\sum_{k=1}^n \mathbb{D}_0(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_k}(\Phi_k)), \sum_{k=1}^n \mathbb{D}_1(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_k}(\Phi_k)), \dots \right. \\ &\quad \left. \dots, \sum_{k=1}^n \mathbb{D}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i)}(\mathcal{E}_{\max_{i \in \mathbb{N} \cap [1, n]} \mathcal{L}(\Phi_i), \Psi_k}(\Phi_k)) \right). \end{aligned} \quad (3.23)$$

This proves item (iii). The proof of Corollary 3.3.4 is thus complete. $\square$

3.4. Linear Combinations of ANNs

Definition 3.4.1. Let $m, n \in \mathbb{N}$. Then we denote by $\mathfrak{S}_{m, n} \in \mathbb{R}^{m \times (mn)} \times \mathbb{R}^m$, $\mathfrak{T}_{m, n} \in \mathbb{R}^{(mn) \times m} \times \mathbb{R}^{mn}$ the ANNs given by

$$\mathfrak{S}_{m, n} = \left(\begin{pmatrix} \mathbf{I}_m & \mathbf{I}_m & \dots & \mathbf{I}_m \end{pmatrix}, 0 \right) \quad \text{and} \quad \mathfrak{T}_{m, n} = \left(\begin{pmatrix} \mathbf{I}_m \\ \mathbf{I}_m \\ \vdots \\ \mathbf{I}_m \end{pmatrix}, 0 \right). \quad (3.24)$$

Definition 3.4.2 (Sums of ANNs with the same length). Let $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, $\Phi_u, \Phi_{u+1}, \dots, \Phi_v \in \mathbf{N}$ satisfy for all $i \in \mathbb{Z} \cap [u, v]$ that $\mathcal{L}(\Phi_i) = \mathcal{L}(\Phi_u)$, $\mathcal{I}(\Phi_i) = \mathcal{I}(\Phi_u)$, $\mathcal{O}(\Phi_i) = \mathcal{O}(\Phi_u)$. Then we denote by $\bigoplus_{i=u}^v \Phi_i$ the ANN given by

$$\bigoplus_{i=u}^v \Phi_i = \left(\mathfrak{S}_{\mathcal{O}(\Phi_u), v-u+1} \bullet [\mathbf{P}_{v-u+1}(\Phi_u, \Phi_{u+1}, \dots, \Phi_v)] \bullet \mathfrak{T}_{\mathcal{I}(\Phi_u), v-u+1} \right) \in \mathbf{N}. \quad (3.25)$$

Definition 3.4.3 (Sums of ANNs with different lengths). Let $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, $\Phi_u, \Phi_{u+1}, \dots, \Phi_v, \Psi \in \mathbf{N}$ satisfy for all $i \in \mathbb{Z} \cap [u, v]$ that $\mathcal{I}(\Phi_i) = \mathcal{I}(\Phi_u)$, $\mathcal{O}(\Phi_i) = \mathcal{I}(\Psi) = \mathcal{O}(\Psi)$, and $\mathcal{L}(\Psi) = 2$. Then denote by $\boxplus_{i=u, \Psi}^v \Phi_i$ the ANN given by

$$\boxplus_{i=u, \Psi}^v \Phi_i = \bigoplus_{i=u}^v \mathcal{E}_{\max_{j \in \mathbb{Z} \cap [u, v]} \mathcal{L}(\Phi_j), \Psi}(\Phi_i). \quad (3.26)$$

Definition 3.4.4 (Scalar multiplications of ANNs). We denote by $\otimes: \mathbb{R} \times \mathbf{N} \rightarrow \mathbf{N}$ the function which satisfies for all $\alpha \in \mathbb{R}$, $\Phi \in \mathbf{N}$ that

$$\alpha \otimes \Phi = (\alpha \mathbf{I}_{\mathcal{O}(\Phi)}, 0) \bullet \Phi. \quad (3.27)$$

Lemma 3.4.5. Let $\alpha \in \mathbb{R}$, $\Phi \in \mathbf{N}$. Then

- (i) it holds that $\mathcal{D}(\alpha \otimes \Phi) = \mathcal{D}(\Phi)$,
- (ii) it holds that $\mathcal{R}_\phi(\alpha \otimes \Phi) \in C(\mathbb{R}^{\mathcal{I}(\Phi)}, \mathbb{R}^{\mathcal{O}(\Phi)})$, and
- (iii) it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi)}$ that $(\mathcal{R}_\phi(\alpha \otimes \Phi))(x) = \alpha(\mathcal{R}_\phi(\Phi))(x)$.

Lemma 3.4.6 (Linear combinations of ANNs with the same length). Let $u \in \mathbb{Z}$, $v \in \mathbb{Z} \cap [u, \infty)$, $h_u, h_{u+1}, \dots, h_v \in \mathbb{R}$, $\phi \in C(\mathbb{R}, \mathbb{R})$, $\Phi_u, \Phi_{u+1}, \dots, \Phi_v, \Psi \in \mathbf{N}$, satisfy for all $i \in \mathbb{Z} \cap [u, v]$ that $\mathcal{I}(\Phi_i) = \mathcal{I}(\Phi_u)$, $\mathcal{O}(\Phi_i) = \mathcal{O}(\Phi_u)$, $\mathcal{L}(\Phi_i) = \mathcal{L}(\Phi_u)$ and $\Psi = \bigoplus_{i=u}^v (h_i \otimes \Phi_i)$. Then

- (i) it holds that

$$\mathcal{D}(\Psi) = \left(\mathcal{I}(\Phi_u), \sum_{i=u}^v \mathbb{D}_1(\Phi_i), \dots, \sum_{i=u}^v \mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_i), \mathcal{O}(\Phi_u) \right), \quad (3.28)$$

(ii) it holds that $\mathcal{R}_\phi(\Psi) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$, and

(iii) it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ that

$$(\mathcal{R}_\phi(\Psi))(x) = \sum_{i=u}^v h_i(\mathcal{R}_\phi(\Phi_i))(x). \quad (3.29)$$

Proof of Lemma 3.4.6. First, note that Lemma 3.4.5 ensures for all $i \in \mathbb{Z} \cap [u, v]$ that $\mathcal{D}(h_i \otimes \Phi_i) = \mathcal{D}(\Phi_i)$. This, the fact that $\mathcal{D}(\mathfrak{S}_{\mathcal{O}(\Phi_u), v-u+1}) = ((v-u+1)\mathcal{O}(\Phi_u), \mathcal{O}(\Phi_u))$, the fact that $\mathcal{D}(\mathfrak{T}_{\mathcal{I}(\Phi_u), v-u+1}) = (\mathcal{I}(\Phi_u), (v-u+1)\mathcal{I}(\Phi_u))$, Lemma 3.2.2, and item (iv) in Lemma 3.3.2 yield

$$\mathcal{D}(\Psi) = \left(\mathcal{I}(\Phi_u), \sum_{i=u}^v \mathbb{D}_1(\Phi_i), \sum_{i=u}^v \mathbb{D}_2(\Phi_i), \dots, \sum_{i=u}^v \mathbb{D}_{\mathcal{L}(\Phi_u)-1}(\Phi_i), \mathcal{O}(\Phi_u) \right). \quad (3.30)$$

This proves item (i). Moreover, observe that Lemma 3.2.2 and item (iii) in Lemma 3.4.5 establish for all $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ that

$$\begin{aligned} (\mathcal{R}_\phi(\Psi))(x) &= (\mathcal{R}_\phi(\mathfrak{S}_{\mathcal{O}(\Phi_u), v-u+1}) \circ \mathbf{P}_{v-u+1}(h_u \otimes \Phi_u, h_{u+1} \otimes \Phi_{u+1}, \dots, h_v \otimes \Phi_v) \circ \mathfrak{T}_{\mathcal{I}(\Phi_u), v-u+1})(x) \\ &= (\mathcal{R}_\phi(\mathfrak{S}_{\mathcal{O}(\Phi_u), v-u+1}) \circ \mathbf{P}_{v-u+1}(h_u \otimes \Phi_u, h_{u+1} \otimes \Phi_{u+1}, \dots, h_v \otimes \Phi_v))(x, x, \dots, x) \\ &= (\mathcal{R}_\phi(\mathfrak{S}_{\mathcal{O}(\Phi_u), v-u+1})) \begin{pmatrix} (\mathcal{R}_\phi(h_u \otimes \Phi_u))(x) \\ (\mathcal{R}_\phi(h_{u+1} \otimes \Phi_{u+1}))(x) \\ \vdots \\ (\mathcal{R}_\phi(h_v \otimes \Phi_v))(x) \end{pmatrix} = \sum_{i=u}^v h_i(\mathcal{R}_\phi(\Phi_i))(x). \end{aligned} \quad (3.31)$$

This proves item (ii) and item (iii). The proof of Lemma 3.4.6 is thus complete. $\square$

Lemma 3.4.7 (Linear combinations of ANNs with different lengths). *Let $L \in \mathbb{N}$, $\phi \in C(\mathbb{R}, \mathbb{R})$, $u \in \mathbb{Z}$, $v \in [u, \infty) \cup \mathbb{Z}$, $h_u, h_{u+1}, \dots, h_v \in \mathbb{R}$, $\Phi_u, \Phi_{u+1}, \dots, \Phi_v, \Psi, \mathfrak{J} \in \mathbf{N}$ satisfy for all $x \in \mathbb{R}^{\mathcal{I}(\mathfrak{J})}$, $i \in \mathbb{Z} \cap [u, v]$ that $\mathcal{I}(\Phi_i) = \mathcal{I}(\Phi_u)$, $\mathcal{O}(\Phi_i) = \mathcal{I}(\mathfrak{J}) = \mathcal{O}(\mathfrak{J})$, $L = \max_{j \in \mathbb{Z} \cap [u, v]} \mathcal{L}(\Phi_j)$, $\mathcal{L}(\mathfrak{J}) = 2$, $(\mathcal{R}_\phi(\mathfrak{J}))(x) = x$, and $\Psi = \boxplus_{i=u, \mathfrak{J}}^v (h_i \otimes \Phi_i)$. Then*

(i) it holds that

$$\mathcal{D}(\Psi) = \left(\mathcal{I}(\Phi_u), \sum_{i=u}^v \mathbb{D}_1(\mathcal{E}_{L, \mathfrak{J}}(\Phi_i)), \dots, \sum_{i=u}^v \mathbb{D}_{L-1}(\mathcal{E}_{L, \mathfrak{J}}(\Phi_i)), \mathcal{O}(\Phi_u) \right), \quad (3.32)$$

(ii) it holds that $\mathcal{R}_\phi(\Psi) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$, and

(iii) it holds for all $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ that

$$(\mathcal{R}_\phi(\Psi))(x) = \sum_{i=u}^v h_i(\mathcal{R}_\phi(\Phi_i))(x). \quad (3.33)$$

Proof of Lemma 3.4.7. First, note that item (i) in Lemma 3.4.5 ensures for all $i \in \mathbb{Z} \cap [u, v]$ that $\mathcal{D}(h_i \otimes \Phi_i) = \mathcal{D}(\Phi_i)$. Combining this, the assumption that it holds that $\mathcal{L}(\mathfrak{J}) = 2$, the assumption that it holds that $\mathcal{I}(\mathfrak{J}) = \mathcal{O}(\mathfrak{J})$, the assumption that for all $x \in \mathbb{R}^{\mathcal{I}(\mathfrak{J})}$ it holds that $(\mathcal{R}_\phi(\mathfrak{J}))(x) = x$, and items (i) to (iii) in Lemma 3.2.7 yields for all $i \in \mathbb{Z} \cap [u, v]$ that $\mathcal{D}(\mathcal{E}_{L, \mathfrak{J}}(h_i \otimes \Phi_i)) = \mathcal{D}(\mathcal{E}_{L, \mathfrak{J}}(\Phi_i))$. This and item (i) in Lemma 3.4.6 demonstrate that

$$\begin{aligned} \mathcal{D}(\Psi) &= \mathcal{D}\left(\boxplus_{i=u, \mathfrak{J}}^v (h_i \otimes \Phi_i)\right) = \mathcal{D}\left(\bigoplus_{i=u}^v \mathcal{E}_{L, \mathfrak{J}}(h_i \otimes \Phi_i)\right) \\ &= (\mathcal{I}(\mathcal{E}_{L, \mathfrak{J}}(h_u \otimes \Phi_u)), \sum_{i=u}^v \mathbb{D}_1(\mathcal{E}_{L, \mathfrak{J}}(h_i \otimes \Phi_i)), \dots \\ &\quad \dots, \sum_{i=u}^v \mathbb{D}_{L-1}(\mathcal{E}_{L, \mathfrak{J}}(h_i \otimes \Phi_i)), \mathcal{O}(\mathcal{E}_{L, \mathfrak{J}}(h_u \otimes \Phi_u))) \\ &= (\mathcal{I}(\Phi_u), \sum_{i=u}^v \mathbb{D}_1(h_i \otimes \Phi_i), \dots, \sum_{i=u}^v \mathbb{D}_{L-1}(h_i \otimes \Phi_i), \mathcal{O}(\Phi_u)) \in \mathbb{N}^{L+2}. \end{aligned} \quad (3.34)$$

Hence, it holds that $\mathcal{R}_\phi(\Psi) \in C(\mathbb{R}^{\mathcal{I}(\Phi_u)}, \mathbb{R}^{\mathcal{O}(\Phi_u)})$. This proves items (i) and (ii). Furthermore, observe that item (iii) in Lemma 3.4.6, item (iii) in Lemma 3.4.5 assures for all $x \in \mathbb{R}^{\mathcal{I}(\Phi_u)}$ that

$$(\mathcal{R}_\phi(\Psi))(x) = \sum_{i=u}^v (\mathcal{R}_\phi(h_i \otimes \Phi_i))(x) = \sum_{i=u}^v h_i(\mathcal{R}_\phi(\Phi_i))(x). \quad (3.35)$$

This proves item (iii). The proof of Lemma 3.4.7 is thus complete. $\square$

Chapter 4.

Deep Neural Network Approximations for Bellman Equations of High-Dimensional Markov Decision Processes

This chapter is based on the joint work [56] with A. Jentzen and T. Kruse. Its structure has been adapted to this thesis. In particular, the general framework and calculus for ANNs used in the underlying article are collected separately in Chapter 3 and will be used throughout the present chapter.

In contrast to the general measurable state spaces that were considered in the analysis of the MLFP approximations in Chapter 2, the ANN approximation results below are formulated on finite-dimensional Euclidean state spaces. In particular, we consider families of functional fixed-point equations of the form (1.3) on $\mathbb{R}^d$, $d \in \mathbb{N}$.

Empirically, the use of DNNs for the approximation of value functions, Q -functions, and control strategies has demonstrated remarkable success in reinforcement learning, including high-dimensional tasks such as in game playing, robotics, and control (see, e.g., [73], [98], [68], and [62]). Apart from this widespread empirical success, there are several theoretical analyses that aim to explain the effectiveness of deep reinforcement learning (see, e.g., [33], [25], and [105]). Yet, a comprehensive theoretical understanding of deep reinforcement learning is still incomplete.

This chapter tries to address a specific theoretical aspect concerning approximation in this broader research area, that is, instead of analyzing neural network training procedures, we establish conditions under which the solutions of functional fixed-point equations can be approximated by ANNs with quantitative error and complexity estimates. To this end, the MLFP approximation scheme developed and analyzed in Chapter 2 as a variant of MLP approximations tailored to functional fixed-point equations of the form (1.3) is a natural starting point. By combining MLP approximation schemes with compositional and approximation properties of ANNs, it has been demonstrated that ANNs can overcome the curse of dimensionality in the numerical approximation of broad classes of PDEs (see, e.g., [1, 2, 28, 50, 76, 77, 78]). Historically, the strategy of combining compositional and approximation properties of DNNs with Monte Carlo methods to overcome the curse of dimensionality was first developed for linear PDEs (see, e.g., [16, 20, 39, 40, 58]) and subsequently used, for example, for two-player zero-sum stochastic differential games (see, e.g., [86]).

However, there is an additional difficulty in the present setting. To obtain ANN approximation results for the solution of a functional fixed-point equation, the nonlinearity and the one-step transitions may be approximated by ANNs themselves. The resulting MLFP scheme thus approximates the solution of a perturbed functional fixed-point equation rather than the solution of the fixed-point equation corresponding to the original problem inputs. Thus, it is necessary to analyze the effect of the perturbations of the nonlinearity and of the one-step transitions on the solutions of the corresponding fixed-point equations.

With this in mind, we first present stability results for solutions of functional fixed-point equations of the form (1.3) where the state spaces are given by $\mathbb{R}^d$, $d \in \mathbb{N}$. The resulting estimates control the difference between two fixed-point solutions in terms of the perturbations

in their nonlinearities and their transition kernels.

Subsequently, we establish ANN representations of the MLFP scheme under the assumption that the nonlinearities and one-step transitions entering the MLFP approximations admit suitable ANN representations. For the resulting ANNs that represent the MLFP scheme we derive explicit bounds on the number of parameters.

Combining the stability results and the representation of the MLFP scheme via ANNs with the MLFP error analysis established in Chapter 2 yields the desired approximation results under suitable assumptions.

To briefly sketch the contribution of this chapter, we now present Theorem 4.0.1, a special case of Theorem 4.3.1, the main result of this chapter. Below Theorem 4.0.1 we explain in words the statement of the result as well as the mathematical objects appearing in it.

Theorem 4.0.1. *Let $\nu \in \{0, 1\}$, $\beta \in [0, \infty) \setminus \{1\}$, let $\phi: \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}$ that $\phi(x) = \nu \ln(1 + \exp(x)) + (1 - \nu) \max\{x, \beta x\}$, for every $d, \mathfrak{d} \in \mathbb{N}$ let*

$$\mathbf{N}_{d, \mathfrak{d}} = \cup_{H \in \mathbb{N}} \cup_{(l_0, l_1, \dots, l_{H+1}) \in \{d\} \times \mathbb{N}^H \times \{\mathfrak{d}\}} (\times_{k=1}^{H+1} (\mathbb{R}^{l_k \times l_{k-1}} \times \mathbb{R}^{l_k})) \subseteq \mathbf{N}, \quad (4.1)$$

let A be a finite set, let $a \in A$, $\delta \in (0, 1)$, $\eta \in (\delta, \infty)$, $\mathfrak{c} \in [1, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, for every $d \in \mathbb{N}$ let $g_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ and $\xi_d: \Omega \rightarrow \mathbb{R}^d$ be measurable, for every $d \in \mathbb{N}$, $a \in A$ let $t_d^{(a)}: \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ be a measurable function which satisfies $\sup_{x \in \mathbb{R}^d} \mathbb{E}[\|t_d^{(a)}(x, \xi_d)\|] < \infty$, for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ let $\mathbf{t}_{d, \varepsilon}^{(a)} \in \mathbf{N}_{2d, d}$ satisfy $\mathcal{D}(\mathbf{t}_{d, \varepsilon}^{(a)}) = \mathcal{D}(\mathbf{t}_{d, \varepsilon}^{(a)})$, and for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ let $\mathbf{G}_{d, \varepsilon} \in \mathbf{N}_{d, |A|}$ satisfy for all $x, y, p \in \mathbb{R}^d$, $a \in A$ that

$$\varepsilon \|g_d(x)\| + \|(\mathcal{R}_\phi(\mathbf{G}_{d, \varepsilon}))(x) - g_d(x)\| + \|(\mathcal{R}_\phi(\mathbf{t}_{d, \varepsilon}^{(a)}))(x, p) - t_d^{(a)}(x, p)\| \leq \varepsilon \mathfrak{c} d^\varepsilon, \quad (4.2)$$

$$\mathfrak{c}^{-1} \|g_d(x) - g_d(y)\| + \eta \|t_d^{(a)}(x, p) - t_d^{(a)}(y, p)\| \leq \|x - y\|, \quad (4.3)$$

and $\mathcal{P}(\mathbf{G}_{d, \varepsilon}) + \mathcal{P}(\mathbf{t}_{d, \varepsilon}^{(a)}) \leq \mathfrak{c} d^\varepsilon \varepsilon^{-\mathfrak{c}}$. Then there exists $c \in \mathbb{R}$ such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ it holds that

- (i) *there exists a unique bounded measurable $Q: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that*

$$(Q(x))(a) = (g_d(x))(a) + \delta \mathbb{E} \left[\max_{b \in A} (Q(t_d^{(a)}(x, \xi_d)))(b) \right] \quad (4.4)$$

and

- (ii) *there exists $\mathbf{Q} \in \mathbf{N}_{d, |A|}$ which satisfies*

$$\left(\int_{\mathbb{R}^d} \|Q(x) - (\mathcal{R}_\phi(\mathbf{Q}))(x)\|^2 \mu(dx) \right)^{1/2} \leq \varepsilon \quad \text{and} \quad \mathcal{P}(\mathbf{Q}) \leq \mathfrak{c} d^c \varepsilon^{-c}. \quad (4.5)$$

In the following we briefly explain the mathematical objects appearing in Theorem 4.0.1 above.

The real numbers $\nu \in \{0, 1\}$, $\beta \in [0, \infty) \setminus \{1\}$ specify the choice of activation function $\phi: \mathbb{R} \rightarrow \mathbb{R}$. In particular, in the case $\nu = 1$ we have that ϕ is the softplus activation function. In the case $\nu = 0 = \beta$ we have that ϕ is the rectified linear unit (ReLU) activation function. In the case $\nu = 0$, $\beta \in (0, 1)$ we have that ϕ is the leaky ReLU activation function with leaky factor β .

The sets $\mathbf{N}_{d, \mathfrak{d}} \subseteq \mathbf{N}$, $d, \mathfrak{d} \in \mathbb{N}$, in (4.1) in Theorem 4.0.1 denote the sets of all ANNs we consider in the approximation of the solutions of the Bellman equations for the Q -function discussed. Note that for all $d, \mathfrak{d} \in \mathbb{N}$ and for every ANN $\Phi \in \mathbf{N}_{d, \mathfrak{d}}$ it holds that $\mathcal{I}(\Phi) = d$ and $\mathcal{O}(\Phi) = \mathfrak{d}$ correspond to the input and output dimensions of the ANN Φ .

The triple $(\Omega, \mathcal{F}, \mathbb{P})$ in Theorem 4.0.1 is the probability space on which all appearing random variables are defined. In Theorem 4.0.1 we consider for every $d \in \mathbb{N}$ an MDP with state space

$(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. The finite set A denotes the common control set for all elements of the sequence of MDPs. For every $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ the one-step transition of the controlled Markov chain of the MDP with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ by choosing control a in state x is given by

$$t_d^{(a)}(x, \xi_d): \Omega \rightarrow \mathbb{R}^d. \quad (4.6)$$

For every $d \in \mathbb{N}$ the function $g_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ in Theorem 4.0.1 denotes the reward function of the MDP with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$. We assume that for all $d \in \mathbb{N}$ the function g_d is uniformly bounded and that the sequence of these bounds grows at most polynomially in the state space dimension $d \in \mathbb{N}$. The real number $\delta \in (0, 1)$ is the common discount factor for all elements of the sequence of MDPs.

The real numbers $\eta \in (\delta, \infty)$ and $\mathfrak{c} \in [1, \infty)$ are constants that we utilize to formulate our regularity and approximation assumptions in Theorem 4.0.1. We assume there exist ANNs $\mathbf{G}_{d,\varepsilon} \in \mathbf{N}_{d,|A|}$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ the realization function $\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}): \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ of the ANN $\mathbf{G}_{d,\varepsilon}$ is a suitable approximation of the function g_d and that the number of parameters $\mathcal{P}(\mathbf{G}_{d,\varepsilon})$ of the approximating ANN $\mathbf{G}_{d,\varepsilon}$ is bounded by $cd^\mathfrak{c}\varepsilon^{-\mathfrak{c}}$. Furthermore, we assume there exist $\mathbf{t}_{d,\varepsilon}^{(a)} \in \mathbf{N}_{2d,d}$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$, such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ the realization function $\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}): \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ of the ANN $\mathbf{t}_{d,\varepsilon}^{(a)}$ is a suitable approximation of the function $t_d^{(a)}$ and that the number of parameters $\mathcal{P}(\mathbf{t}_{d,\varepsilon}^{(a)})$ of the approximating ANN $\mathbf{t}_{d,\varepsilon}^{(a)}$ is bounded by $cd^\mathfrak{c}\varepsilon^{-\mathfrak{c}}$.

Theorem 4.0.1 establishes in item (i) the essentially well-known result that under the above assumptions it holds that for every $d \in \mathbb{N}$ the Bellman equation (4.4) for the Q -function associated to the MDP with state space $(\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$ has a unique solution $Q: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$.

Moreover, Theorem 4.0.1 concludes in item (ii) that there exists a constant $c \in \mathbb{R}$ such that for every state space dimension $d \in \mathbb{N}$, every prescribed approximation accuracy $\varepsilon \in (0, 1]$, and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ we have that there exists an ANN $\mathbf{Q} \in \mathbf{N}_{d,|A|}$ such that the realization

$$\mathcal{R}_\phi(\mathbf{Q}): \mathbb{R}^d \rightarrow \mathbb{R}^{|A|} \quad (4.7)$$

of the ANN $\mathbf{Q}$ approximates the solution $Q: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ of the Bellman equation for the Q -function in the L^2 -sense with respect to the probability measure μ on $\mathbb{R}^d$ with prescribed approximation accuracy ε and such that the number of parameters $\mathcal{P}(\mathbf{Q})$ of the approximating ANN $\mathbf{Q}$ is bounded by $cd^\mathfrak{c}\varepsilon^{-\mathfrak{c}}$.

The remainder of this chapter is structured as follows. In Section 4.1 below we study stability properties of solutions of functional fixed-point equations of the form (1.3). In Section 4.2 we establish ANN representations for MLFP approximations for certain stochastic fixed-point equations. In Section 4.3 we combine the stability result in Section 4.1 and the ANN representations of the MLFP approximation scheme in Section 4.2 to prove Theorem 4.3.1 in Subsection 4.3.1 and its specializations in Subsection 4.3.2 and Subsection 4.3.3.

4.1. A Stability Result for Solutions of Certain Stochastic Fixed-Point Equations

In this section, we establish stability properties of solutions of functional fixed-point equations of the form (1.3) with respect to perturbations in the nonlinearity and in the transition kernel. In Definitions 4.1.1 to 4.1.3 we introduce the Wasserstein-1 distance as a measure of perturbations of the transition kernels, while in Setting 4.1.4 we recall the function space the considered solutions of functional fixed-point equations are elements of. In Lemma 4.1.5, we establish a stability result for perturbations in the nonlinearity of functional fixed-point equations. While Lemma 4.1.6 establishes Lipschitz properties of solutions of functional fixed-point equations, Lemma 4.1.7 provides a stability result for perturbations with respect to the transition kernels.

Finally, Proposition 4.1.8 combines Lemmas 4.1.5 and 4.1.7 to obtain a stability result for solutions of functional fixed-point equations of the form (1.3) encompassing both perturbations in the nonlinearity and the transition kernel.

Definition 4.1.1 (Couplings). Let $d \in \mathbb{N}$, let $\mu, \nu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ be probability measures. Then we denote by $\Pi_d(\mu, \nu)$ the set of all couplings of μ and ν given by

$$\Pi_d(\mu, \nu) = \left\{ \gamma : \begin{array}{l} \gamma \text{ is a probability measure on } (\mathbb{R}^d \times \mathbb{R}^d, \mathcal{B}(\mathbb{R}^d) \otimes \mathcal{B}(\mathbb{R}^d)), \\ \text{for all } A \in \mathcal{B}(\mathbb{R}^d) \text{ it holds that } \mu(A) = \gamma(A \times \mathbb{R}^d) \text{ and } \nu(A) = \gamma(\mathbb{R}^d \times A) \end{array} \right\}. \quad (4.8)$$

Definition 4.1.2 (Wasserstein-1 space). Let $d \in \mathbb{N}$. Then we denote by $\mathcal{M}_d$ the set given by

$$\mathcal{M}_d = \left\{ (\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]) : \mu \text{ is a probability measure with } \int_{\mathbb{R}^d} \|x\| \mu(dx) < \infty \right\}. \quad (4.9)$$

Definition 4.1.3 (Wasserstein-1 distance). Let $d \in \mathbb{N}$. Then we denote by $\mathbb{W}_d: \mathcal{M}_d \times \mathcal{M}_d \rightarrow [0, \infty)$ the function which satisfies for all $\mu, \nu \in \mathcal{M}_d$ that

$$\mathbb{W}_d(\mu, \nu) = \inf_{\pi \in \Pi_d(\mu, \nu)} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|x - y\| \pi(d(x, y)). \quad (4.10)$$

Setting 4.1.4. Let $d \in \mathbb{N}$, let $\mathbf{w}: \mathbb{R}^d \rightarrow (0, \infty)$ be measurable, let A be a nonempty finite set, and let

$$\mathfrak{W} = \left\{ (u: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}) : u \text{ is measurable, } \sup_{x \in \mathbb{R}^d, a \in A} \frac{|(u(x))(a)|}{|\mathbf{w}(x)|} < \infty \right\}. \quad (4.11)$$

Lemma 4.1.5. Assume Setting 4.1.4, let $c, L_i \in [0, \infty)$, $i \in \{1, 2\}$, satisfy for all $i \in \{1, 2\}$ that $cL_i < 1$, let $f_i: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$, $i \in \{1, 2\}$, be measurable, assume for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ that $|f_i(x, r) - f_i(x, s)| \leq L_i \max_{a \in A} |r(a) - s(a)|$, let $\kappa^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $a \in A$, be stochastic kernels, assume for all $x \in \mathbb{R}^d$, $a \in A$ that $\int_{\mathbb{R}^d} |\mathbf{w}(y)| \kappa^{(a)}(x, dy) \leq c \mathbf{w}(x)$, assume for all $i \in \{1, 2\}$ that

$$\sup_{(x, a) \in \mathbb{R}^d \times A} |\mathbf{w}(x)|^{-1} \int_{\mathbb{R}^d} |f_i(y, 0)| \kappa^{(a)}(x, dy) < \infty. \quad (4.12)$$

Then

(i) there exist unique $u_i \in \mathfrak{W}$, $i \in \{1, 2\}$, such that for every $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that

$$\int_{\mathbb{R}^d} |f_i(y, u_i(y))| \kappa^{(a)}(x, dy) < \infty \quad \text{and} \quad (u_i(x))(a) = \int_{\mathbb{R}^d} f_i(y, u_i(y)) \kappa^{(a)}(x, dy) \quad (4.13)$$

and

(ii) it holds that

$$\left[\sup_{(x, a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathbf{w}(x)|} \right] \leq \frac{c}{1 - c(\min\{L_1, L_2\})} \left[\sup_{(y, r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(y, r) - f_2(y, r)|}{|\mathbf{w}(y)|} \right]. \quad (4.14)$$

Proof of Lemma 4.1.5. First, note that the assumption that $\mathbf{w}$ is measurable, the assumption that for all $i \in \{1, 2\}$ it holds that $cL_i < 1$, the assumption that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f_i(x, r) - f_i(x, s)| \leq L_i \max_{a \in A} |r(a) - s(a)|$, the assumption that for all $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathbf{w}(y)| \kappa^{(a)}(x, dy) \leq c \mathbf{w}(x)$, the assumption that for all $i \in \{1, 2\}$ it holds that $\sup_{(x, a) \in \mathbb{R}^d \times A} |\mathbf{w}(x)|^{-1} \int_{\mathbb{R}^d} |f_i(y, 0)| \kappa^{(a)}(x, dy) < \infty$, and Lemma 2.1.2 (applied for every $i \in \{1, 2\}$ with $c \curvearrowright c$, $L \curvearrowright L_i$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $(\kappa_a)_{a \in A} \curvearrowright (\kappa^{(a)})_{a \in A}$, $f \curvearrowright f_i$, $\mathbf{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathbf{w}(x) \in (0, \infty)]] \in (0, \infty)^{|A|}$) in the notation of Lemma 2.1.2)

establish item (i). Combining this, the assumption that for all $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f_1(x, r) - f_1(x, s)| \leq L_1 \max_{a \in A} |r(a) - s(a)|$, the assumption that for all $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa^{(a)}(x, dy) \leq c \mathfrak{w}(x)$, and the triangle inequality yields for all $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned}
\frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} &= \frac{1}{|\mathfrak{w}(x)|} \left| \int_{\mathbb{R}^d} f_1(y, u_1(y)) \kappa^{(a)}(x, dy) - \int_{\mathbb{R}^d} f_2(y, u_2(y)) \kappa^{(a)}(x, dy) \right| \\
&= \frac{1}{|\mathfrak{w}(x)|} \left| \int_{\mathbb{R}^d} (f_1(y, u_1(y)) - f_2(y, u_2(y))) \kappa^{(a)}(x, dy) \right| \\
&\leq \frac{1}{|\mathfrak{w}(x)|} \left(\int_{\mathbb{R}^d} |f_1(y, u_1(y)) - f_1(y, u_2(y))| \kappa^{(a)}(x, dy) \right. \\
&\quad \left. + \int_{\mathbb{R}^d} |f_1(y, u_2(y)) - f_2(y, u_2(y))| \kappa^{(a)}(x, dy) \right) \\
&\leq \frac{1}{|\mathfrak{w}(x)|} \left(L_1 \int_{\mathbb{R}^d} \max_{b \in A} |(u_1(y))(b) - (u_2(y))(b)| \kappa^{(a)}(x, dy) \right. \\
&\quad \left. + \int_{\mathbb{R}^d} |f_1(y, u_2(y)) - f_2(y, u_2(y))| \kappa^{(a)}(x, dy) \right) \\
&\leq \frac{L_1}{|\mathfrak{w}(x)|} \left[\sup_{(z, c) \in \mathbb{R}^d \times A} \frac{|(u_1(z))(c) - (u_2(z))(c)|}{|\mathfrak{w}(z)|} \right] \int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa^{(a)}(x, dy) \\
&\quad + \frac{1}{|\mathfrak{w}(x)|} \left[\sup_{(z, r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(z, r) - f_2(z, r)|}{|\mathfrak{w}(z)|} \right] \int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa^{(a)}(x, dy) \\
&\leq c L_1 \left[\sup_{(z, c) \in \mathbb{R}^d \times A} \frac{|(u_1(z))(c) - (u_2(z))(c)|}{|\mathfrak{w}(z)|} \right] + c \left[\sup_{(z, r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(z, r) - f_2(z, r)|}{|\mathfrak{w}(z)|} \right].
\end{aligned} \tag{4.15}$$

Moreover, note that item (i) proves that

$$\left[\sup_{(y, b) \in \mathbb{R}^d \times A} \frac{|(u_1(y))(b) - (u_2(y))(b)|}{|\mathfrak{w}(y)|} \right] < \infty. \tag{4.16}$$

This and (4.15) imply that

$$(1 - c L_1) \left[\sup_{(x, a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] \leq c \left[\sup_{(y, r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(y, r) - f_2(y, r)|}{|\mathfrak{w}(y)|} \right]. \tag{4.17}$$

This and the assumption that $c L_1 < 1$ demonstrate that

$$\left[\sup_{(x, a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] \leq \frac{c}{1 - c L_1} \left[\sup_{(y, r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(y, r) - f_2(y, r)|}{|\mathfrak{w}(y)|} \right]. \tag{4.18}$$

Symmetry yields

$$\left[\sup_{(x, a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] \leq \frac{c}{1 - c L_2} \left[\sup_{(y, r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(y, r) - f_2(y, r)|}{|\mathfrak{w}(y)|} \right]. \tag{4.19}$$

This and (4.18) establish that

$$\left[\sup_{(x, a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] \leq \frac{c}{1 - c(\min\{L_1, L_2\})} \left[\sup_{(y, r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(y, r) - f_2(y, r)|}{|\mathfrak{w}(y)|} \right]. \tag{4.20}$$

This proves item (ii). The proof of Lemma 4.1.5 is thus complete. $\square$

Lemma 4.1.6. Assume Setting 4.1.4, let $\eta, c, L, K \in [0, \infty)$, assume $cL < 1$ and $\eta L < 1$, let $f: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ be measurable, assume for all $x, y \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$ and $|f(x, r) - f(y, r)| \leq K \|x - y\|$, let $\kappa^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $a \in A$, be stochastic kernels, assume for all $x \in \mathbb{R}^d$, $a \in A$ that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa^{(a)}(x, dy) \leq c \mathfrak{w}(x)$ and $\int_{\mathbb{R}^d} \|y\| \kappa^{(a)}(x, dy) < \infty$, assume

$$\sup_{(x, a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} |f(y, 0)| \kappa^{(a)}(x, dy) < \infty, \quad (4.21)$$

assume for all $x, y \in \mathbb{R}^d$, $a \in A$ that $\mathbb{W}_d(\kappa^{(a)}(x, \cdot), \kappa^{(a)}(y, \cdot)) \leq \eta \|x - y\|$. Then

(i) there exists a unique $u \in \mathfrak{W}$ such that for every $x \in \mathbb{R}^d$, $a \in A$ it holds that

$$\int_{\mathbb{R}^d} |f(y, u(y))| \kappa^{(a)}(x, dy) < \infty \quad \text{and} \quad (u(x))(a) = \int_{\mathbb{R}^d} f(y, u(y)) \kappa^{(a)}(x, dy) \quad (4.22)$$

and

(ii) it holds for all $x, y \in \mathbb{R}^d$, $a \in A$ that

$$|(u(x))(a) - (u(y))(a)| \leq \frac{\eta K}{1 - \eta L} \|x - y\|. \quad (4.23)$$

Proof of Lemma 4.1.6. First, observe that the assumption that $\mathfrak{w}$ is measurable, the assumption that $cL < 1$, the assumption that for all $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, the assumption that for all $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa^{(a)}(x, dy) \leq c \mathfrak{w}(x)$, the assumption that $\sup_{(x, a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} |f(y, 0)| \kappa^{(a)}(x, dy) < \infty$, and Lemma 2.1.2 (applied with $c \curvearrowright c$, $L \curvearrowright L$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $(\kappa_a)_{a \in A} \curvearrowright (\kappa^{(a)})_{a \in A}$, $f \curvearrowright f$, $\mathfrak{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathfrak{w}(x) \in (0, \infty)]] \in (0, \infty)^{|A|}$) in the notation of Lemma 2.1.2) demonstrates item (i). Throughout the remainder of this proof let $u_n: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$, $n \in \mathbb{N}_0$, satisfy for all $n \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ that $u_0 = 0$ and

$$(u_n(x))(a) = \int_{\mathbb{R}^d} f(y, u_{n-1}(y)) \kappa^{(a)}(x, dy). \quad (4.24)$$

Next, note that the assumption that for all $a \in A$ it holds that $\kappa^{(a)}$ is a stochastic kernel, the assumption that for all $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, the assumption that $cL \in [0, 1)$, the assumption that for all $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa^{(a)}(x, dy) \leq c \mathfrak{w}(x)$, the assumption that $\sup_{(x, a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} |f(y, 0)| \kappa^{(a)}(x, dy) < \infty$, and induction ensure for all $n \in \mathbb{N}_0$ that $u_n \in \mathfrak{W}$. Combining the assumption that for all $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$ and the assumption that $cL \in [0, 1)$ with item (i) yields that $\lim_{n \rightarrow \infty} (\sup_{(x, a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} |(u_n(x))(a) - (u(x))(a)|) = 0$. We prove by induction that for all $n \in \mathbb{N}_0$, $x, y \in \mathbb{R}^d$, $a \in A$ it holds that

$$|(u_n(x))(a) - (u_n(y))(a)| \leq \left[\eta K \sum_{i=0}^{n-1} (\eta L)^i \right] \|x - y\|. \quad (4.25)$$

Note that for all $x, y \in \mathbb{R}^d$, $a \in A$ it holds that $|(u_0(x))(a) - (u_0(y))(a)| = 0 = [\eta K \sum_{i=0}^{-1} (\eta L)^i] \|x - y\|$. This proves (4.25) for the base case $n = 0$. For the induction step $\mathbb{N}_0 \ni (n-1) \dashrightarrow n \in \mathbb{N}$ fix $n \in \mathbb{N}$ and assume for all $l \in \mathbb{N}_0 \cap [0, n-1]$, $x, y \in \mathbb{R}^d$, $a \in A$ that

$$|(u_l(x))(a) - (u_l(y))(a)| \leq \left[\eta K \sum_{i=0}^{l-1} (\eta L)^i \right] \|x - y\|. \quad (4.26)$$

Moreover, note that for all $x, y \in \mathbb{R}^d$, $a \in A$, $\gamma \in \Pi_d(\kappa^{(a)}(x, \cdot), \kappa^{(a)}(y, \cdot))$, $B \in \mathcal{B}(\mathbb{R}^d)$ it holds that $\gamma(B \times \mathbb{R}^d) = \kappa^{(a)}(x, B)$ and $\gamma(\mathbb{R}^d \times B) = \kappa^{(a)}(y, B)$. Combining this, (4.24), the triangle

inequality, the assumption that for all $x, y \in \mathbb{R}^d$, $r, s \in \mathbb{R}^A$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$ and $|f(x, r) - f(y, r)| \leq K \|x - y\|$, and the induction hypothesis that for all $l \in \mathbb{N}_0 \cap [0, n-1]$, $x, y \in \mathbb{R}^d$, $a \in A$ it holds that $|(u_l(x))(a) - (u_l(y))(a)| \leq [\eta K \sum_{i=0}^{l-1} (\eta L)^i] \|x - y\|$ imply for all $x, y \in \mathbb{R}^d$, $a \in A$, $\gamma \in \Pi_d(\kappa^{(a)}(x, \cdot), \kappa^{(a)}(y, \cdot))$ that

$$\begin{aligned}
|(u_n(x))(a) - (u_n(y))(a)| &= \left| \int_{\mathbb{R}^d} f(\mathbf{x}, u_{n-1}(\mathbf{x})) \kappa^{(a)}(x, d\mathbf{x}) - \int_{\mathbb{R}^d} f(\mathbf{y}, u_{n-1}(\mathbf{y})) \kappa^{(a)}(y, d\mathbf{y}) \right| \\
&= \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} f(\mathbf{x}, u_{n-1}(\mathbf{x})) \gamma(d(\mathbf{x}, \mathbf{y})) - \int_{\mathbb{R}^d \times \mathbb{R}^d} f(\mathbf{y}, u_{n-1}(\mathbf{y})) \gamma(d(\mathbf{x}, \mathbf{y})) \right| \\
&\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(\mathbf{x}, u_{n-1}(\mathbf{x})) - f(\mathbf{y}, u_{n-1}(\mathbf{y}))| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(\mathbf{x}, u_{n-1}(\mathbf{x})) - f(\mathbf{x}, u_{n-1}(\mathbf{y}))| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\quad + \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(\mathbf{x}, u_{n-1}(\mathbf{y})) - f(\mathbf{y}, u_{n-1}(\mathbf{y}))| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\leq L \int_{\mathbb{R}^d \times \mathbb{R}^d} \max_{b \in A} |(u_{n-1}(\mathbf{x}))(b) - (u_{n-1}(\mathbf{y}))(b)| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\quad + K \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\leq L \left[\eta K \sum_{i=0}^{n-2} (\eta L)^i \right] \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) + K \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&= \left[K \sum_{i=0}^{n-1} (\eta L)^i \right] \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})). \tag{4.27}
\end{aligned}$$

This and the assumption that for all $x, y \in \mathbb{R}^d$, $a \in A$ it holds that $\mathbb{W}_d(\kappa^{(a)}(x, \cdot), \kappa^{(a)}(y, \cdot)) \leq \eta \|x - y\|$ assure for all $x, y \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned}
|(u_n(x))(a) - (u_n(y))(a)| &\leq \left[K \sum_{i=0}^{n-1} (\eta L)^i \right] \inf_{\gamma \in \Pi_d(\kappa^{(a)}(x, \cdot), \kappa^{(a)}(y, \cdot))} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \tag{4.28} \\
&= \left[K \sum_{i=0}^{n-1} (\eta L)^i \right] \mathbb{W}_d(\kappa^{(a)}(x, \cdot), \kappa^{(a)}(y, \cdot)) \leq \left[\eta K \sum_{i=0}^{n-1} (\eta L)^i \right] \|x - y\|.
\end{aligned}$$

This and induction prove (4.25). Furthermore, note that the assumption that $\eta L \in [0, 1]$ ensures for all $n \in \mathbb{N}_0$ that $\sum_{i=0}^{n-1} (\eta L)^i \leq \sum_{i=0}^{\infty} (\eta L)^i = \frac{1}{1-\eta L}$. This yields that for all $n \in \mathbb{N}_0$, $x, y \in \mathbb{R}^d$, $a \in A$ it holds that

$$|(u_n(x))(a) - (u_n(y))(a)| \leq \frac{\eta K}{1-\eta L} \|x - y\|. \tag{4.29}$$

Combining this with the fact that $\lim_{n \rightarrow \infty} (\sup_{x \in \mathbb{R}^d, a \in A} |\mathfrak{w}(x)|^{-1} |(u_n(x))(a) - (u(x))(a)|) = 0$ establishes for all $x, y \in \mathbb{R}^d$, $a \in A$ that

$$|(u(x))(a) - (u(y))(a)| = \lim_{n \rightarrow \infty} |(u_n(x))(a) - (u_n(y))(a)| \leq \frac{\eta K}{1-\eta L} \|x - y\|. \tag{4.30}$$

This proves item (ii). The proof of Lemma 4.1.6 is thus complete. $\square$

Lemma 4.1.7. Assume Setting 4.1.4, let $\eta, c, L, K \in [0, \infty)$, assume $cL < 1$ and $\eta L < 1$, let $f: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ be measurable, assume for all $x, y \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$ and $|f(x, r) - f(y, r)| \leq K \|x - y\|$, let $\kappa_i^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $i \in \{1, 2\}$, $a \in A$, be stochastic kernels, assume for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa_i^{(a)}(x, dy) \leq c \mathfrak{w}(x)$ and $\int_{\mathbb{R}^d} \|y\| \kappa_i^{(a)}(x, dy) < \infty$, assume for all $i \in \{1, 2\}$ that

$$\sup_{(x, a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} |f(y, 0)| \kappa_i^{(a)}(x, dy) < \infty, \tag{4.31}$$

assume for all $x, y \in \mathbb{R}^d$, $a \in A$ that $\mathbb{W}_d(\kappa_2^{(a)}(x, \cdot), \kappa_2^{(a)}(y, \cdot)) \leq \eta \|x - y\|$, Then

(i) there exist unique $u_i \in \mathfrak{W}$, $i \in \{1, 2\}$, such that for every $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that

$$\int_{\mathbb{R}^d} |f(y, u_i(y))| \kappa_i^{(a)}(x, dy) < \infty \quad \text{and} \quad (u_i(x))(a) = \int_{\mathbb{R}^d} f(y, u_i(y)) \kappa_i^{(a)}(x, dy) \quad (4.32)$$

and

(ii) it holds that

$$\left[\sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] \leq \frac{K}{(1-\eta L)(1-cL)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\kappa_1^{(b)}(y, \cdot), \kappa_2^{(b)}(y, \cdot))}{|\mathfrak{w}(y)|} \right]. \quad (4.33)$$

Proof of Lemma 4.1.7. First, note that the assumption that $\mathfrak{w}$ is measurable, the assumption that $cL < 1$, the assumption that for all $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, the assumption that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa_i^{(a)}(x, dy) \leq c \mathfrak{w}(x)$, the assumption that for all $i \in \{1, 2\}$ it holds that $\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} |f(y, 0)| \kappa_i^{(a)}(x, dy) < \infty$, and Lemma 2.1.2 (applied for every $i \in \{1, 2\}$ with $c \curvearrowright c$, $L \curvearrowright L$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $(\kappa_a)_{a \in A} \curvearrowright (\kappa_i^{(a)})_{a \in A}$, $f \curvearrowright f$, $\mathfrak{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathfrak{w}(x) \in (0, \infty)] \in (0, \infty)^{|A|}]$ in the notation of Lemma 2.1.2) prove item (i). Next, observe that item (ii) in Lemma 4.1.6 (applied with $\eta \curvearrowright \eta$, $c \curvearrowright c$, $L \curvearrowright L$, $K \curvearrowright K$, $f \curvearrowright f$, $(\kappa^{(a)})_{a \in A} \curvearrowright (\kappa_2^{(a)})_{a \in A}$ in the notation of Lemma 4.1.6) yields for all $x, y \in \mathbb{R}^d$, $a \in A$ that

$$|(u_2(x))(a) - (u_2(y))(a)| \leq \frac{\eta K}{1-\eta L} \|x - y\|. \quad (4.34)$$

This, the triangle inequality, and the assumption that for all $x, y \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$ and $|f(x, r) - f(y, r)| \leq K \|x - y\|$ demonstrate for all $x, y \in \mathbb{R}^d$ that

$$\begin{aligned} |f(x, u_2(x)) - f(y, u_2(y))| &\leq |f(x, u_2(x)) - f(x, u_2(y))| + |f(x, u_2(y)) - f(y, u_2(y))| \\ &\leq L \max_{a \in A} |(u_2(x))(a) - (u_2(y))(a)| + K \|x - y\| \\ &\leq \left(L \frac{\eta K}{1-\eta L} + K \right) \|x - y\| = \frac{K}{1-\eta L} \|x - y\|. \end{aligned} \quad (4.35)$$

Combining this, the fact that for all $x \in \mathbb{R}^d$, $a \in A$, $\gamma \in \Pi_d(\kappa_1^{(a)}(x, \cdot), \kappa_2^{(a)}(x, \cdot))$, $B \in \mathcal{B}(\mathbb{R}^d)$ it holds that $\gamma(B \times \mathbb{R}^d) = \kappa_1^{(a)}(x, B)$ and $\gamma(\mathbb{R}^d \times B) = \kappa_2^{(a)}(x, B)$, the triangle inequality, the assumption that for all $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f(x, r) - f(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$, and the assumption that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathfrak{w}(\mathbf{x})| \kappa_i^{(a)}(x, d\mathbf{x}) \leq$

$c\mathfrak{w}(x)$ with item (i) assures for all $x \in \mathbb{R}^d$, $a \in A$, $\gamma \in \Pi_d(\kappa_1^{(a)}(x, \cdot), \kappa_2^{(a)}(x, \cdot))$ that

$$\begin{aligned}
\frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} &= \frac{1}{|\mathfrak{w}(x)|} \left| \int_{\mathbb{R}^d} f(\mathbf{x}, u_1(\mathbf{x})) \kappa_1^{(a)}(x, d\mathbf{x}) - \int_{\mathbb{R}^d} f(\mathbf{y}, u_2(\mathbf{y})) \kappa_2^{(a)}(x, d\mathbf{y}) \right| \\
&= \frac{1}{|\mathfrak{w}(x)|} \left| \int_{\mathbb{R}^d \times \mathbb{R}^d} f(\mathbf{x}, u_1(\mathbf{x})) - f(\mathbf{y}, u_2(\mathbf{y})) \gamma(d(\mathbf{x}, \mathbf{y})) \right| \\
&\leq \frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(\mathbf{x}, u_1(\mathbf{x})) - f(\mathbf{y}, u_2(\mathbf{y}))| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\leq \frac{1}{|\mathfrak{w}(x)|} \left(\int_{\mathbb{R}^d \times \mathbb{R}^d} |f(\mathbf{x}, u_1(\mathbf{x})) - f(\mathbf{x}, u_2(\mathbf{x}))| \gamma(d(\mathbf{x}, \mathbf{y})) \right. \\
&\quad \left. + \int_{\mathbb{R}^d \times \mathbb{R}^d} |f(\mathbf{x}, u_2(\mathbf{x})) - f(\mathbf{y}, u_2(\mathbf{y}))| \gamma(d(\mathbf{x}, \mathbf{y})) \right) \\
&\leq \frac{1}{|\mathfrak{w}(x)|} \left(L \int_{\mathbb{R}^d \times \mathbb{R}^d} \max_{b \in A} |(u_1(\mathbf{x}))(b) - (u_2(\mathbf{x}))(b)| \gamma(d(\mathbf{x}, \mathbf{y})) \right. \\
&\quad \left. + \frac{K}{1-\eta L} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \right) \\
&\leq \frac{L}{|\mathfrak{w}(x)|} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{|(u_1(y))(b) - (u_2(y))(b)|}{|\mathfrak{w}(y)|} \right] \int_{\mathbb{R}^d \times \mathbb{R}^d} |\mathfrak{w}(\mathbf{x})| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\quad + \frac{K}{(1-\eta L)|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \\
&\leq cL \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{|(u_1(y))(b) - (u_2(y))(b)|}{|\mathfrak{w}(y)|} \right] + \frac{K}{(1-\eta L)|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})).
\end{aligned} \tag{4.36}$$

Hence, it holds for all $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned}
\frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} &- cL \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{|(u_1(y))(b) - (u_2(y))(b)|}{|\mathfrak{w}(y)|} \right] \\
&\leq \frac{K}{(1-\eta L)|\mathfrak{w}(x)|} \left(\inf_{\gamma \in \Pi_d(\kappa_1^{(a)}(x, \cdot), \kappa_2^{(a)}(x, \cdot))} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \right) \\
&= \frac{K}{1-\eta L} \frac{\mathbb{W}_d(\kappa_1^{(a)}(x, \cdot), \kappa_2^{(a)}(x, \cdot))}{|\mathfrak{w}(x)|}.
\end{aligned} \tag{4.37}$$

Combining this, the assumption that $cL < 1$, and item (i) shows that

$$\left[\sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] \leq \frac{K}{(1-\eta L)(1-cL)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\kappa_1^{(b)}(y, \cdot), \kappa_2^{(b)}(y, \cdot))}{|\mathfrak{w}(y)|} \right]. \tag{4.38}$$

This proves item (ii). The proof of Lemma 4.1.7 is thus complete. $\square$

Proposition 4.1.8. Assume Setting 4.1.4, let $\eta, c, K \in [0, \infty)$, let $L_i \in [0, \infty)$, $i \in \{1, 2\}$, assume for all $i \in \{1, 2\}$ that $cL_i < 1$ and $\eta L_2 < 1$, let $f_i: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$, $i \in \{1, 2\}$, be measurable, assume for all $i \in \{1, 2\}$, $x, y \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ that $|f_i(x, r) - f_i(x, s)| \leq L_i \max_{a \in A} |r(a) - s(a)|$ and $|f_2(x, r) - f_2(y, r)| \leq K\|x - y\|$, let $\kappa_i^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $i \in \{1, 2\}$, $a \in A$, be stochastic kernels, assume for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa_i^{(a)}(x, dy) \leq c\mathfrak{w}(x)$, assume for all $i \in \{1, 2\}$ that

$$\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} |f_i(y, 0)| \kappa_i^{(a)}(x, dy) < \infty, \tag{4.39}$$

assume for all $i \in \{1, 2\}$ that $\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} \|y\| \kappa_i^{(a)}(x, dy) < \infty$, assume for all $x, y \in \mathbb{R}^d$, $a \in A$ that $\mathbb{W}_d(\kappa_2^{(a)}(x, \cdot), \kappa_2^{(a)}(y, \cdot)) \leq \eta\|x - y\|$. Then

(i) there exist unique $u_i \in \mathfrak{W}$, $i \in \{1, 2\}$, such that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that

$$\int_{\mathbb{R}^d} |f_i(y, u_i(y))| \kappa_i^{(a)}(x, dy) < \infty \quad \text{and} \quad (u_i(x))(a) = \int_{\mathbb{R}^d} f_i(y, u_i(y)) \kappa_i^{(a)}(x, dy) \quad (4.40)$$

and

(ii) it holds that

$$\begin{aligned} \left[\sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] &\leq \frac{c}{1 - c(\min\{L_1, L_2\})} \left[\sup_{(y,r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(y, r) - f_2(y, r)|}{|\mathfrak{w}(y)|} \right] \\ &+ \frac{K}{(1 - \eta L_2)(1 - c L_2)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\kappa_1^{(b)}(y, \cdot), \kappa_2^{(b)}(y, \cdot))}{|\mathfrak{w}(y)|} \right]. \end{aligned} \quad (4.41)$$

Proof of Proposition 4.1.8. First, observe that the assumption that $\mathfrak{w}$ is measurable, the assumption that for all $i \in \{1, 2\}$ it holds that $cL_i < 1$, the assumption that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f_i(x, r) - f_i(x, s)| \leq L_i \max_{a \in A} |r(a) - s(a)|$, the assumption that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathfrak{w}(y)| \kappa_i^{(a)}(x, dy) \leq c\mathfrak{w}(x)$, the assumption that for all $i \in \{1, 2\}$ it holds that $\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} |f_i(y, u_i(y))| \kappa_i^{(a)}(x, dy) < \infty$, and Lemma 2.1.2 (applied for every $i \in \{1, 2\}$ with $c \curvearrowright c$, $L \curvearrowright L_i$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $(\kappa_a)_{a \in A} \curvearrowright (\kappa_i^{(a)})_{a \in A}$, $f \curvearrowright f_i$, $\mathfrak{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathfrak{w}(x) \in (0, \infty)]] \in (0, \infty)^{|A|}$ in the notation of Lemma 2.1.2) establishes item (i). Next, note that the assumption that for all $i \in \{1, 2\}$ it holds that $\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathfrak{w}(x)|^{-1} \int_{\mathbb{R}^d} \|y\| \kappa_i^{(a)}(x, dy) < \infty$ and the fact that for all $x \in \mathbb{R}^d$ it holds that $\mathfrak{w}(x) > 0$ imply that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that

$$\begin{aligned} \int_{\mathbb{R}^d} \|y\| \kappa_i^{(a)}(x, dy) &= \frac{|\mathfrak{w}(x)|}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d} \|y\| \kappa_i^{(a)}(x, dy) \\ &\leq |\mathfrak{w}(x)| \left[\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathfrak{w}(z)|^{-1} \int_{\mathbb{R}^d} \|y\| \kappa_i^{(b)}(z, dy) \right] < \infty. \end{aligned} \quad (4.42)$$

Moreover, note that Fubini's theorem, the triangle inequality, and the assumption that for all $x, y \in \mathbb{R}^d$, $r \in \mathbb{R}^{|A|}$ it holds that $|f_2(x, r) - f_2(y, r)| \leq K\|x - y\|$ ensure for all $x \in \mathbb{R}^d$, $a \in A$

that

$$\begin{aligned}
\frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d} |f_2(\mathbf{x}, 0)| \kappa_1^{(a)}(x, d\mathbf{x}) &= \frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |f_2(\mathbf{x}, 0)| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&\leq \frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |f_2(\mathbf{x}, 0) - f_2(\mathbf{y}, 0)| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&\quad + \frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |f_2(\mathbf{y}, 0)| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&\leq \frac{K}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&\quad + \frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |f_2(\mathbf{y}, 0)| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&\leq \frac{K}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x}\| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&\quad + \frac{K}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{y}\| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&\quad + \frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d \times \mathbb{R}^d} |f_2(\mathbf{y}, 0)| (\kappa_1^{(a)}(x, d\mathbf{x}) \otimes \kappa_2^{(a)}(x, d\mathbf{y})) \\
&= \frac{K}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d} \|\mathbf{x}\| \kappa_1^{(a)}(x, d\mathbf{x}) + \frac{K}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d} \|\mathbf{y}\| \kappa_2^{(a)}(x, d\mathbf{y}) \\
&\quad + \frac{1}{|\mathfrak{w}(x)|} \int_{\mathbb{R}^d} |f_2(\mathbf{y}, 0)| \kappa_2^{(a)}(x, d\mathbf{y}) \\
&\leq K \left(\left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\int_{\mathbb{R}^d} \|\mathbf{x}\| \kappa_1^{(b)}(y, d\mathbf{x})}{|\mathfrak{w}(y)|} \right] + \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\int_{\mathbb{R}^d} \|\mathbf{y}\| \kappa_2^{(b)}(y, d\mathbf{y})}{|\mathfrak{w}(y)|} \right] \right) \\
&\quad + \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\int_{\mathbb{R}^d} |f_2(\mathbf{y}, 0)| \kappa_2^{(b)}(y, d\mathbf{y})}{|\mathfrak{w}(y)|} \right].
\end{aligned} \tag{4.43}$$

This, the assumption that for all $i \in \{1, 2\}$ it holds that $\sup_{(y,b) \in \mathbb{R}^d \times A} |\mathfrak{w}(y)|^{-1} \int_{\mathbb{R}^d} |f_i(\mathbf{y}, 0)| \kappa_i^{(b)}(y, d\mathbf{y}) < \infty$, and the assumption that for all $i \in \{1, 2\}$ it holds that $\sup_{(y,b) \in \mathbb{R}^d \times A} |\mathfrak{w}(y)|^{-1} \int_{\mathbb{R}^d} \|\mathbf{y}\| \kappa_i^{(b)}(y, d\mathbf{y}) < \infty$, show that $\sup_{(y,b) \in \mathbb{R}^d \times A} |\mathfrak{w}(y)|^{-1} \int_{\mathbb{R}^d} |f_2(\mathbf{y}, 0)| \kappa_1^{(b)}(y, d\mathbf{y}) < \infty$. Combining this with Lemma 2.1.2 (applied with $c \curvearrowright c$, $L \curvearrowright L_2$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $(\kappa_a)_{a \in A} \curvearrowright (\kappa_1^{(a)})_{a \in A}$, $f \curvearrowright f_2$, $\mathfrak{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathfrak{w}(x) \in (0, \infty)]] \in (0, \infty)^{|A|}$ in the notation of Lemma 2.1.2) proves that there exists a unique $u \in \mathfrak{W}$ such that for all $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |f_2(y, u(y))| \kappa_1^{(a)}(x, d\mathbf{y}) < \infty$ and $(u(x))(a) = \int_{\mathbb{R}^d} f_2(y, u(y)) \kappa_1^{(a)}(x, d\mathbf{y})$. This, (4.42), the triangle inequality, item (ii) in Lemma 4.1.5 (applied with $c \curvearrowright c$, $L_1 \curvearrowright L_1$, $L_2 \curvearrowright L_2$, $f_1 \curvearrowright f_1$, $f_2 \curvearrowright f_2$, $(\kappa^{(a)})_{a \in A} \curvearrowright (\kappa_1^{(a)})_{a \in A}$ in the notation of Lemma 4.1.5), and item (ii) Lemma 4.1.7 (applied with $\eta \curvearrowright \eta$, $c \curvearrowright c$, $L \curvearrowright L_2$, $K \curvearrowright K$, $f \curvearrowright f_2$, $(\kappa_1^{(a)})_{a \in A} \curvearrowright (\kappa_1^{(a)})_{a \in A}$, $(\kappa_2^{(a)})_{a \in A} \curvearrowright (\kappa_2^{(a)})_{a \in A}$ in the notation of Lemma 4.1.7) demonstrate that

$$\begin{aligned}
\left[\sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] &\leq \left[\sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(u_1(x))(a) - (u(x))(a)|}{|\mathfrak{w}(x)|} \right] + \left[\sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(u(x))(a) - (u_2(x))(a)|}{|\mathfrak{w}(x)|} \right] \\
&\leq \frac{c}{1 - c(\min\{L_1, L_2\})} \left[\sup_{(y,r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|f_1(y, r) - f_2(y, r)|}{|\mathfrak{w}(y)|} \right] \\
&\quad + \frac{K}{(1 - \eta L_2)(1 - c L_2)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\kappa_1^{(b)}(y, \cdot), \kappa_2^{(b)}(y, \cdot))}{|\mathfrak{w}(y)|} \right].
\end{aligned} \tag{4.44}$$

This proves item (ii). The proof of Proposition 4.1.8 is thus complete. $\square$

4.2. Auxiliary ANN Representations and Approximations

In this section, we develop auxiliary representation and approximation results for ANNs. In Subsection 4.2.1 we establish an ANN representation for the maximum function with leaky ReLU activation function as well as an ANN approximation of the maximum function with softplus activation function. To this end, we refer to the related construction in [57, Section 4.2.3] for the special case of the ReLU activation function. In Subsection 4.2.2 we develop representations for MLFP approximations through ANNs with a general activation function.

4.2.1. ANN Representations and Approximations for the Maximum Function

Lemma 4.2.1. *Let $\beta \in [0, \infty) \setminus \{1\}$, let $\phi \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\phi(x) = \max\{x, \beta x\}$. Then there exists $\Phi_n \in \mathbf{N}$, $n \in \mathbb{N} \cap [2, \infty)$, such that*

(i) *it holds for all $n \in \mathbb{N} \cap [2, \infty)$ that*

$$\mathcal{R}_\phi(\Phi_n) \in \begin{cases} C(\mathbb{R}^n, \mathbb{R}^{n/2}) & : n \in \{2, 4, \dots\} \\ C(\mathbb{R}^n, \mathbb{R}^{(n+1)/2}) & : n \in \{3, 5, \dots\}, \end{cases} \quad (4.45)$$

(ii) *it holds for all $n \in \mathbb{N} \cap [2, \infty)$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ that*

$$(\mathcal{R}_\phi(\Phi_n))(x) = \begin{cases} (\max\{x_1, x_2\}, \dots, \max\{x_{n-1}, x_n\}) \in \mathbb{R}^{n/2} & : n \in \{2, 4, \dots\} \\ (\max\{x_1, x_2\}, \dots, \max\{x_{n-2}, x_{n-1}\}, x_n) \in \mathbb{R}^{(n+1)/2} & : n \in \{3, 5, \dots\}, \end{cases} \quad (4.46)$$

(iii) *it holds for all $n \in \mathbb{N} \cap [2, \infty)$ that*

$$\mathcal{D}(\Phi_n) = \begin{cases} (n, 2n, \frac{n}{2}) & : n \in \{2, 4, \dots\} \\ (n, 2n, \frac{n+1}{2}) & : n \in \{3, 5, \dots\}, \end{cases} \quad (4.47)$$

and

(iv) *it holds for all $n \in \mathbb{N} \cap [2, \infty)$ that $\mathcal{L}(\Phi_n) = 2$.*

Proof of Lemma 4.2.1. Throughout this proof let $\gamma, \delta \in \mathbb{R}$ satisfy $\gamma = \frac{|1-\beta|}{(1-\beta)(1-\beta^2)}$ and $\delta = \frac{|1-\beta|}{1-\beta}$, let $T \in \mathbb{R}^{1 \times 2}$, $w_1 \in \mathbb{R}^{4 \times 2}$, $w_2 \in \mathbb{R}^{1 \times 4}$ satisfy

$$T = ((1+\beta)^{-1} \quad -(1+\beta)^{-1}), \quad w_1 = \begin{pmatrix} \delta & -\delta \\ -\delta & \delta \\ 0 & \delta \\ 0 & -\delta \end{pmatrix}, \quad \text{and} \quad (4.48)$$

$$w_2 = (\gamma \quad \gamma\beta \quad \gamma(1-\beta) \quad \gamma(\beta-1)),$$

let $\mathfrak{J}, \mathbf{M} \in \mathbf{N}$ satisfy

$$\mathfrak{J} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), (T, 0) \right) \quad \text{and} \quad \mathbf{M} = ((w_1, 0), (w_2, 0)), \quad (4.49)$$

let $\Phi_n \in \mathbf{N}$, $n \in \mathbb{N} \cap [2, \infty)$, satisfy for all $n \in \mathbb{N} \cap [2, \infty)$ that

$$\Phi_n = \begin{cases} \mathbf{P}_{\frac{n}{2}}(\mathbf{M}, \mathbf{M}, \dots, \mathbf{M}) & : n \in \{2, 4, \dots\} \\ \mathbf{P}_{\frac{n+1}{2}}(\mathbf{M}, \mathbf{M}, \dots, \mathbf{M}, \mathfrak{J}) & : n \in \{3, 5, \dots\}. \end{cases} \quad (4.50)$$

First, note that [1, Equation (4.99)] ensures for all $x \in \mathbb{R}$ that $\max\{x, 0\} = \gamma\beta\phi(-\delta x) + \gamma\phi(\delta x)$. This, (4.48), and (4.49), imply that for all $x \in \mathbb{R}$, $y = (y_1, y_2) \in \mathbb{R}^2$ it holds that $\mathcal{L}(\mathfrak{J}) = \mathcal{L}(\mathbf{M}) = 2$, $\mathcal{D}(\mathfrak{J}) = (1, 2, 1) \in \mathbb{N}^3$, $\mathcal{D}(\mathbf{M}) = (2, 4, 1) \in \mathbb{N}^3$, $\mathcal{R}_\phi(\mathfrak{J}) \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{R}_\phi(\mathbf{M}) \in C(\mathbb{R}^2, \mathbb{R})$,

$$\begin{aligned} (\mathcal{R}_\phi(\mathfrak{J}))(x) &= T\mathfrak{M}_{\phi,2} \left(\begin{pmatrix} 1 \\ -1 \end{pmatrix} x \right) = (1 + \beta)^{-1}(\phi(x) - \phi(-x)) \\ &= (1 + \beta)^{-1}(\max\{x, \beta x\} - \max\{-x, -\beta x\}) = \frac{x + \beta x}{1 + \beta} = x, \quad \text{and} \end{aligned} \quad (4.51)$$

$$\begin{aligned} (\mathcal{R}_\phi(\mathbf{M}))(y) &= w_2\mathfrak{M}_{\phi,4} \left(w_1 \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) = w_2\mathfrak{M}_{\phi,4} \left(\begin{pmatrix} \delta & -\delta \\ -\delta & \delta \\ 0 & \delta \\ 0 & -\delta \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \end{pmatrix} \right) \\ &= (\gamma \quad \gamma\beta \quad \gamma(1 - \beta) \quad \gamma(\beta - 1)) \begin{pmatrix} \phi(\delta(y_1 - y_2)) \\ \phi(-\delta(y_1 - y_2)) \\ \phi(\delta y_2) \\ \phi(-\delta y_2) \end{pmatrix} \\ &= \gamma\phi(\delta(y_1 - y_2)) + \gamma\beta\phi(-\delta(y_1 - y_2)) + \gamma\phi(\delta y_2) + \gamma\beta\phi(-\delta y_2) \\ &\quad - \gamma\beta\phi(\delta y_2) - \gamma\phi(-\delta y_2) \\ &= \max\{y_1 - y_2, 0\} + \max\{y_2, 0\} - \max\{-y_2, 0\} = \max\{y_1, y_2\}. \end{aligned} \quad (4.52)$$

Combining this, (4.50), and Lemma 3.3.2 therefore demonstrates for all $n \in \mathbb{N} \cap [2, \infty)$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ that $\mathcal{L}(\Phi_n) = 2$,

$$\mathcal{D}(\Phi_n) = \begin{cases} (n, 2n, \frac{n}{2}) & : n \in \{2, 4, \dots\} \\ (n, 2n, \frac{n+1}{2}) & : n \in \{3, 5, \dots\}, \end{cases} \quad (4.53)$$

$$\mathcal{R}_\phi(\Phi_n) \in \begin{cases} C(\mathbb{R}^n, \mathbb{R}^{n/2}) & : n \in \{2, 4, \dots\} \\ C(\mathbb{R}^n, \mathbb{R}^{(n+1)/2}) & : n \in \{3, 5, \dots\}, \end{cases} \quad (4.54)$$

and

$$\begin{aligned} (\mathcal{R}_\phi(\Phi_n))(x) &= \begin{cases} (\mathcal{R}_\phi(\mathbf{P}_{\frac{n}{2}}(\mathbf{M}, \mathbf{M}, \dots, \mathbf{M}))(x_1, x_2, \dots, x_n) & : n \in \{2, 4, \dots\} \\ (\mathcal{R}_\phi(\mathbf{P}_{\frac{n+1}{2}}(\mathbf{M}, \mathbf{M}, \dots, \mathbf{M}, \mathfrak{J}))(x_1, x_2, \dots, x_n) & : n \in \{3, 5, \dots\} \end{cases} \\ &= \begin{cases} \begin{pmatrix} (\mathcal{R}_\phi(\mathbf{M}))(x_1, x_2) \\ (\mathcal{R}_\phi(\mathbf{M}))(x_3, x_4) \\ \vdots \\ (\mathcal{R}_\phi(\mathbf{M}))(x_{n-1}, x_n) \end{pmatrix} & : n \in \{2, 4, \dots\} \\ \begin{pmatrix} (\mathcal{R}_\phi(\mathbf{M}))(x_1, x_2) \\ (\mathcal{R}_\phi(\mathbf{M}))(x_3, x_4) \\ \vdots \\ (\mathcal{R}_\phi(\mathbf{M}))(x_{n-2}, x_{n-1}) \\ (\mathcal{R}_\phi(\mathfrak{J}))(x_n) \end{pmatrix} & : n \in \{3, 5, \dots\} \end{cases} \\ &= \begin{cases} (\max\{x_1, x_2\}, \dots, \max\{x_{n-1}, x_n\}) & : n \in \{2, 4, \dots\} \\ (\max\{x_1, x_2\}, \dots, \max\{x_{n-2}, x_{n-1}\}, x_n) & : n \in \{3, 5, \dots\}. \end{cases} \end{aligned} \quad (4.55)$$

This proves items (i) to (iv). The proof of Lemma 4.2.1 is thus complete. $\square$

Lemma 4.2.2 (Maximum of finite set of real numbers (leaky ReLU ANN)). *Let $\beta \in [0, \infty) \setminus \{1\}$, for every $n \in \mathbb{N}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\lfloor x \rfloor \in \mathbb{R}$ satisfy $\lfloor x \rfloor = \max_{i \in \{1, \dots, n\}} |x_i|$, let $\lceil \cdot \rceil : \mathbb{R} \rightarrow \mathbb{Z}$ satisfy for all $x \in \mathbb{R}$ that $\lceil x \rceil = \min\{l \in \mathbb{Z} : l \geq x\}$, let $\phi \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\phi(x) = \max\{x, \beta x\}$. Then there exists $\Psi_m \in \mathbf{N}$, $m \in \mathbb{N}$, such that*

(i) *it holds for all $m \in \mathbb{N}$ that $\mathcal{R}_\phi(\Psi_m) \in C(\mathbb{R}^m, \mathbb{R})$,*

(ii) *it holds for all $m \in \mathbb{N}$, $x = (x_1, \dots, x_m) \in \mathbb{R}^m$ that*

$$(\mathcal{R}_\phi(\Psi_m))(x) = \max_{i \in [1, m] \cap \mathbb{N}} x_i, \quad (4.56)$$

and

(iii) *it holds for all $m \in \mathbb{N}$ that*

$$\|\mathcal{D}(\Psi_m)\| \leq 2m \quad \text{and} \quad \mathcal{L}(\Psi_m) = \lceil \log_2(m) \rceil + 1. \quad (4.57)$$

Proof of Lemma 4.2.2. First, note that Lemma 4.2.1 assures that there exist $\Phi_n \in \mathbf{N}$, $n \in \mathbb{N} \cap [2, \infty)$, such that items (i) to (iv) in Lemma 4.2.1 hold true. Throughout the remainder of this proof let $\Psi_m \in \mathbf{N}$, $m \in \mathbb{N}$, satisfy for all $m \in \mathbb{N} \cap [3, \infty)$ that $\Psi_1 = (1, 0) \in \mathbb{R}^{1 \times 1} \times \mathbb{R}^1$, $\Psi_2 = \Phi_2$, and $\Psi_m = \Psi_{\lceil \frac{m}{2} \rceil} \bullet \Phi_m$. The fact that $\Psi_1 = (1, 0)$ implies for all $x \in \mathbb{R}$ that $\mathcal{D}(\Psi_1) = (1, 1)$, $\mathcal{R}_\phi(\Psi_1) \in C(\mathbb{R}, \mathbb{R})$, $(\mathcal{R}_\phi(\Psi_1))(x) = x$, $\mathcal{L}(\Psi_1) = 1 = \lceil \log_2(1) \rceil + 1$, and $\|\mathcal{D}(\Psi_1)\| = 1 \leq 2$. This proves items (i) to (iii) for the case $m = 1$. Next, we prove items (i) to (iii) for the case $m \in \mathbb{N} \cap [2, \infty)$ by induction on $m \in \mathbb{N} \cap [2, \infty)$. For the base case $m = 2$ observe that the fact that $\Psi_2 = \Phi_2$ and items (i) to (iii) in Lemma 4.2.1 imply for all $x = (x_1, x_2) \in \mathbb{R}^2$ that $\mathcal{R}_\phi(\Psi_2) \in C(\mathbb{R}^2, \mathbb{R})$, $(\mathcal{R}_\phi(\Psi_2))(x_1, x_2) = \max\{x_1, x_2\}$, $\|\mathcal{D}(\Psi_2)\| = 4$, and $\mathcal{L}(\Psi_2) = 2 = \lceil \log_2(2) \rceil + 1$. This proves items (i) to (iii) for the base case $m = 2$. For the induction step $\mathbb{N} \cap [2, \infty) \ni (m-1) \dashrightarrow m \in \mathbb{N} \cap [3, \infty)$ let $m \in \mathbb{N} \cap [3, \infty)$ and assume that items (i) to (iii) hold true for all $l \in \mathbb{N} \cap [2, m-1]$. Observe that the fact that for all $l \in \mathbb{Z}$ it holds that

$$\lceil \frac{l}{2} \rceil = \begin{cases} \frac{l}{2} & : l \in \{2, 4, \dots\} \\ \frac{l+1}{2} & : l \in \{3, 5, \dots\} \end{cases} \quad (4.58)$$

and item (iii) in Lemma 4.2.1 demonstrate for all $l \in \mathbb{N} \cap [2, \infty)$ that

$$\mathcal{D}(\Phi_l) = \left\{ \begin{array}{ll} (l, 2l, l/2) & : l \in \{2, 4, \dots\} \\ (l, 2l, (l+1)/2) & : l \in \{3, 5, \dots\} \end{array} \right\} = (l, 2l, \lceil \frac{l}{2} \rceil) \in \mathbb{N}^3. \quad (4.59)$$

This, the fact that for all $k \in \mathbb{N} \cap [3, \infty)$ it holds that $\lceil \frac{k}{2} \rceil \in \mathbb{N} \cap [2, \infty)$, Lemma 3.2.2, and the induction hypothesis establish that

$$\begin{aligned} \mathcal{I}(\Psi_m) &= \mathcal{I}(\Phi_m) = m, & \mathcal{O}(\Psi_m) &= \mathcal{O}(\Psi_{\lceil \frac{m}{2} \rceil}) = 1, \\ \text{and} \quad \mathcal{R}_\phi(\Psi_m) &= \mathcal{R}_\phi(\Psi_{\lceil \frac{m}{2} \rceil}) \circ \mathcal{R}_\phi(\Phi_m) \in C(\mathbb{R}^m, \mathbb{R}). \end{aligned} \quad (4.60)$$

Moreover, note that Lemma 3.2.2, item (ii) in Lemma 4.2.1, the fact that for all $l \in \mathbb{N} \cap [2, \infty)$ it holds that $\lceil \frac{l}{2} \rceil < l$, and the induction hypothesis that for all $l \in \mathbb{N} \cap [2, m-1]$, $x = (x_1, \dots, x_l) \in \mathbb{R}^l$ it holds that $(\mathcal{R}_\phi(\Psi_l))(x) = \max_{i \in \mathbb{N} \cap [1, l]} x_i$ show for all $x = (x_1, \dots, x_m) \in \mathbb{R}^m$ that

$$\begin{aligned} (\mathcal{R}_\phi(\Psi_m))(x) &= (\mathcal{R}_\phi(\Psi_{\lceil \frac{m}{2} \rceil}) \circ \mathcal{R}_\phi(\Phi_m))(x_1, \dots, x_m) \\ &= \begin{cases} \max\{\max\{x_1, x_2\}, \dots, \max\{x_{m-1}, x_m\}\} & : m \in \{4, 6, \dots\} \\ \max\{\max\{x_1, x_2\}, \dots, \max\{x_{m-2}, x_{m-1}\}, x_m\} & : m \in \{3, 5, \dots\} \end{cases} \\ &= \max\{x_1, \dots, x_m\}. \end{aligned} \quad (4.61)$$

In addition, observe that Lemma 3.2.2, (4.50), the fact that for all $l \in \mathbb{N} \cap [2, \infty)$ it holds that $\lceil \frac{l}{2} \rceil < l$, and the induction hypothesis assure that

$$\|\mathcal{D}(\Psi_m)\| \leq \max \{ \|\mathcal{D}(\Psi_{\lceil \frac{m}{2} \rceil})\|, \|\mathcal{D}(\Phi_m)\| \} \leq \max \{ 2\lceil \frac{m}{2} \rceil, 2m \} = 2m. \quad (4.62)$$

Furthermore, note that the fact that for all $l \in \mathbb{N} \cap [2, \infty)$ it holds that $\lceil \log_2(2\lceil l/2 \rceil) \rceil = \lceil \log_2(l) \rceil$, Lemma 3.2.2, item (iv) in Lemma 4.2.1, and the induction hypothesis imply that

$$\begin{aligned} \mathcal{L}(\Psi_m) &= \mathcal{L}(\Psi_{\lceil \frac{m}{2} \rceil}) + \mathcal{L}(\Phi_m) - 1 = \mathcal{L}(\Psi_{\lceil \frac{m}{2} \rceil}) + 1 = (\lceil \log_2(\lceil m/2 \rceil) \rceil + 1) + 1 \\ &= (\lceil \log_2(2\lceil m/2 \rceil) \rceil) + 1 = \lceil \log_2(m) \rceil + 1. \end{aligned} \quad (4.63)$$

Combining this, (4.60), (4.61), and (4.62) completes the induction step. Induction hence proves items (i) to (iii). The proof of Lemma 4.2.2 is thus complete. $\square$

Lemma 4.2.3. *Let $\phi \in C(\mathbb{R}, \mathbb{R})$ satisfy for all $x \in \mathbb{R}$ that $\phi(x) = \ln(1 + \exp(x))$, let $\lceil \cdot \rceil : \mathbb{R} \rightarrow \mathbb{Z}$ satisfy for all $x \in \mathbb{R}$ that $\lceil x \rceil = \min \{ l \in \mathbb{Z} : l \geq x \}$, and for every $n \in \mathbb{N}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\|x\| \in \mathbb{R}$ satisfy $\|x\| = \max_{i \in \{1, \dots, n\}} |x_i|$. Then there exists $\Psi_{n, \delta} \in \mathbf{N}$, $n \in \mathbb{N}$, $\delta \in (0, 1]$, such that*

- (i) *it holds for all $n \in \mathbb{N}$, $\delta \in (0, 1]$ that $\mathcal{R}_\phi(\Psi_{n, \delta}) \in C(\mathbb{R}^n, \mathbb{R})$,*
- (ii) *it holds for all $n \in \mathbb{N}$, $\delta \in (0, 1]$ that $\|\mathcal{D}(\Psi_{n, \delta})\| \leq \lceil \frac{3n}{2} \rceil$ and $\mathcal{L}(\Psi_{n, \delta}) = \lceil \log_2(n) \rceil + 1$,*
- (iii) *it holds for all $n \in \mathbb{N}$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\delta \in (0, 1]$ that*

$$|(\mathcal{R}_\phi(\Psi_{n, \delta}))(x) - \max\{x_1, \dots, x_n\}| \leq 3 \ln(2) \lceil \log_2(n) \rceil \delta, \quad (4.64)$$

and

- (iv) *it holds for all $n \in \mathbb{N}$, $x, y \in \mathbb{R}^n$, $\delta \in (0, 1]$ that*

$$|(\mathcal{R}_\phi(\Psi_{n, \delta}))(x) - (\mathcal{R}_\phi(\Psi_{n, \delta}))(y)| \leq \|x - y\|. \quad (4.65)$$

Proof of Lemma 4.2.3. For every $n \in \mathbb{N} \cap [2, \infty)$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\mathbf{m}_n(x) \in \mathbb{R}^{\lceil \frac{n}{2} \rceil}$ satisfy

$$\mathbf{m}_n(x) = \begin{cases} (\max\{x_1, x_2\}, \dots, \max\{x_{n-1}, x_n\}) & : n \in \{2, 4, \dots\} \\ (\max\{x_1, x_2\}, \dots, \max\{x_{n-2}, x_{n-1}\}, x_n) & : n \in \{3, 5, \dots\}, \end{cases} \quad (4.66)$$

let $\mathfrak{J} \in \mathbf{N}$ satisfy

$$\mathfrak{J} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), ((1 \quad -1), 0) \right) \in (\mathbb{R}^{2 \times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1 \times 2} \times \mathbb{R}^1), \quad (4.67)$$

let $(w_1^{(\delta)}, w_2^{(\delta)}, b_1^{(\delta)}, b_2^{(\delta)}) \in \mathbb{R}^{3 \times 2} \times \mathbb{R}^{1 \times 3} \times \mathbb{R}^3 \times \mathbb{R}$, $\delta \in (0, 1]$, satisfy for all $\delta \in (0, 1]$ that

$$w_1^{(\delta)} = \delta^{-1} \begin{pmatrix} 1 & -1 \\ 0 & 1 \end{pmatrix}, \quad w_2^{(\delta)} = \delta \begin{pmatrix} 1 & 1 & -1 \end{pmatrix}, \quad b_1^{(\delta)} = 0, \quad b_2^{(\delta)} = 0, \quad (4.68)$$

for every $\delta \in (0, 1]$ let $\mathbf{M}_\delta \in \mathbf{N}$ satisfy that

$$\mathbf{M}_\delta = ((w_1^{(\delta)}, b_1^{(\delta)}), (w_2^{(\delta)}, b_2^{(\delta)})), \quad (4.69)$$

let $\Phi_{n, \delta} \in \mathbf{N}$, $n \in \mathbb{N} \cap [2, \infty)$, $\delta \in (0, 1]$, satisfy for all $n \in \mathbb{N} \cap [2, \infty)$, $\delta \in (0, 1]$ that

$$\Phi_{n, \delta} = \begin{cases} \mathbf{P}_{\frac{n}{2}}(\mathbf{M}_\delta, \mathbf{M}_\delta, \dots, \mathbf{M}_\delta) & : n \in \{2, 4, \dots\} \\ \mathbf{P}_{\frac{n+1}{2}}(\mathbf{M}_\delta, \mathbf{M}_\delta, \dots, \mathbf{M}_\delta, \mathfrak{J}) & : n \in \{3, 5, \dots\}, \end{cases} \quad (4.70)$$

and let $\Psi_{n,\delta} \in \mathbf{N}$, $n \in \mathbb{N}$, $\delta \in (0, 1]$, satisfy for all $n \in \mathbb{N} \cap [3, \infty)$, $\delta \in (0, 1]$ that $\Psi_{1,\delta} = (1, 0) \in \mathbb{R}^{1 \times 1} \times \mathbb{R}^1$, $\Psi_{2,\delta} = \Phi_{2,\delta}$, and $\Psi_{n,\delta} = \Psi_{\lceil \frac{n}{2} \rceil, \delta} \bullet \Phi_{n,\delta}$. First, observe that the fact that for all $\delta \in (0, 1]$ it holds that $\Psi_{1,\delta} = (1, 0)$ ensures for all $x, y \in \mathbb{R}$, $\delta \in (0, 1]$ that $\mathcal{D}(\Psi_{1,\delta}) = (1, 1)$, $\mathcal{R}_\phi(\Psi_{1,\delta}) \in C(\mathbb{R}, \mathbb{R})$, $(\mathcal{R}_\phi(\Psi_{1,\delta}))(x) = x$, $\|\mathcal{D}(\Psi_{1,\delta})\| = 1 \leq \lceil \frac{3}{2} \rceil$, $\mathcal{L}(\Psi_{1,\delta}) = 1 = \lceil \log_2(1) \rceil + 1$, $|(\mathcal{R}_\phi(\Psi_{1,\delta}))(x) - x| = 0 = \lceil \log_2(1) \rceil = 3 \ln(2) \lceil \log_2(1) \rceil \delta$, and $|(\mathcal{R}_\phi(\Psi_{1,\delta}))(x) - (\mathcal{R}_\phi(\Psi_{1,\delta}))(y)| = |x - y| = \lfloor x - y \rfloor$. This proves items (i) to (iv) for the case $n = 1$. Next, note that (4.66) and the fact that for all $x, y, z \in \mathbb{R}$ it holds that $\max\{x, y, z\} = \max\{\max\{x, y\}, z\}$ imply that for all $n \in \mathbb{N} \cap [2, \infty)$, $x = (x_1, \dots, x_n) \in \mathbb{R}^n$ it holds that $\max(\mathbf{m}_n(x)) = \max\{x_1, \dots, x_n\}$. Next, observe that (4.67) and [1, Lemma 3.8] ensure for all $x \in \mathbb{R}^d$ that $\mathcal{R}_\phi(\mathfrak{J}) \in C(\mathbb{R}, \mathbb{R})$, $\mathcal{D}(\mathfrak{J}) = (1, 2, 1)$, and $(\mathcal{R}_\phi(\mathfrak{J}))(x) = x$. Furthermore, note that for all $x_1, x_2 \in \mathbb{R}$, $\delta \in (0, 1]$ it holds that $\mathcal{R}_\phi(\mathbf{M}_\delta) \in C(\mathbb{R}^2, \mathbb{R})$, $\mathcal{D}(\mathbf{M}_\delta) = (2, 3, 1)$, and

$$\begin{aligned} (\mathcal{R}_\phi(\mathbf{M}_\delta))(x_1, x_2) &= w_2^{(\delta)} \mathfrak{M}_{\phi,3} \left(w_1^{(\delta)} \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \right) \\ &= \begin{pmatrix} \delta & \delta & -\delta \end{pmatrix} \begin{pmatrix} \phi(\delta^{-1}(x_1 - x_2)) \\ \phi(\delta^{-1}x_2) \\ \phi(-\delta^{-1}x_2) \end{pmatrix} \\ &= \delta(\phi(\delta^{-1}(x_1 - x_2)) + \phi(\delta^{-1}x_2) - \phi(-\delta^{-1}x_2)). \end{aligned} \quad (4.71)$$

Combining this, the fact that for all $x \in \mathbb{R}^d$, $\delta \in (0, 1]$ it holds that $|\delta\phi(\delta^{-1}x) - \max\{x, 0\}| \leq \ln(2)\delta$, the fact that for all $x_1, x_2 \in \mathbb{R}$ it holds that $\max\{x_1, x_2\} = \max\{x_1 - x_2, 0\} + \max\{x_2, 0\} - \max\{-x_2, 0\}$, and the triangle inequality implies for all $x_1, x_2 \in \mathbb{R}$, $\delta \in (0, 1]$ that

$$|(\mathcal{R}_\phi(\mathbf{M}_\delta))(x_1, x_2) - \max\{x_1, x_2\}| \leq 3 \ln(2)\delta. \quad (4.72)$$

This, (4.66), and (4.70) demonstrate for all $n \in \mathbb{N} \cap [2, \infty)$, $x \in \mathbb{R}^n$, $\delta \in (0, 1]$ that

$$\|(\mathcal{R}_\phi(\Phi_{n,\delta}))(x) - \mathbf{m}_n(x)\| \leq 3 \ln(2)\delta. \quad (4.73)$$

Moreover, note that (4.71) ensures that for all $x_1, x_2 \in \mathbb{R}$, $\delta \in (0, 1]$ it holds that

$$(\mathcal{R}_\phi(\mathbf{M}_\delta))(x_1, x_2) = \delta \ln(\exp(\delta^{-1}x_1) + \exp(\delta^{-1}x_2)). \quad (4.74)$$

This yields for every $x_1, x_2 \in \mathbb{R}$, $\delta \in (0, 1]$ that $\mathcal{R}_\phi(\mathbf{M}_\delta) \in C^1(\mathbb{R}^2, \mathbb{R})$ and

$$\nabla(\mathcal{R}_\phi(\mathbf{M}_\delta))(x_1, x_2) = \left(\frac{\frac{\exp(\delta^{-1}x_1)}{\exp(\delta^{-1}x_1) + \exp(\delta^{-1}x_2)}}{\frac{\exp(\delta^{-1}x_2)}{\exp(\delta^{-1}x_1) + \exp(\delta^{-1}x_2)}} \right). \quad (4.75)$$

This, the fact that for all $x_1, x_2 \in \mathbb{R}$ it holds that $\frac{\exp(x_1)}{\exp(x_1) + \exp(x_2)} \in (0, 1)$, and the mean value theorem demonstrate for all $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$, $\delta \in (0, 1]$ that

$$\begin{aligned} &|(\mathcal{R}_\phi(\mathbf{M}_\delta))(x) - (\mathcal{R}_\phi(\mathbf{M}_\delta))(y)| \\ &\leq \sup_{p_1, p_2 \in \mathbb{R}} \left(\left| \frac{\exp(p_1/\delta)}{\exp(p_1/\delta) + \exp(p_2/\delta)} \right| + \left| \frac{\exp(p_2/\delta)}{\exp(p_1/\delta) + \exp(p_2/\delta)} \right| \right) \lfloor x - y \rfloor = \lfloor x - y \rfloor. \end{aligned} \quad (4.76)$$

This, (4.70), and Lemma 3.3.2 establish for all $l \in \mathbb{N} \cap [2, \infty)$, $x, y \in \mathbb{R}^l$, $\delta \in (0, 1]$ that

$$\|(\mathcal{R}_\phi(\Phi_{l,\delta}))(x) - (\mathcal{R}_\phi(\Phi_{l,\delta}))(y)\| \leq \lfloor x - y \rfloor. \quad (4.77)$$

Next, note that the fact that for all $\delta \in (0, 1]$ it holds that $\mathcal{D}(\mathbf{M}_\delta) = (2, 3, 1)$, the fact that for all $l \in \mathbb{Z}$ it holds that

$$\lceil \frac{l}{2} \rceil = \begin{cases} \frac{l}{2} & : l \in \{0, 2, -2, 4, -4, \dots\} \\ \frac{l+1}{2} & : l \in \{1, -1, 3, -3, \dots\}, \end{cases} \quad (4.78)$$

(4.70), and Lemma 3.3.2 establish for all $n \in \mathbb{N} \cap [2, \infty)$, $\delta \in (0, 1]$ that

$$\mathcal{D}(\Phi_{n,\delta}) = \left\{ \begin{array}{ll} (n, \frac{3n}{2}, \frac{n}{2}) & : n \in \{2, 4, \dots\} \\ (n, \frac{3n+1}{2}, \frac{n+1}{2}) & : n \in \{3, 5, \dots\} \end{array} \right\} = (n, \lceil \frac{3n}{2} \rceil, \lceil \frac{n}{2} \rceil). \quad (4.79)$$

Next, we prove items (i) to (iv) for the case $n \in \mathbb{N} \cap [2, \infty)$ by induction on $n \in \mathbb{N} \cap [2, \infty)$. For the base case $n = 2$ observe that the fact that for all $\delta \in (0, 1]$ it holds that $\Psi_{2,\delta} = \Phi_{2,\delta} = \mathbf{M}_\delta$, (4.69), (4.71), (4.72), (4.76), (4.79), and Lemma 3.3.2 demonstrate that for all $x = (x_1, x_2), y = (y_1, y_2) \in \mathbb{R}^2$, $\delta \in (0, 1]$ it holds that $\mathcal{D}(\Psi_{2,\delta}) = (2, 3, 1)$, $\lceil \mathcal{D}(\Psi_{2,\delta}) \rceil = 3$, $\mathcal{L}(\Psi_{2,\delta}) = 2$, $\mathcal{R}_\phi(\Psi_{2,\delta}) \in C(\mathbb{R}^2, \mathbb{R})$, $|(\mathcal{R}_\phi(\Psi_{2,\delta}))(x) - \max\{x_1, x_2\}| \leq 3 \ln(2)\delta$, and

$$|(\mathcal{R}_\phi(\Psi_{2,\delta}))(x) - (\mathcal{R}_\phi(\Psi_{2,\delta}))(y)| \leq \lceil x - y \rceil. \quad (4.80)$$

This proves items (i) to (iv) in the base case $n = 2$. For the induction step $\mathbb{N} \cap [2, \infty) \ni (n-1) \dashrightarrow n \in \mathbb{N} \cap [3, \infty)$ let $n \in \mathbb{N} \cap [3, \infty)$ and assume that items (i) to (iv) hold true for all $l \in \mathbb{N} \cap [2, n-1]$. First, observe that the fact that for all $k \in \mathbb{N} \cap [3, \infty)$ it holds that $\lceil \frac{k}{2} \rceil \in \mathbb{N} \cap [2, \infty)$, the fact that for all $l \in \mathbb{N} \cap [2, \infty)$ it holds that $\lceil \frac{l}{2} \rceil < l$, the fact that for all $l \in \mathbb{N} \cap [2, \infty)$ it holds that $\lceil \log_2(2^{\lceil l/2 \rceil}) \rceil = \lceil \log_2(l) \rceil$, (4.79), the fact that for all $\Lambda, \Xi \in \mathbf{N}$ with $\mathcal{I}(\Lambda) = \mathcal{O}(\Xi)$ it holds that $\lceil \mathcal{D}(\Lambda \bullet \Xi) \rceil \leq \max\{\lceil \mathcal{D}(\Lambda) \rceil, \lceil \mathcal{D}(\Xi) \rceil\}$, Lemma 3.2.2, and the induction hypothesis yield for all $\delta \in (0, 1]$ that $\mathcal{I}(\Psi_{n,\delta}) = \mathcal{I}(\Phi_{n,\delta}) = n$, $\mathcal{O}(\Psi_{n,\delta}) = \mathcal{O}(\Psi_{\lceil n/2 \rceil, \delta}) = 1$,

$$\lceil \mathcal{D}(\Psi_{n,\delta}) \rceil \leq \max\{\lceil \mathcal{D}(\Psi_{\lceil n/2 \rceil, \delta}) \rceil, \lceil \mathcal{D}(\Phi_{n,\delta}) \rceil\} \leq \max\left\{\left\lceil \frac{3\lceil n/2 \rceil}{2} \right\rceil, \left\lceil \frac{3n}{2} \right\rceil\right\} \leq \left\lceil \frac{3n}{2} \right\rceil, \quad (4.81)$$

and $\mathcal{L}(\Psi_{n,\delta}) = \mathcal{L}(\Psi_{\lceil n/2 \rceil, \delta}) + \mathcal{L}(\Phi_{n,\delta}) - 1 = \lceil \log_2(\lceil n/2 \rceil) \rceil + 1 + 1 = \lceil \log_2(\lceil n/2 \rceil) + 1 \rceil + 1 = \lceil \log_2(2^{\lceil n/2 \rceil}) \rceil + 1 = \lceil \log_2(n) \rceil + 1$. Next, note that the fact that for all $l \in \mathbb{N} \cap [2, \infty)$, $x = (x_1, \dots, x_l) \in \mathbb{R}^l$ it holds that $\max(\mathbf{m}_l(x)) = \max\{x_1, \dots, x_l\}$, the fact that for all $l \in \mathbb{N} \cap [2, \infty)$ it holds that $\lceil \log_2(2^{\lceil l/2 \rceil}) \rceil = \lceil \log_2(l) \rceil$, (4.73), the triangle inequality, and the induction hypothesis show for all $x = (x_1, \dots, x_n) \in \mathbb{R}^n$, $\delta \in (0, 1]$ that

$$\begin{aligned} & |(\mathcal{R}_\phi(\Psi_{n,\delta}))(x_1, x_2, \dots, x_n) - \max\{x_1, x_2, \dots, x_n\}| \\ & \leq |(\mathcal{R}_\phi(\Psi_{\lceil n/2 \rceil, \delta}))((\mathcal{R}_\phi(\Phi_{n,\delta}))(x)) - (\mathcal{R}_\phi(\Psi_{\lceil n/2 \rceil, \delta}))(\mathbf{m}_n(x))| \\ & \quad + |(\mathcal{R}_\phi(\Psi_{\lceil n/2 \rceil, \delta}))(\mathbf{m}_n(x)) - \max(\mathbf{m}_n(x))| \\ & \leq \lceil \mathcal{R}_\phi(\Phi_{n,\delta})(x) - \mathbf{m}_n(x) \rceil + 3\lceil \log_2(\lceil n/2 \rceil) \rceil \ln(2)\delta \\ & \leq 3 \ln(2)\delta(1 + \lceil \log_2(\lceil n/2 \rceil) \rceil) = 3\lceil \log_2(n) \rceil \ln(2)\delta. \end{aligned} \quad (4.82)$$

Moreover, observe that (4.77) and the induction hypothesis demonstrate for all $x, y \in \mathbb{R}^n$, $\delta \in (0, 1]$ that

$$\begin{aligned} & |(\mathcal{R}_\phi(\Psi_{n,\delta}))(x) - (\mathcal{R}_\phi(\Psi_{n,\delta}))(y)| \\ & = |(\mathcal{R}_\phi(\Psi_{\lceil n/2 \rceil, \delta}))((\mathcal{R}_\phi(\Phi_{n,\delta}))(x)) - (\mathcal{R}_\phi(\Psi_{\lceil n/2 \rceil, \delta}))((\mathcal{R}_\phi(\Phi_{n,\delta}))(y))| \\ & \leq \lceil \mathcal{R}_\phi(\Phi_{n,\delta})(x) - \mathcal{R}_\phi(\Phi_{n,\delta})(y) \rceil \leq \lceil x - y \rceil. \end{aligned} \quad (4.83)$$

This, (4.81), (4.82), and induction prove items (i) to (iv). The proof of Lemma 4.2.3 is thus complete. $\square$

4.2.2. ANN Representations for MLFP Approximations

Lemma 4.2.4 (ANN representations for MLFP approximations). *Let $d, \mathfrak{d} \in \mathbb{N}$, let $M \in \mathbb{N} \cap [2, \infty)$, let $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$, let A be a nonempty finite set, let $\mathfrak{a} \in A$, let $\phi \in C(\mathbb{R}, \mathbb{R})$, let $\mathfrak{J}, \mathbf{F} \in \mathbf{N}$ satisfy $\mathcal{D}(\mathfrak{J}) = (1, \mathfrak{d}, 1)$, $\mathcal{R}_\phi(\mathfrak{J}) = \text{id}_{\mathbb{R}}$, and $\mathcal{R}_\phi(\mathbf{F}) \in C(\mathbb{R}^{d+|A|}, \mathbb{R})$, let $\mathbf{X}_a^\theta \in \mathbf{N}$, $\theta \in \Theta$, $a \in A$, satisfy for all $\theta \in \Theta$, $a \in A$ that $\mathcal{D}(\mathbf{X}_a^\theta) = \mathcal{D}(\mathbf{X}_a^0)$ and $\mathcal{R}_\phi(\mathbf{X}_a^\theta) \in C(\mathbb{R}^d, \mathbb{R}^d)$,*

for every $d \in \mathbb{N}$, $x \in (x_1, \dots, x_d) \in \mathbb{R}^d$ let $\mathbf{|x|} \in \mathbb{R}$ satisfy $\mathbf{|x|} = \max_{i \in \{1, \dots, d\}} |x_i|$, let $c = 2 \max\{\mathbf{|D(F)|}, \mathbf{|D(X_a^0)|}, \mathfrak{d}d + \mathfrak{d}|A|\}$, let $\mathcal{U}_n^\theta: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned} (\mathcal{U}_n^\theta(x))(a) = & \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \left[(\mathcal{R}_\phi(\mathbf{F}))((\mathcal{R}_\phi(\mathbf{X}_a^{(\theta, l, i)}))(x), \mathcal{U}_l^{(\theta, l, i)}((\mathcal{R}_\phi(\mathbf{X}_a^{(\theta, l, i)}))(x))) \right. \\ & \left. - \mathbb{1}_{\mathbb{N}}(l)(\mathcal{R}_\phi(\mathbf{F}))((\mathcal{R}_\phi(\mathbf{X}_a^{(\theta, l, i)}))(x), \mathcal{U}_{\max\{l-1, 0\}}^{(\theta, -l, i)}((\mathcal{R}_\phi(\mathbf{X}_a^{(\theta, l, i)}))(x))) \right]. \end{aligned} \quad (4.84)$$

Then there exists $\Phi_n^\theta \in \mathbf{N}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, such that

- (i) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that $\mathcal{R}_\phi(\Phi_n^\theta) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$,
- (ii) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_\phi(\Phi_n^\theta))(x) = \mathcal{U}_n^\theta(x) \in \mathbb{R}^{|A|}$,
- (iii) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that $\mathcal{D}(\Phi_n^\theta) = \mathcal{D}(\Phi_n^0)$,
- (iv) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that

$$\mathcal{L}(\Phi_n^\theta) \leq n(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 1) + 1, \quad (4.85)$$

- (v) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that

$$\mathbf{|D(\Phi_n^\theta)|} \leq c(4|A|M)^n, \quad (4.86)$$

and

- (vi) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that

$$\mathcal{P}(\Phi_n^\theta) \leq 2(n(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 1) + 1)c^2(4|A|M)^{2n}. \quad (4.87)$$

Proof of Lemma 4.2.4. Throughout this proof let $\mathbb{I}_k \in \mathbf{N}$, $k \in \mathbb{N}$, satisfy for all $k \in \mathbb{N}$ that

$$\mathbb{I}_k = \mathbf{P}_k(\mathfrak{J}, \dots, \mathfrak{J}), \quad (4.88)$$

let $\mathbb{J} \in \mathbf{N}^2$ satisfy $\mathbb{J} = (\mathbb{I}_d, \mathbb{I}_{|A|})$. Next, observe that Lemma 3.3.2 ensures for all $k \in \mathbb{N}$, $x \in \mathbb{R}^k$ that $\mathcal{D}(\mathbb{I}_k) = (k, k\mathfrak{d}, k) \in \mathbb{N}^3$, $\mathcal{L}(\mathbb{I}_k) = 2$, $\mathcal{R}_\phi(\mathbb{I}_k) \in C(\mathbb{R}^k, \mathbb{R}^k)$, and $(\mathcal{R}_\phi(\mathbb{I}_k))(x) = x$. Throughout the remainder of this proof let $\Lambda_n^{\theta, a}, \Phi_n^\theta, \Xi_n^\theta \in \mathbf{N}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$, $a \in A$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}$, $a \in A$ that

$$\begin{aligned} \Lambda_0^{\theta, a} &= ((0 \ 0 \ \dots \ 0), 0) \in \mathbb{R}^{1 \times d} \times \mathbb{R}^1, & \Phi_0^\theta &= \mathbf{P}_{|A|}((\Lambda_0^{\theta, a})_{a \in A}) \bullet \mathfrak{T}_{d, |A|}, \\ \Xi_0^\theta &= (0, 0) \in \mathbb{R}^{(|A|+d) \times d} \times \mathbb{R}^{|A|+d}, \end{aligned} \quad (4.89)$$

$$\begin{aligned} \Lambda_n^{\theta, a} &= \left[\bigoplus_{l=0, \mathfrak{J}}^{n-1} \left[\frac{1}{M^{n-l}} \circledast \left(\bigoplus_{i=1, \mathfrak{J}}^{M^{n-l}} \left(\mathbf{F} \bullet \Xi_l^{(\theta, l, i)} \bullet \mathbf{X}_a^{(\theta, l, i)} \right) \right) \right] \right] \\ &\quad \bigoplus_{\mathfrak{J}} \left[\bigoplus_{l=0, \mathfrak{J}}^{n-1} \left[\frac{-\mathbb{1}_{\mathbb{N}}(l)}{M^{n-l}} \circledast \left(\bigoplus_{i=1, \mathfrak{J}}^{M^{n-l}} \left(\mathbf{F} \bullet \Xi_{\max\{l-1, 0\}}^{(\theta, -l, i)} \bullet \mathbf{X}_a^{(\theta, l, i)} \right) \right) \right] \right], \end{aligned} \quad (4.90)$$

$$\Phi_n^\theta = \mathbf{P}_{|A|}((\Lambda_n^{\theta, a})_{a \in A}) \bullet \mathfrak{T}_{d, |A|}, \quad (4.91)$$

$$\Xi_n^\theta = \mathbf{P}_{2, \mathbb{J}}(\mathbb{I}_d, \Phi_n^\theta) \bullet \mathfrak{T}_{d, 2}, \quad (4.92)$$

let $\Psi_{n,l}^{\theta,j,a} \in \mathbf{N}$, $\theta \in \Theta$, $n \in \mathbb{N}$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $a \in A$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $a \in A$ that

$$\Psi_{n,l}^{\theta,j,a} = \bigsqcup_{i=1, \mathfrak{J}}^{M^{n-l}} \left(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^{(\theta, (-1)^j l, i)} \bullet \mathbf{X}_a^{(\theta, l, i)} \right) \in \mathbf{N}, \quad (4.93)$$

and let $\Gamma_n^{\theta,j,a} \in \mathbf{N}$, $\theta \in \Theta$, $n \in \mathbb{N}$, $j \in \{0, 1\}$, $a \in A$, satisfy for all $\theta \in \Theta$, $n \in \mathbb{N}$, $j \in \{0, 1\}$, $a \in A$ that

$$\Gamma_n^{\theta,j,a} = \bigsqcup_{l=0, \mathfrak{J}}^{n-1} \left[\frac{(-1)^j \mathbf{1}_{\mathbb{N}}(l+1-j)}{M^{n-l}} \circledast \Psi_{n,l}^{\theta,j,a} \right], \quad (4.94)$$

let $L_l^j \in \mathbf{N}$, $l \in \mathbb{N}_0 \cap [0, \infty)$, $j \in \{0, 1\}$, satisfy for all $l \in \mathbb{N}_0 \cap [0, \infty)$, $j \in \{0, 1\}$ that

$$L_l^j = \mathcal{L}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^0 \bullet \mathbf{X}_a^0), \quad (4.95)$$

let $\mathfrak{L}_n^j \in \mathbf{N}$, $n \in \mathbb{N}$, $j \in \{0, 1\}$, satisfy for all $n \in \mathbb{N}$, $j \in \{0, 1\}$ that

$$\mathfrak{L}_n^j = \max_{l \in \mathbb{N}_0 \cap [0, n-1]} L_l^j, \quad (4.96)$$

let $\mathbb{L}_n \in \mathbf{N}$, $n \in \mathbb{N}$, satisfy for all $n \in \mathbb{N}$ that

$$\mathbb{L}_n = \max_{j \in \{0, 1\}} \mathfrak{L}_n^j. \quad (4.97)$$

Observe that for $\theta \in \Theta$, $n \in \mathbb{N}$, $a \in A$ it holds that

$$\Lambda_n^{\theta,a} = \Gamma_n^{\theta,0,a} \bigsqcup \mathfrak{I}_n^{\theta,1,a}. \quad (4.98)$$

We prove items (i) to (vi) by induction on $n \in \mathbb{N}_0$. For the base case $n = 0$ note that (4.89) ensures for all $\theta \in \Theta$, $a \in A$ that $\Lambda_0^{\theta,a} = (0, 0) \in \mathbb{R}^{1 \times d} \times \mathbb{R}^1$. This establishes for all $\theta \in \Theta$, $a \in A$ that $\mathcal{D}(\Lambda_0^{\theta,a}) = \mathcal{D}(\Lambda_0^{0,a}) = (d, 1)$. Combining this, the fact that for all $j, k \in \mathbb{N}$ it holds that $\mathcal{D}(\mathfrak{T}_{k,j}) = (k, kj)$, (4.89), Lemma 3.2.2, and Lemma 3.3.2 proves for all $\theta \in \Theta$ that $\mathcal{D}(\Phi_0^\theta) = \mathcal{D}(\Phi_0^0) = (d, |A|)$. Hence, item (iii) holds for the base case $n = 0$. Moreover, note that (4.89) ensures for all $\theta \in \Theta$ that $\mathcal{D}(\Xi_0^\theta) = \mathcal{D}(\Xi_0^0) = (d, d + |A|)$. Thus for all $\theta \in \Theta$, $a \in A$ it holds that

$$\mathcal{L}(\Xi_0^\theta) = \mathcal{L}(\Phi_0^\theta) = \mathcal{L}(\Lambda_0^{\theta,a}) = 0 \cdot (\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 1) + 1. \quad (4.99)$$

This proves item (iv) for the base case $n = 0$. The fact that for all $\theta \in \Theta$ it holds that $\mathcal{D}(\Phi_0^\theta) = (d, |A|)$ implies for all $\theta \in \Theta$ that $\|\mathcal{D}(\Phi_0^\theta)\| = \max\{d, |A|\} \leq d + |A| \leq c$ and $\mathcal{P}(\Phi_0^\theta) = |A|(d + 1) \leq c^2 \leq 2c^2$. This establishes items (v) and (vi) for the base case $n = 0$. Furthermore, note that (4.89) implies for all $\theta \in \Theta$, $x \in \mathbb{R}^d$, $a \in A$ that $\mathcal{R}_\phi(\Lambda_0^{\theta,a}) \in C(\mathbb{R}^d, \mathbb{R})$ and $(\mathcal{R}_\phi(\Lambda_0^{\theta,a}))(x) = 0$. This, (4.89), Lemma 3.2.2, Lemma 3.3.2, and the fact that for all $\theta \in \Theta$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $(\mathcal{U}_0^\theta(x))(a) = 0$ demonstrate for all $\theta \in \Theta$, $x \in \mathbb{R}^d$ that $\mathcal{R}_\phi(\Phi_0^\theta) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$ and

$$\begin{aligned} (\mathcal{R}_\phi(\Phi_0^\theta))(x) &= (\mathcal{R}_\phi(\mathbf{P}_{|A|}((\Lambda_0^{\theta,a})_{a \in A}) \bullet \mathfrak{T}_{d,|A|}))(x) = (\mathcal{R}_\phi(\mathbf{P}_{|A|}((\Lambda_0^{\theta,a})_{a \in A}))) \begin{pmatrix} x \\ x \\ \vdots \\ x \end{pmatrix} \\ &= \left((\mathcal{R}_\phi(\Lambda_0^{\theta,a}))(x) \right)_{a \in A} = 0 = \mathcal{U}_0^\theta(x). \end{aligned} \quad (4.100)$$

Hence items (i) and (ii) hold true for the base case $n = 0$. Thus items (i) to (vi) hold for the base case $n = 0$. For the induction step $\mathbb{N}_0 \ni (n-1) \dashrightarrow n \in \mathbb{N}$ let $n \in \mathbb{N}$ and assume that items (i) to (vi) hold for all $l \in \mathbb{N}_0 \cap [0, n-1]$. First, note that (4.92), Lemma 3.2.2,

Corollary 3.3.4, and the induction hypothesis imply for all $\theta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} (\mathcal{R}_\phi(\Xi_l^\theta))(x) &= (\mathcal{R}_\phi(P_{2,\mathbb{J},(\mathbb{I}_d, \Phi_l^\theta)} \bullet \mathfrak{T}_{d,2}))(x) = (\mathcal{R}_\phi(P_{2,\mathbb{J},(\mathbb{I}_d, \Phi_l^\theta)}) \circ \mathcal{R}_\phi(\mathfrak{T}_{d,2}))(x) \\ &= (\mathcal{R}_\phi(P_{2,\mathbb{J},(\mathbb{I}_d, \Phi_l^\theta)}))(x, x) = \left(\begin{pmatrix} (\mathcal{R}_\phi(\mathbb{I}_d))(x) \\ (\mathcal{R}_\phi(\Phi_l^\theta))(x) \end{pmatrix} \right) = \begin{pmatrix} x \\ \mathcal{U}_l^\theta(x) \end{pmatrix} \in \mathbb{R}^{d+|A|}. \end{aligned} \quad (4.101)$$

This and Lemma 3.2.2 demonstrate for all $\theta, \vartheta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned} (\mathcal{R}_\phi(\mathbf{F} \bullet \Xi_l^\vartheta \bullet \mathbf{X}_a^\theta))(x) &= (\mathcal{R}_\phi(\mathbf{F}) \circ \mathcal{R}_\phi(\Xi_l^\vartheta))((\mathcal{R}_\phi(\mathbf{X}_a^\theta))(x)) \\ &= (\mathcal{R}_\phi(\mathbf{F}))\left((\mathcal{R}_\phi(\mathbf{X}_a^\theta))(x), \mathcal{U}_l^\vartheta((\mathcal{R}_\phi(\mathbf{X}_a^\theta))(x))\right). \end{aligned} \quad (4.102)$$

Moreover, observe that (4.92) and the induction hypothesis establish for all $\theta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$ that $\mathcal{D}(\Xi_l^\theta) = \mathcal{D}(\Xi_l^0)$. Combining this, Lemma 3.2.2, and the assumption that for all $\theta \in \Theta$, $a \in A$ it holds that $\mathcal{D}(\mathbf{X}_a^\theta) = \mathcal{D}(\mathbf{X}_a^0)$ proves for all $\theta, \vartheta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $a \in A$ that

$$\mathcal{D}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^\vartheta \bullet \mathbf{X}_a^\theta) = \mathcal{D}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^0 \bullet \mathbf{X}_a^0). \quad (4.103)$$

This implies for all $\theta, \vartheta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $a \in A$ that

$$\mathcal{L}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^\vartheta \bullet \mathbf{X}_a^\theta) = \mathcal{L}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^0 \bullet \mathbf{X}_a^0) = L_l^j. \quad (4.104)$$

This, Lemma 3.4.7 (applied for every $\theta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $a \in A$ with

$$\begin{aligned} u \curvearrowright 1, \quad v \curvearrowright M^{n-l}, \quad \mathfrak{J} \curvearrowright \mathfrak{J}, \quad (h_k)_{k \in \mathbb{Z} \cap [u, v]} \curvearrowright (1)_{i \in \mathbb{N} \cap [1, M^{n-l}]}, \\ L \curvearrowright L_l^j, \quad \phi \curvearrowright \phi, \quad \Psi \curvearrowright \Psi_{n,l}^{\theta, j, a}, \\ (\Phi_k)_{k \in \mathbb{Z} \cap [u, v]} \curvearrowright (\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^{(\theta, (-1)^{jl, i})} \bullet \mathbf{X}_a^{(\theta, l, i)})_{i \in \mathbb{N} \cap [1, M^{n-l}]} \end{aligned} \quad (4.105)$$

in the notation of Lemma 3.4.7), and item (iii) in Lemma 3.2.7 yield for all $\theta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned} &\mathcal{D}(\Psi_{n,l}^{\theta, j, a}) \\ &= \left(d, \sum_{i=1}^{M^{n-l}} \mathbb{D}_1(\mathcal{E}_{L_l^j, \mathfrak{J}}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^{(\theta, (-1)^{jl, i})} \bullet \mathbf{X}_a^{(\theta, l, i)})), \sum_{i=1}^{M^{n-l}} \mathbb{D}_2(\mathcal{E}_{L_l^j, \mathfrak{J}}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^{(\theta, (-1)^{jl, i})} \bullet \mathbf{X}_a^{(\theta, l, i)})), \dots \right. \\ &\quad \left. \dots, \sum_{i=1}^{M^{n-l}} \mathbb{D}_{L_l^j-1}(\mathcal{E}_{L_l^j, \mathfrak{J}}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^{(\theta, (-1)^{jl, i})} \bullet \mathbf{X}_a^{(\theta, l, i)})), 1 \right) \\ &= \left(d, M^{n-l} \mathbb{D}_1(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^0 \bullet \mathbf{X}_a^0), M^{n-l} \mathbb{D}_2(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^0 \bullet \mathbf{X}_a^0), \dots \right. \\ &\quad \left. \dots, M^{n-l} \mathbb{D}_{L_l^j-1}(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^0 \bullet \mathbf{X}_a^0), 1 \right) \in \mathbb{N}^{L_l^j+1} \end{aligned} \quad (4.106)$$

and

$$\begin{aligned} (\mathcal{R}_\phi(\Psi_{n,l}^{\theta, j, a}))(x) &= \sum_{i=1}^{M^{n-l}} \left(\mathcal{R}_\phi(\mathbf{F} \bullet \Xi_{\max\{l-j, 0\}}^{(\theta, (-1)^{jl, i})} \bullet \mathbf{X}_a^{(\theta, l, i)}) \right)(x) \\ &= \sum_{i=1}^{M^{n-l}} \left(\mathcal{R}_\phi(\mathbf{F}) \right) \left((\mathcal{R}_\phi(\mathbf{X}_a^{(\theta, l, i)}))(x), \mathcal{U}_{\max\{l-j, 0\}}^{(\theta, (-1)^{jl, i})}((\mathcal{R}_\phi(\mathbf{X}_a^{(\theta, l, i)}))(x)) \right). \end{aligned} \quad (4.107)$$

Furthermore, observe that (4.106) and Lemma 3.4.5 assure for all $\theta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $a \in A$ that

$$\mathcal{D}\left(\frac{(-1)^j \mathbf{1}_{\mathbb{N}}(l+1-j)}{M^{n-l}} \circledast \Psi_{n,l}^{\theta,j,a}\right) = \mathcal{D}(\Psi_{n,l}^{\theta,j,a}) = \mathcal{D}(\Psi_{n,l}^{0,j,a}). \quad (4.108)$$

Hence, it holds for all $\theta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$, $j \in \{0, 1\}$, $a \in A$ that $\mathcal{L}(\Psi_{n,l}^{\theta,j,a}) = \mathcal{L}(\Psi_{n,l}^{0,j,a}) = L_l^j$. This ensures for all $\theta \in \Theta$, $a \in A$, $j \in \{0, 1\}$ that $\max_{l \in \mathbb{N}_0 \cap [0, n-1]} \mathcal{L}(\Psi_{n,l}^{\theta,j,a}) = \max_{l \in \mathbb{N}_0 \cap [0, n-1]} L_l^j = \mathfrak{L}_n^j$. Moreover, note that this, (4.94), and Lemma 3.4.7 (applied for every $\theta \in \Theta$, $j \in \{0, 1\}$, $a \in A$ with

$$\begin{aligned} u \curvearrowright 0, \quad v \curvearrowright n-1, \quad \mathfrak{J} \curvearrowright \mathfrak{J}, \quad (h_k)_{k \in \mathbb{Z} \cap [u,v]} \curvearrowright \left(\frac{(-1)^j \mathbf{1}_{\mathbb{N}}(l+1-j)}{M^{n-l}} \right)_{l \in \mathbb{N}_0 \cap [0, n-1]}, \\ L \curvearrowright \mathfrak{L}_n^j, \quad \phi \curvearrowright \phi, \quad \Psi \curvearrowright \Gamma_n^{\theta,j,a}, \quad (\Phi_k)_{k \in \mathbb{Z} \cap [u,v]} \curvearrowright (\Phi_{n,l}^{\theta,j,a})_{l \in \mathbb{N}_0 \cap [0, n-1]} \end{aligned} \quad (4.109)$$

in the notation of Lemma 3.4.7) establish for all $\theta \in \Theta$, $j \in \{0, 1\}$, $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned} \mathcal{D}(\Gamma_n^{\theta,j,a}) &= \left(d, \sum_{l=0}^{n-1} \mathbb{D}_1(\mathcal{E}_{\mathfrak{L}_n^j, \mathfrak{J}}(\Psi_{n,l}^{\theta,j,a})), \sum_{l=0}^{n-1} \mathbb{D}_2(\mathcal{E}_{\mathfrak{L}_n^j, \mathfrak{J}}(\Psi_{n,l}^{\theta,j,a})), \dots, \sum_{l=0}^{n-1} \mathbb{D}_{\mathfrak{L}_n^j-1}(\mathcal{E}_{\mathfrak{L}_n^j, \mathfrak{J}}(\Psi_{n,l}^{\theta,j,a})), 1 \right) \\ &= \left(d, \sum_{l=0}^{n-1} \mathbb{D}_1(\mathcal{E}_{\mathfrak{L}_n^j, \mathfrak{J}}(\Psi_{n,l}^{0,j,a})), \sum_{l=0}^{n-1} \mathbb{D}_2(\mathcal{E}_{\mathfrak{L}_n^j, \mathfrak{J}}(\Psi_{n,l}^{0,j,a})), \dots, \sum_{l=0}^{n-1} \mathbb{D}_{\mathfrak{L}_n^j-1}(\mathcal{E}_{\mathfrak{L}_n^j, \mathfrak{J}}(\Psi_{n,l}^{0,j,a})), 1 \right) \\ &= \mathcal{D}(\Gamma_n^{0,j,a}) \in \mathbb{N}^{\mathfrak{L}_n^j+1} \end{aligned} \quad (4.110)$$

and

$$\begin{aligned} \left(\mathcal{R}_\phi(\Gamma_n^{\theta,j,a}) \right)(x) &= \sum_{l=0}^{n-1} \frac{(-1)^j \mathbf{1}_{\mathbb{N}}(l+1-j)}{M^{n-l}} \left(\mathcal{R}_\phi(\Psi_{n,l}^{\theta,j,a}) \right)(x) \\ &= \sum_{l=0}^{n-1} \frac{(-1)^j \mathbf{1}_{\mathbb{N}}(l+1-j)}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \left(\mathcal{R}_\phi(\mathbf{F}) \circ \mathcal{R}_\phi(\Xi_{\max\{l-j, 0\}}^{(\theta, (-1)^j l, i)}) \circ \mathcal{R}_\phi(\mathbf{X}_a^{(\theta, l, i)}) \right)(x). \end{aligned} \quad (4.111)$$

Note that (4.110) implies for all $\theta \in \Theta$, $a \in A$ that

$$\max\{\mathcal{L}(\Gamma_n^{\theta,0,a}), \mathcal{L}(\Gamma_n^{\theta,1,a})\} = \max\{\mathcal{L}(\Gamma_n^{0,0,a}), \mathcal{L}(\Gamma_n^{0,1,a})\} = \max\{\mathfrak{L}_n^0, \mathfrak{L}_n^1\} = \mathbb{L}_n. \quad (4.112)$$

Combining this, (4.94), (4.101), (4.102), (4.110), (4.111), and Lemma 3.4.7 (applied for every $\theta \in \Theta$, $a \in A$ with $u \curvearrowright 1$, $v \curvearrowright 2$, $\mathfrak{J} \curvearrowright \mathfrak{J}$, $h_1 \curvearrowright 1$, $h_2 \curvearrowright 1$, $L \curvearrowright \mathbb{L}_n$, $\phi \curvearrowright \phi$, $\Phi_1 \curvearrowright \Gamma_n^{\theta,0,a}$, $\Phi_2 \curvearrowright \Gamma_n^{\theta,1,a}$, $\Psi \curvearrowright \Lambda_n^{\theta,a}$ in the notation of Lemma 3.4.7) proves for all $\theta \in \Theta$, $a \in A$ that

$$\begin{aligned} \mathcal{D}(\Lambda_n^{\theta,a}) &= \mathcal{D}(\Gamma_n^{\theta,0,a} \boxplus \Gamma_n^{\theta,1,a}) \\ &= \left(d, \mathbb{D}_1(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{\theta,0,a})) + \mathbb{D}_1(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{\theta,1,a})), \mathbb{D}_2(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{\theta,0,a})) + \mathbb{D}_2(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{\theta,1,a})), \dots \right. \\ &\quad \left. \dots, \mathbb{D}_{\mathbb{L}_n-1}(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{\theta,0,a})) + \mathbb{D}_{\mathbb{L}_n-1}(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{\theta,1,a})), 1 \right) \\ &= \left(d, \mathbb{D}_1(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,0,a})) + \mathbb{D}_1(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,1,a})), \mathbb{D}_2(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,0,a})) + \mathbb{D}_2(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,1,a})), \dots \right. \\ &\quad \left. \dots, \mathbb{D}_{\mathbb{L}_n-1}(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,0,a})) + \mathbb{D}_{\mathbb{L}_n-1}(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,1,a})), 1 \right) \\ &= \mathcal{D}(\Lambda_n^{0,a}) \in \mathbb{N}^{\mathbb{L}_n+1}. \end{aligned} \quad (4.113)$$

and

$$\begin{aligned}
(\mathcal{R}_\phi(\Lambda_n^{\theta,a}))(x) &= (\mathcal{R}_\phi(\Gamma_n^{\theta,0,a}))(x) + (\mathcal{R}_\phi(\Gamma_n^{\theta,1,a}))(x) \\
&= \left[\sum_{l=0}^{n-1} \frac{1}{M^{n-l}} (\mathcal{R}_\phi(\Psi_{n,l}^{\theta,0,a}))(x) \right] + \left[\sum_{l=0}^{n-1} \frac{\mathbb{1}_{\mathbb{N}}(l)}{M^{n-l}} (\mathcal{R}_\phi(\Psi_{n,\max\{l-1,0\}}^{\theta,1,a}))(x) \right] \\
&= \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \left(\mathcal{R}_\phi(\mathbf{F}) \circ \mathcal{R}_\phi(\Xi_l^{(\theta,l,i)}) \circ \mathcal{R}_\phi(\mathbf{X}_a^{(\theta,l,i)}) \right)(x) \\
&\quad - \sum_{l=0}^{n-1} \frac{\mathbb{1}_{\mathbb{N}}(l)}{M^{n-l}} \sum_{i=1}^{M^{n-l}} \left(\mathcal{R}_\phi(\mathbf{F}) \circ \mathcal{R}_\phi(\Xi_{\max\{l-1,0\}}^{(\theta,-l,i)}) \circ \mathcal{R}_\phi(\mathbf{X}_a^{(\theta,l,i)}) \right)(x) \\
&= (\mathcal{U}_n^\theta(x))(a).
\end{aligned} \tag{4.114}$$

This, (4.113), (4.91), Lemma 3.2.2, Lemma 3.3.2 demonstrate for all $\theta \in \Theta$, $x \in \mathbb{R}^d$ that

$$\begin{aligned}
\mathcal{D}(\Phi_n^\theta) &= \mathcal{D}(\mathbf{P}_{|A|}((\Lambda_n^{\theta,a})_{a \in A}) \bullet \mathfrak{T}_{d,|A|}) \\
&= \left(d, \sum_{a \in A} \mathbb{D}_1(\Lambda_n^{\theta,a}), \sum_{a \in A} \mathbb{D}_2(\Lambda_n^{\theta,a}), \dots, \sum_{a \in A} \mathbb{D}_{\mathbb{L}_n-1}(\Lambda_n^{\theta,a}), \sum_{a \in A} \mathbb{D}_{\mathbb{L}_n}(\Lambda_n^{\theta,a}) \right) \\
&= \left(d, |A| \mathbb{D}_1(\Lambda_n^{0,\varepsilon}), |A| \mathbb{D}_2(\Lambda_n^{0,\varepsilon}), \dots, |A| \mathbb{D}_{\mathbb{L}_n-1}(\Lambda_n^{0,\varepsilon}), |A| \right) \in \mathbb{N}^{\mathbb{L}_n+1}
\end{aligned} \tag{4.115}$$

and

$$\begin{aligned}
(\mathcal{R}_\phi(\Phi_n^\theta))(x) &= (\mathcal{R}_\phi(\mathbf{P}_{|A|}((\Lambda_n^{\theta,a})_{a \in A}) \bullet \mathfrak{T}_{d,|A|}))(x) = \left(\mathcal{R}_\phi(\mathbf{P}_{|A|}((\Lambda_n^{\theta,a})_{a \in A})) \right) \begin{pmatrix} x \\ x \\ \vdots \\ x \end{pmatrix} \\
&= \left((\mathcal{R}_\phi(\Lambda_n^{\theta,a}))(x) \right)_{a \in A} = \left((\mathcal{U}_n^\theta(x))(a) \right)_{a \in A} = \mathcal{U}_n^\theta(x) \in \mathbb{R}^{|A|}.
\end{aligned} \tag{4.116}$$

This and induction prove items (i) to (iii). Moreover, note that (4.89), (4.92), the fact that for all $j, k \in \mathbb{N}$ it holds that $\mathcal{L}(\mathfrak{T}_{k,j}) = 1$, Lemma 3.2.2, item (iii) in Lemma 3.2.7, Definition 3.3.3, Corollary 3.3.4, and the induction hypothesis yield for all $\theta \in \Theta$, $l \in \mathbb{N} \cap [1, n-1]$, that

$$\mathcal{L}(\Xi_0^\theta) = \mathcal{L}(\Xi_0^0) = 1 \tag{4.117}$$

and

$$\begin{aligned}
\mathcal{L}(\Xi_l^\theta) &= \mathcal{L}(\Xi_l^0) = \mathcal{L}(\mathbf{P}_{2,\mathbb{J}}(\mathbb{I}_d, \Phi_l^0) \bullet \mathfrak{T}_{d,2}) = \mathcal{L}(\mathbf{P}_{2,\mathbb{J}}(\mathbb{I}_d, \Phi_l^0)) \\
&= \mathcal{L}(\mathbf{P}_2(\mathcal{E}_{\max\{\mathcal{L}(\mathbb{I}_d), \mathcal{L}(\Phi_l^0)\}, \mathbb{I}_d}(\mathbb{I}_d), \mathcal{E}_{\max\{\mathcal{L}(\mathbb{I}_d), \mathcal{L}(\Phi_l^0)\}, \mathbb{I}_{|A|}}(\Phi_l^0))) \\
&= \max\{\mathcal{L}(\mathbb{I}_d), \mathcal{L}(\Phi_l^0)\} = \max\{2, \mathcal{L}(\Phi_l^0)\} \\
&\leq \max\{2, l(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_\varepsilon^0) - 1) + 1\} = l(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_\varepsilon^0) - 1) + 1.
\end{aligned} \tag{4.118}$$

This ensures for all $\theta \in \Theta$, $l \in \mathbb{N}_0 \cap [0, n-1]$ that

$$\mathcal{L}(\Xi_l^\theta) = \mathcal{L}(\Xi_l^0) \leq l(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_\varepsilon^0) - 1) + 1. \tag{4.119}$$

This, (4.97), (4.115), and Lemma 3.2.2 establish for all $\theta \in \Theta$ that

$$\begin{aligned}
\mathcal{L}(\Phi_n^\theta) &= \mathbb{L}_n \\
&= \max_{j \in \{0,1\}} \mathfrak{L}_n^j \\
&= \max_{\substack{l \in \{0,1,\dots,n-1\}, \\ j \in \{0,1\}}} \mathcal{L}(\mathbf{F} \bullet \Xi_{\max\{l-j,0\}}^0 \bullet \mathbf{X}_a^0) \\
&= \max_{\substack{l \in \{0,1,\dots,n-1\}, \\ j \in \{0,1\}}} [\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) + \mathcal{L}(\Xi_{\max\{l-j,0\}}^0) - 2] \\
&\leq \max_{\substack{l \in \{0,1,\dots,n-1\}, \\ j \in \{0,1\}}} [\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 2 + \max\{l-j,0\}(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 1) + 1] \\
&= \mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 2 + (n-1)(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 1) + 1 \\
&\leq n(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_a^0) - 1) + 1.
\end{aligned} \tag{4.120}$$

This and induction prove item (iv). Next, observe that (4.106), (4.110), (4.113), and item (ii) in Lemma 3.2.8 demonstrate for all $\theta \in \Theta$, $a \in A$ that

$$\begin{aligned}
|\mathcal{D}(\Lambda_n^{\theta,a})| &= |\mathcal{D}(\Lambda_n^{0,a})| \leq |\mathcal{D}(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,0,a}))| + |\mathcal{D}(\mathcal{E}_{\mathbb{L}_n, \mathfrak{J}}(\Gamma_n^{0,1,a}))| \\
&\leq \max\{|\mathcal{D}(\mathfrak{J})|, |\mathcal{D}(\Gamma_n^{0,0,a})|\} + \max\{|\mathcal{D}(\mathfrak{J})|, |\mathcal{D}(\Gamma_n^{0,1,a})|\} \\
&\leq \max\left\{\mathfrak{d}, \sum_{l=0}^{n-1} |\mathcal{D}(\mathcal{E}_{\mathfrak{L}_n^0, \mathfrak{J}}(\Psi_{n,l}^{0,0,a}))|\right\} + \max\left\{\mathfrak{d}, \sum_{l=0}^{n-1} |\mathcal{D}(\mathcal{E}_{\mathfrak{L}_n^1, \mathfrak{J}}(\Psi_{n,l}^{0,1,a}))|\right\} \\
&\leq \max\left\{\mathfrak{d}, \sum_{l=0}^{n-1} \max\{\mathfrak{d}, |\mathcal{D}(\Psi_{n,l}^{0,0,a})|\}\right\} + \max\left\{\mathfrak{d}, \sum_{l=0}^{n-1} \max\{\mathfrak{d}, |\mathcal{D}(\Psi_{n,l}^{0,1,a})|\}\right\} \\
&\leq \sum_{l=0}^{n-1} \max\{\mathfrak{d}, |\mathcal{D}(\Psi_{n,l}^{0,0,a})|\} + \sum_{l=0}^{n-1} \max\{\mathfrak{d}, |\mathcal{D}(\Psi_{n,l}^{0,1,a})|\} \\
&\leq \sum_{l=0}^{n-1} \max\{\mathfrak{d}, M^{n-l} \max\{d, |\mathcal{D}(\mathbf{F} \bullet \Xi_l^0 \bullet \mathbf{X}_a^0)|\}\} \\
&\quad + \sum_{l=0}^{n-1} \max\{\mathfrak{d}, M^{n-l} \max\{d, |\mathcal{D}(\mathbf{F} \bullet \Xi_{\max\{l-1,0\}}^0 \bullet \mathbf{X}_a^0)|\}\} \\
&\leq \sum_{l=0}^{n-1} M^{n-l} \max\{\mathfrak{d}, |\mathcal{D}(\mathbf{F} \bullet \Xi_l^0 \bullet \mathbf{X}_a^0)|\} \\
&\quad + \sum_{l=0}^{n-1} M^{n-l} \max\{\mathfrak{d}, |\mathcal{D}(\mathbf{F} \bullet \Xi_{\max\{l-1,0\}}^0 \bullet \mathbf{X}_a^0)|\} \\
&\leq \sum_{l=0}^{n-1} M^{n-l} \max\{\mathfrak{d}, |\mathcal{D}(\mathbf{F})|, |\mathcal{D}(\mathbf{X}_a^0)|, |\mathcal{D}(\Xi_l^0)|\} \\
&\quad + \sum_{l=0}^{n-1} M^{n-l} \max\{\mathfrak{d}, |\mathcal{D}(\mathbf{F})|, |\mathcal{D}(\mathbf{X}_a^0)|, |\mathcal{D}(\Xi_{\max\{l-1,0\}}^0)|\}.
\end{aligned} \tag{4.121}$$

Furthermore, note that the fact that for all $j, k \in \mathbb{N}$ it holds that $\mathcal{D}(\mathfrak{T}_{k,j}) = (k, jk)$, (4.89), (4.92), Lemma 3.2.2, Lemma 3.3.2, and Definition 3.3.3 ensure for all $l \in \mathbb{N} \cap [1, n-1]$ that $|\mathcal{D}(\Xi_0^0)| = d + |A| \leq \mathfrak{d}d + \mathfrak{d}|A| + |\mathcal{D}(\Phi_0^0)|$ and

$$\begin{aligned}
|\mathcal{D}(\Xi_l^0)| &= |\mathcal{D}(\mathbb{P}_{2,\mathbb{J}}(\mathbb{I}_d, \Phi_l^0) \bullet \mathfrak{T}_{d,2})| \\
&\leq \max\{2d, |\mathcal{D}(\mathbb{P}_{2,\mathbb{J}}(\mathbb{I}_d, \Phi_l^0))|\} \\
&\leq \max\left\{2d, |\mathcal{D}(\mathcal{E}_{\max\{\mathcal{L}(\mathbb{I}_d), \mathcal{L}(\Phi_l^0)\}, \mathbb{I}_d}(\mathbb{I}_d))| + |\mathcal{D}(\mathcal{E}_{\max\{\mathcal{L}(\mathbb{I}_d), \mathcal{L}(\Phi_l^0)\}, \mathbb{I}_{|A|}}(\Phi_l^0))|\right\} \\
&\leq \max\{2d, \mathfrak{d}d + \max\{\mathfrak{d}|A|, |\mathcal{D}(\Phi_l^0)|\}\} \\
&= \mathfrak{d}d + \max\{\mathfrak{d}|A|, |\mathcal{D}(\Phi_l^0)|\} \leq \mathfrak{d}d + \mathfrak{d}|A| + |\mathcal{D}(\Phi_l^0)|.
\end{aligned} \tag{4.122}$$

This implies for all $l \in \mathbb{N}_0 \cap [0, n-1]$ that $\|\mathcal{D}(\Xi_l^0)\| \leq \mathfrak{d}d + \mathfrak{d}|A| + \|\mathcal{D}(\Phi_l^0)\|$. Combining this, (4.121), the assumption that $M \geq 2$, and the induction hypothesis yields that

$$\begin{aligned}
\|\mathcal{D}(\Lambda_n^{0,\varepsilon})\| &\leq \sum_{l=0}^{n-1} M^{n-l} \max \{ \mathfrak{d}, \|\mathcal{D}(\mathbf{F})\|, \|\mathcal{D}(\mathbf{X}_\varepsilon^0)\|, \|\mathcal{D}(\Xi_l^0)\| \} \\
&\quad + \sum_{l=0}^{n-1} M^{n-l} \max \{ \mathfrak{d}, \|\mathcal{D}(\mathbf{F})\|, \|\mathcal{D}(\mathbf{X}_\varepsilon^0)\|, \|\mathcal{D}(\Xi_{\max\{l-1,0\}}^0)\| \} \\
&\leq \sum_{l=0}^{n-1} M^{n-l} \max \{ \mathfrak{d}, \|\mathcal{D}(\mathbf{F})\|, \|\mathcal{D}(\mathbf{X}_\varepsilon^0)\|, \mathfrak{d}d + \mathfrak{d}|A| + \|\mathcal{D}(\Phi_l^0)\| \} \\
&\quad + \sum_{l=0}^{n-1} M^{n-l} \max \{ \mathfrak{d}, \|\mathcal{D}(\mathbf{F})\|, \|\mathcal{D}(\mathbf{X}_\varepsilon^0)\|, \mathfrak{d}d + \mathfrak{d}|A| + \|\mathcal{D}(\Phi_{\max\{l-1,0\}}^0)\| \} \\
&\leq \sum_{l=0}^{n-1} M^{n-l} \left[\max \{ \|\mathcal{D}(\mathbf{F})\|, \|\mathcal{D}(\mathbf{X}_\varepsilon^0)\|, \mathfrak{d}d + \mathfrak{d}|A| \} + \|\mathcal{D}(\Phi_l^0)\| \right] \\
&\quad + \sum_{l=0}^{n-1} M^{n-l} \left[\max \{ \|\mathcal{D}(\mathbf{F})\|, \|\mathcal{D}(\mathbf{X}_\varepsilon^0)\|, \mathfrak{d}d + \mathfrak{d}|A| \} + \|\mathcal{D}(\Phi_{\max\{l-1,0\}}^0)\| \right] \\
&= c \left[\sum_{l=0}^{n-1} M^{n-l} \right] + \left[\sum_{l=0}^{n-1} M^{n-l} \left(\|\mathcal{D}(\Phi_l^0)\| + \|\mathcal{D}(\Phi_{\max\{l-1,0\}}^0)\| \right) \right] \\
&\leq c \left[\sum_{l=0}^{n-1} M^{n-l} \right] + \left[\sum_{l=0}^{n-1} M^{n-l} (4|A|M)^l \right] + \left[\sum_{l=0}^{n-1} M^{n-l} (4|A|M)^{\max\{l-1,0\}} \right] \quad (4.123) \\
&= c \left[M^{\frac{M^n-1}{M-1}} + M^n \left[\sum_{l=0}^{n-1} (4|A|)^l \right] + M^n + M^{n-1} \left[\sum_{l=1}^{n-1} (4|A|)^{l-1} \right] \right] \\
&\leq c \left[3M^n + M^n \frac{(4|A|)^n - 1}{4|A| - 1} + M^{n-1} \frac{(4|A|)^{n-1} - 1}{4|A| - 1} \right] \\
&= cM^n \left[3 + \frac{(4|A|)^n - 1}{4|A| - 1} + \frac{1}{M} \frac{(4|A|)^{n-1} - 1}{4|A| - 1} \right] \\
&\leq cM^n \left[\frac{(4|A|)^n}{4|A| - 1} + \frac{(4|A|)^{n-1}}{4|A| - 1} + \frac{3(4|A| - 1) - 2}{4|A| - 1} \right] \\
&= c(4|A|M)^n \left[\frac{1}{4|A| - 1} + \frac{1}{4|A|(4|A| - 1)} + \frac{3(4|A| - 1) - 2}{(4|A|)^n(4|A| - 1)} \right] \\
&\leq c(4|A|M)^n \left[\frac{4|A| + 1}{4|A|(4|A| - 1)} + \frac{3(4|A| - 1) - 2}{4|A|(4|A| - 1)} \right] \\
&= c(4|A|M)^n \left[\frac{4(4|A| - 1)}{4|A|(4|A| - 1)} \right] = c(4|A|M)^n \frac{1}{|A|}.
\end{aligned}$$

Combining this and (4.91) demonstrates that

$$\|\mathcal{D}(\Phi_n^0)\| \leq |A| \|\mathcal{D}(\Lambda_n^{0,\varepsilon})\| \leq c(4|A|M)^n. \quad (4.124)$$

This and induction prove item (v). This, item (iv), and the fact that for all $L \in \mathbb{N}$, $\zeta \in \mathbf{N}$, $l_0, l_1, \dots, l_L \in \mathbb{N}$ with $\mathcal{D}(\zeta) = (l_0, l_1, \dots, l_L) \in \mathbb{N}^{L+1}$ it holds that $\mathcal{P}(\zeta) = \sum_{k=1}^L l_k(l_{k-1} + 1)$ establish that

$$\begin{aligned}
\mathcal{P}(\Phi_n^0) &\leq \sum_{k=1}^{\mathcal{L}(\Phi_n^0)} \|\mathcal{D}(\Phi_n^0)\| (\|\mathcal{D}(\Phi_n^0)\| + 1) \leq 2\mathcal{L}(\Phi_n^0) \|\mathcal{D}(\Phi_n^0)\|^2 \\
&\leq 2[n(\mathcal{L}(\mathbf{F}) + \mathcal{L}(\mathbf{X}_\varepsilon^0) - 1) + 1] c^2 (4|A|M)^{2n}.
\end{aligned} \quad (4.125)$$

This and induction prove item (vi). Hence, items (i) to (vi) hold true. The proof of Lemma 4.2.4 is thus complete. $\square$

4.3. ANN Approximations for Functional Fixed-Point Equations

In this section, we establish ANN approximation results for solutions of functional fixed-point equations of the form (1.3). To this end, we combine the representation of the MLFP scheme by means of ANNS in Section 4.2 to prove in Subsection 4.3.1 the main contribution of this chapter, Theorem 4.3.1, which establishes ANN approximation results for solutions of functional fixed-point equations with general activation functions. In Subsection 4.3.2, we specialize these ANN approximation results to specific activation functions and one-step transitions. Finally, in Subsection 4.3.3, the results of Subsection 4.3.2 are specialized to the case of Bellman equations.

4.3.1. ANN Approximations for Functional Fixed-Point Equations with General Activation Functions

Theorem 4.3.1. *Let $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$, let A be a nonempty finite set, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathfrak{c} \in [1, \infty)$, for every $d \in \mathbb{N}$ let $\lambda_d, \eta_d, K_d, L_d \in (0, \infty)$, for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ let $\ell_{d,\varepsilon} \in (0, \infty)$, assume $\sup_{d \in \mathbb{N}} \lambda_d L_d < 1$, $\sup_{d \in \mathbb{N}} \eta_d L_d < 1$, $\sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \lambda_d \ell_{d,\varepsilon} < 1$, for every $d \in \mathbb{N}$ let $\mathfrak{w}_d: \mathbb{R}^d \rightarrow (0, \infty)$ be measurable, for every $d \in \mathbb{N}$, $x \in (x_1, \dots, x_d) \in \mathbb{R}^d$ let $\mathbf{x} \in \mathbb{R}$ satisfy $\mathbf{x} = \max_{i \in \{1, \dots, d\}} |x_i|$, for every $d \in \mathbb{N}$ let $f_d: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ be measurable, for every $d \in \mathbb{N}$ let $\kappa_d^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $a \in A$, be stochastic kernels, let $a \in A$, let $\bar{\omega} \in \Omega$, let $\phi \in C(\mathbb{R}, \mathbb{R})$, let $\mathfrak{J} \in \mathbb{N}$ satisfy $\mathcal{R}_\phi(\mathfrak{J}) = \text{id}_{\mathbb{R}}$ and $\mathcal{L}(\mathfrak{J}) = 2$, let $(\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \subseteq \mathbb{N}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in \mathbb{N}$ that $\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}) \in C(\mathbb{R}^{d+|A|}, \mathbb{R})$, let $\mathbf{X}_{d,\varepsilon}^{\theta,a}: \Omega \rightarrow \mathbb{N}$, $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$, satisfy for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$, $\omega \in \Omega$ that $\mathcal{D}(\mathbf{X}_{d,\varepsilon}^{\theta,a}(\omega)) = \mathcal{D}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega}))$ and $\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{0,a}(\omega)) \in C(\mathbb{R}^d, \mathbb{R}^d)$, for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ let $\rho_{d,\varepsilon}^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $a \in A$, satisfy for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$, $Y \in \mathcal{B}(\mathbb{R}^d)$ that $\rho_{d,\varepsilon}^{(a)}(x, Y) = \mathbb{P}[(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{\theta,a}))(x) \in Y]$, for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ let $(\mathcal{F}_{d,\varepsilon}^\theta)_{\theta \in \Theta}$ be independent sub- σ -algebras of $\mathcal{F}$, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{\theta,a}))_{a \in A}: \mathbb{R}^d \times \Omega \rightarrow (\mathbb{R}^d)^{|A|}$, $\theta \in \Theta$, are i.i.d. random fields, assume for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{\theta,a}))_{a \in A}$ is $(\mathcal{B}(\mathbb{R}^d) \otimes \mathcal{F}_{d,\varepsilon}^\theta) / \mathcal{B}((\mathbb{R}^d)^{|A|})$ -measurable, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $\max\{\lambda_d, K_d\} \leq \mathfrak{c}d^\mathfrak{c}$ and*

$$\max\{\mathcal{L}(\mathbf{F}_{d,\varepsilon}), \mathcal{L}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega})), \|\mathcal{D}(\mathbf{F}_{d,\varepsilon})\|, \|\mathcal{D}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega}))\|\} \leq \mathfrak{c}d^\mathfrak{c}\varepsilon^{-\mathfrak{c}}, \quad (4.126)$$

assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, y \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$, $a \in A$ that

$$|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - f_d(x, r)| \leq \varepsilon \mathfrak{c}d^\mathfrak{c} |\mathfrak{w}_d(x)|, \quad \mathbb{W}_d(\rho_{d,\varepsilon}^{(a)}(x, \cdot), \kappa_d^{(a)}(x, \cdot)) \leq \varepsilon \mathfrak{c}d^\mathfrak{c} |\mathfrak{w}_d(x)|, \quad (4.127)$$

$$|f_d(x, r) - f_d(x, s)| \leq L_d \max_{b \in A} |r(b) - s(b)|, \quad |f_d(x, r) - f_d(y, r)| \leq K_d \|x - y\|, \quad (4.128)$$

$$|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, s)| \leq \ell_{d,\varepsilon} \max_{b \in A} |r(b) - s(b)|, \quad \text{and}$$

$$\mathbb{W}_d(\kappa_d^{(a)}(x, \cdot), \kappa_d^{(a)}(y, \cdot)) \leq \eta_d \|x - y\|,$$

assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ that

$$\sup_{x \in \mathbb{R}^d} |\mathfrak{w}_d(x)|^{-1} \int_{\mathbb{R}^d} \|y\| \kappa_d^{(a)}(x, dy) < \infty \quad \text{and} \quad \sup_{x \in \mathbb{R}^d} |\mathfrak{w}_d(x)|^{-1} \int_{\mathbb{R}^d} \|y\| \rho_{d,\varepsilon}^{(a)}(x, dy) < \infty, \quad (4.129)$$

assume for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $a \in A$ that

$$\left(\int_{\mathbb{R}^d} |\mathfrak{w}_d(y)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy) \right)^{1/2} \leq \lambda_d |\mathfrak{w}_d(x)|, \quad (4.130)$$

$$\left(\int_{\mathbb{R}^d} |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, 0)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy) \right)^{1/2} \leq c d^c |\mathfrak{w}_d(x)|,$$

$$\int_{\mathbb{R}^d} |\mathfrak{w}_d(y)| \kappa_d^{(a)}(x, dy) \leq \lambda_d |\mathfrak{w}_d(x)|, \quad \text{and} \quad \sup_{\substack{z \in \mathbb{R}^d, \\ b \in A}} |\mathfrak{w}_d(z)|^{-1} \int_{\mathbb{R}^d} |f_d(y, 0)| \kappa_d^{(b)}(z, dy) < \infty. \quad (4.131)$$

Then there exists $c \in \mathbb{R}$ such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq c d^c$ it holds that

(i) there exists a unique measurable $u: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that $\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathfrak{w}_d(z)|^{-1} |(u(z))(b)| < \infty$, $\int_{\mathbb{R}^d} |f_d(y, u(y))| \kappa_d^{(a)}(x, dy) < \infty$, and

$$(u(x))(a) = \int_{\mathbb{R}^d} f_d(y, u(y)) \kappa_d^{(a)}(x, dy) \quad (4.132)$$

and

(ii) there exists $\Psi \in \mathbf{N}$ which satisfies that $\mathcal{H}(\Psi) > 0$, $\mathcal{P}(\Psi) \leq c d^c \varepsilon^{-c}$, $\mathcal{R}_\phi(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, and

$$\left(\int_{\mathbb{R}^d} \|u(x) - (\mathcal{R}_\phi(\Psi))(x)\|^2 \mu(dx) \right)^{1/2} \leq \varepsilon. \quad (4.133)$$

Proof of Theorem 4.3.1. Throughout this proof let $\mathfrak{d} \in \mathbb{N}$ satisfy $\mathcal{D}(\mathfrak{J}) = (1, \mathfrak{d}, 1)$, let $\mathfrak{K} \in \mathbb{R}$ satisfy $\mathfrak{K} = \max\{\mathfrak{c}, \mathfrak{d}(1 + |A|)\}$, let $M \in \mathbb{N} \cap [2, \infty)$ satisfy

$$M > \left(\frac{1 + (\sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \lambda_d \ell_{d,\varepsilon})(2|A| - 1)}{1 - (\sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \lambda_d \ell_{d,\varepsilon})} \right)^2, \quad (4.134)$$

let $\alpha_{d,\varepsilon}, \gamma_{d,\varepsilon} \in [0, \infty)$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\alpha_{d,\varepsilon} = \frac{\lambda_d \ell_{d,\varepsilon} (1 + |A| M^{-1/2}) + M^{-1/2} + \sqrt{(\lambda_d \ell_{d,\varepsilon} (1 + |A| M^{-1/2}) + M^{-1/2})^2 + 4 M^{-1/2} \lambda_d \ell_{d,\varepsilon} (|A| - 1)}}{2} \quad \text{and}$$

$$\gamma_{d,\varepsilon} = \frac{3}{2} \max \left\{ \frac{\mathfrak{c}}{1 - \lambda_d \ell_{d,\varepsilon}}, \frac{|A| \mathfrak{c}}{|A| \lambda_d \ell_{d,\varepsilon} + 1} \right\}, \quad (4.135)$$

let $\alpha, \gamma \in [0, \infty]$ satisfy $\alpha = \sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \alpha_{d,\varepsilon}$ and $\gamma = \sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \gamma_{d,\varepsilon}$, let $\mathfrak{n}: (0, 1] \rightarrow \mathbb{N} \cup \{+\infty\}$ satisfy for all $\varepsilon \in (0, 1]$ that $\mathfrak{n}(\varepsilon) = \min(\{n \in \mathbb{N} : \gamma \alpha^n \leq \varepsilon\} \cup \{+\infty\})$. First, observe that the assumption that $\sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \lambda_d \ell_{d,\varepsilon} < 1$ and (4.134) ensure that $\alpha \in [M^{-1/2}, 1)$. Next, note that the assumption that $\sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \lambda_d \ell_{d,\varepsilon} < 1$ and the assumption that $\mathfrak{c} \in [1, \infty)$ imply that $\gamma \in [3/2, \infty)$. This and the fact that $\alpha \in (0, 1)$ show for all $\varepsilon \in (0, 1]$ that $\mathfrak{n}(\varepsilon) \in \mathbb{N}$. Furthermore, let $\mathcal{V}_{d,\varepsilon,n}^\theta: \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^{|A|}$, $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $n \in \mathbb{N}_0$ satisfy for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned} & \left(\mathcal{V}_{d,\varepsilon,n}^\theta(x) \right)(a) \\ &= \sum_{l=0}^{n-1} \frac{1}{M^{n-l}} \left[\sum_{i=1}^{M^{n-l}} \left[\left(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}) \right) \left(\left(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{(\theta,l,i),a}) \right)(x), \mathcal{V}_{d,\varepsilon,l}^{(\theta,l,i)} \left(\left(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{(\theta,l,i),a}) \right)(x) \right) \right) \right. \right. \\ & \quad \left. \left. - \mathbb{1}_{\mathbb{N}}(l) \left(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}) \right) \left(\left(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{(\theta,l,i),a}) \right)(x), \mathcal{V}_{d,\varepsilon,\max\{l-1,0\}}^{(\theta,-l,i)} \left(\left(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{(\theta,l,i),a}) \right)(x) \right) \right) \right] \right]. \end{aligned} \quad (4.136)$$

Note that the assumption that for all $d \in \mathbb{N}$ it holds that $\lambda_d L_d < 1$, the assumption that for all $d \in \mathbb{N}$ it holds that $\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathbf{w}_d(x)|^{-1} \int_{\mathbb{R}^d} f_d(y, 0) |\kappa_d^{(a)}(x, dy)| < \infty$, the assumption that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $\int_{\mathbb{R}^d} |\mathbf{w}_d(y)| |\kappa_d^{(a)}(x, dy)| \leq \lambda_d |\mathbf{w}_d(x)|$, and Lemma 2.1.2 (applied for every $d \in \mathbb{N}$ with $c \curvearrowright \lambda_d$, $L \curvearrowright L_d$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $(\kappa_a)_{a \in A} \curvearrowright (\kappa_d^{(a)})_{a \in A}$, $f \curvearrowright f_d$, $\mathbf{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathbf{w}_d(x) \in (0, \infty)]] \in (0, \infty)^{|A|}$ in the notation of Lemma 2.1.2) yield that for every $d \in \mathbb{N}$ there exists a unique measurable $u_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that $\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathbf{w}_d(z)|^{-1} |(u_d(z))(b)| < \infty$, $\int_{\mathbb{R}^d} f_d(y, u_d(y)) |\kappa_d^{(a)}(x, dy)| < \infty$, and

$$(u_d(x))(a) = \int_{\mathbb{R}^d} f_d(y, u_d(y)) \kappa_d^{(a)}(x, dy). \quad (4.137)$$

This proves item (i). Next, observe that the assumptions that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\lambda_d \ell_{d,\varepsilon} < 1$, the assumptions that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $(\int_{\mathbb{R}^d} |\mathbf{w}_d(y)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy))^{1/2} \leq \lambda_d |\mathbf{w}_d(x)|$ and $(\int_{\mathbb{R}^d} |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, 0)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy))^{1/2} \leq c d^c |\mathbf{w}_d(x)|$, and Lemma 2.1.2 (applied for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ with $c \curvearrowright \lambda_d$, $L \curvearrowright \ell_{d,\varepsilon}$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $(\kappa_a)_{a \in A} \curvearrowright (\rho_{d,\varepsilon}^{(a)})_{a \in A}$, $f \curvearrowright \mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon})$, $\mathbf{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathbf{w}_d(x) \in (0, \infty)]] \in (0, \infty)^{|A|}$ in the notation of Lemma 2.1.2) imply that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ there exists a unique measurable $v_{d,\varepsilon}: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that $\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathbf{w}_d(z)|^{-1} |(v_{d,\varepsilon}(z))(b)| < \infty$, $\int_{\mathbb{R}^d} |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, v_{d,\varepsilon}(y))| \rho_{d,\varepsilon}^{(a)}(x, dy) < \infty$, and

$$(v_{d,\varepsilon}(x))(a) = \int_{\mathbb{R}^d} (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, v_{d,\varepsilon}(y)) \rho_{d,\varepsilon}^{(a)}(x, dy). \quad (4.138)$$

This and the assumption that for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$, $Y \in \mathcal{B}(\mathbb{R}^d)$ it holds that $\rho_{d,\varepsilon}^{(a)}(x, Y) = \mathbb{P}[(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{\theta,a}))(x) \in Y]$ ensure for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ that

$$\begin{aligned} (v_{d,\varepsilon}(x))(a) &= \int_{\mathbb{R}^d} (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, v_{d,\varepsilon}(y)) \rho_{d,\varepsilon}^{(a)}(x, dy) \\ &= \mathbb{E}\left[(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))\left(\left(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{0,a})\right)(x), v_{d,\varepsilon}\left(\left(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{0,a})\right)(x)\right)\right)\right]. \end{aligned} \quad (4.139)$$

Moreover, observe that the triangle inequality demonstrates for all $d, n \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} &\left(\mathbb{E}\left[\max_{a \in A} |(u_d(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2\right]\right)^{1/2} \\ &\leq \left(\mathbb{E}\left[\left(\max_{a \in A} |(u_d(x))(a) - (v_{d,\varepsilon}(x))(a)| + \max_{b \in A} |(v_{d,\varepsilon}(x))(b) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(b)|\right)^2\right]\right)^{1/2} \\ &\leq \left(\mathbb{E}\left[\max_{a \in A} |(u_d(x))(a) - (v_{d,\varepsilon}(x))(a)|^2\right]\right)^{1/2} + \left(\mathbb{E}\left[\max_{a \in A} |(v_{d,\varepsilon}(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2\right]\right)^{1/2} \\ &= \max_{a \in A} |(u_d(x))(a) - (v_{d,\varepsilon}(x))(a)| + \left(\mathbb{E}\left[\max_{a \in A} |(v_{d,\varepsilon}(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2\right]\right)^{1/2}. \end{aligned} \quad (4.140)$$

Hence, it holds for all $d, n \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$ that

$$\begin{aligned} &\mathbb{E}\left[\max_{a \in A} |(u_d(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2\right] \\ &\leq \left(\max_{a \in A} |(u_d(x))(a) - (v_{d,\varepsilon}(x))(a)| + \left(\mathbb{E}\left[\max_{a \in A} |(v_{d,\varepsilon}(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2\right]\right)^{1/2}\right)^2 \\ &= |\mathbf{w}_d(x)|^2 \left(\max_{a \in A} \frac{|(u_d(x))(a) - (v_{d,\varepsilon}(x))(a)|}{|\mathbf{w}_d(x)|} + \left(\frac{\mathbb{E}\left[\max_{a \in A} |(v_{d,\varepsilon}(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2\right]}{|\mathbf{w}_d(x)|^2}\right)^{1/2}\right)^2. \end{aligned} \quad (4.141)$$

Furthermore, note that the assumptions that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, y \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|f_d(x, r) - f_d(x, s)| \leq L_d \max_{b \in A} |r(b) - s(b)|$, $|f_d(x, r) - f_d(y, r)| \leq K_d \|x - y\|$, $|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, s)| \leq \ell_{d,\varepsilon} \max_{b \in A} |r(b) - s(b)|$, and $\mathbb{W}_d(\kappa_d^{(a)}(x, \cdot), \kappa_d^{(a)}(y, \cdot)) \leq \eta_d \|x - y\|$, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $(\int_{\mathbb{R}^d} |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, 0)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy))^{1/2} \leq \mathbf{cd}^c |\mathbf{w}_d(x)|$, (4.137), (4.139), and item (ii) in Proposition 4.1.8 (applied for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ with $A \curvearrowright A$, $\eta \curvearrowright \eta_d$, $c \curvearrowright \lambda_d$, $K \curvearrowright K_d$, $L_1 \curvearrowright \ell_{d,\varepsilon}$, $L_2 \curvearrowright L_d$, $\mathbf{w} \curvearrowright \mathbf{w}_d$, $f_1 \curvearrowright \mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon})$, $f_2 \curvearrowright f_d$, $(\kappa_1^{(a)})_{a \in A} \curvearrowright (\rho_{d,\varepsilon}^{(a)})_{a \in A}$, $(\kappa_2^{(a)})_{a \in A} \curvearrowright (\kappa_d^{(a)})_{a \in A}$, $u_1 \curvearrowright v_{d,\varepsilon}$, $u_2 \curvearrowright u_d$ in the notation of Proposition 4.1.8) yield for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\begin{aligned} \sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(v_{d,\varepsilon}(x))(a) - (u_d(x))(a)|}{|\mathbf{w}_d(x)|} &\leq \frac{\lambda_d}{1 - \lambda_d(\min\{L_d, \ell_{d,\varepsilon}\})} \left[\sup_{(y,r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, r) - f_d(y, r)|}{|\mathbf{w}_d(y)|} \right] \\ &+ \frac{K_d}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\rho_{d,\varepsilon}^{(b)}(y, \cdot), \kappa_d^{(b)}(y, \cdot))}{|\mathbf{w}_d(y)|} \right]. \end{aligned} \quad (4.142)$$

Next, note that the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\lambda_d \ell_{d,\varepsilon} < 1$, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $(\int_{\mathbb{R}^d} |\mathbf{w}_d(y)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy))^{1/2} \leq \lambda_d |\mathbf{w}_d(x)|$ and $(\int_{\mathbb{R}^d} |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, 0)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy))^{1/2} \leq \mathbf{cd}^c |\mathbf{w}_d(x)|$, the fact that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathbf{w}_d(x)|^{-1} |(v_{d,\varepsilon}(x))(a)| < \infty$, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that $|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, s)| \leq \ell_{d,\varepsilon} \max_{b \in A} |r(b) - s(b)|$, (4.139), and Lemma 2.1.4 (applied for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ with $c_f \curvearrowright \mathbf{cd}^c$, $c_{\mathbf{w}} \curvearrowright \lambda_d$, $L \curvearrowright \ell_{d,\varepsilon}$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $(\Omega, \mathcal{F}, \mathbb{P}) \curvearrowright (\Omega, \mathcal{F}, \mathbb{P})$, $A \curvearrowright A$, $X \curvearrowright (\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{0,a}))_{a \in A}$, $f \curvearrowright \mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon})$, $\mathbf{w} \curvearrowright [\mathbb{R}^d \ni x \mapsto [A \ni a \mapsto \mathbf{w}_d(x) \in (0, \infty)]] \in (0, \infty)^{|A|}$, $v \curvearrowright v_{d,\varepsilon}$ in the notation of Lemma 2.1.4) prove that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that

$$\begin{aligned} \sup_{(x,a) \in \mathbb{R}^d \times A} \frac{|(v_{d,\varepsilon}(x))(a)|}{|\mathbf{w}_d(x)|} &\leq \frac{\mathbf{cd}^c}{1 - \lambda_d \ell_{d,\varepsilon}} \quad \text{and} \\ \sup_{(x,a) \in \mathbb{R}^d \times A} \frac{(\mathbb{E}[\max_{b \in A} |(v_{d,\varepsilon}(\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{0,a}))(x)))(b)|^2])^{1/2}}{|\mathbf{w}_d(x)|} &\leq \frac{\lambda_d \mathbf{cd}^c}{1 - \lambda_d \ell_{d,\varepsilon}}. \end{aligned} \quad (4.143)$$

Combining this, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $(\int_{\mathbb{R}^d} |\mathbf{w}_d(y)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy))^{1/2} \leq \lambda_d |\mathbf{w}_d(x)|$ and $(\int_{\mathbb{R}^d} |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y, 0)|^2 \rho_{d,\varepsilon}^{(a)}(x, dy))^{1/2} \leq \mathbf{cd}^c |\mathbf{w}_d(x)|$, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$ it holds that

$$|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, s)| \leq \ell_{d,\varepsilon} \max_{b \in A} |r(b) - s(b)|, \quad (4.144)$$

(4.139), and Proposition 2.2.8 (applied for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ with $M \curvearrowright M$, $\Theta \curvearrowright \Theta$, $(\mathbb{X}, \mathcal{X}) \curvearrowright (\mathbb{R}^d, \mathcal{B}(\mathbb{R}^d))$, $A \curvearrowright A$, $\mathbf{w} \curvearrowright \mathbf{w}_d$, $f \curvearrowright \mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon})$, $v \curvearrowright v_{d,\varepsilon}$, $(\mathbb{F}^\theta)_{\theta \in \Theta} \curvearrowright (\mathcal{F}_{d,\varepsilon}^\theta)_{\theta \in \Theta}$, $(X^\theta)_{\theta \in \Theta} \curvearrowright ((\mathcal{R}_\phi(\mathbf{X}_{d,\varepsilon}^{\theta,a}))_{a \in A})_{\theta \in \Theta}$, $(V_n^\theta)_{(\theta,n) \in \Theta \times \mathbb{N}_0} \curvearrowright (\mathcal{V}_{d,\varepsilon,n}^\theta)_{(\theta,n) \in \Theta \times \mathbb{N}_0}$, $c \curvearrowright \gamma_{d,\varepsilon}$, $c_{\mathbf{w}} \curvearrowright \lambda_d$, $L \curvearrowright \ell_{d,\varepsilon}$, $c_f \curvearrowright \mathbf{cd}^c$, $c_v \curvearrowright \frac{\lambda_d \mathbf{cd}^c}{1 - \lambda_d \ell_{d,\varepsilon}}$ in the notation of Proposition 2.2.8) demonstrates for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $n \in \mathbb{N}$, $x \in \mathbb{R}^d$ that

$$\left(\frac{\mathbb{E}[\max_{a \in A} |(v_{d,\varepsilon}(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2]}{|\mathbf{w}_d(x)|^2} \right)^{1/2} \leq d^c \gamma_{d,\varepsilon} \alpha_{d,\varepsilon}^n \leq d^c \gamma \alpha^n. \quad (4.145)$$

This, (4.141), and (4.142), ensure for every $d, n \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure

$\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^\mathbf{c}$ that

$$\begin{aligned}
& \left(\int_{\mathbb{R}^d} \mathbb{E} \left[\max_{a \in A} |(u_d(x))(a) - (\mathcal{V}_{d,\varepsilon,n}^0(x))(a)|^2 \right] \mu(dx) \right)^{1/2} \\
& \leq \left(d^\mathbf{c} \gamma \alpha^n + \frac{\lambda_d}{1 - \lambda_d(\min\{L_d, \ell_{d,\varepsilon}\})} \left[\sup_{(y,r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y,r) - f_d(y,r)|}{|\mathbf{w}(y)|} \right] \right. \\
& \quad \left. + \frac{K_d}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\rho_{d,\varepsilon}^{(b)}(y, \cdot), \kappa_d^{(b)}(y, \cdot))}{|\mathbf{w}_d(y)|} \right] \right) \left(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx) \right)^{1/2} \\
& \leq \left(d^\mathbf{c} \gamma \alpha^n + \frac{\lambda_d}{1 - \lambda_d(\min\{L_d, \ell_{d,\varepsilon}\})} \left[\sup_{(y,r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y,r) - f_d(y,r)|}{|\mathbf{w}(y)|} \right] \right. \\
& \quad \left. + \frac{K_d}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\rho_{d,\varepsilon}^{(b)}(y, \cdot), \kappa_d^{(b)}(y, \cdot))}{|\mathbf{w}_d(y)|} \right] \right) \mathbf{c}d^\mathbf{c}.
\end{aligned} \tag{4.146}$$

Combining this, the assumption that $\mathbf{c} \in [1, \infty)$, the assumptions that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$, $r \in \mathbb{R}^{|A|}$ it holds that $|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - f_d(x, r)| \leq \varepsilon \mathbf{c}d^\mathbf{c} |\mathbf{w}_d(x)|$ and $\mathbb{W}_d(\rho_{d,\varepsilon}^{(a)}(x, \cdot), \kappa_d^{(a)}(x, \cdot)) \leq \varepsilon \mathbf{c}d^\mathbf{c} |\mathbf{w}_d(x)|$, and the assumption that for all $d \in \mathbb{N}$ it holds that $\max\{\lambda_d, K_d\} \leq \mathbf{c}d^\mathbf{c}$ establishes for every $d \in \mathbb{N}$, $\delta, \varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^\mathbf{c}$ that

$$\begin{aligned}
& \left(\int_{\mathbb{R}^d} \mathbb{E} \left[\max_{a \in A} |(u_d(x))(a) - (\mathcal{V}_{d,\varepsilon,n(\delta)}^0(x))(a)|^2 \right] \mu(dx) \right)^{1/2} \\
& \leq \mathbf{c}d^\mathbf{c} \left(d^\mathbf{c} \gamma \alpha^{n(\delta)} + \frac{\lambda_d}{1 - \lambda_d(\min\{L_d, \ell_{d,\varepsilon}\})} \left[\sup_{(y,r) \in \mathbb{R}^d \times \mathbb{R}^{|A|}} \frac{|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(y,r) - f_d(y,r)|}{|\mathbf{w}(y)|} \right] \right. \\
& \quad \left. + \frac{K_d}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \left[\sup_{(y,b) \in \mathbb{R}^d \times A} \frac{\mathbb{W}_d(\rho_{d,\varepsilon}^{(b)}(y, \cdot), \kappa_d^{(b)}(y, \cdot))}{|\mathbf{w}_d(y)|} \right] \right) \\
& \leq \mathbf{c}d^\mathbf{c} \left(d^\mathbf{c} \delta + \frac{\varepsilon \mathbf{c}^2 d^{2\mathbf{c}}}{1 - \lambda_d(\min\{L_d, \ell_{d,\varepsilon}\})} + \frac{\varepsilon \mathbf{c}^2 d^{2\mathbf{c}}}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \right) \\
& \leq \mathbf{c}^3 d^{3\mathbf{c}} \left(\delta + \frac{\varepsilon}{1 - \lambda_d(\min\{L_d, \ell_{d,\varepsilon}\})} + \frac{\varepsilon}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \right).
\end{aligned} \tag{4.147}$$

This and the fact that for all $k \in \mathbb{N}$, $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ it holds that $\|x\| \leq k^{1/2} \max_{i \in \{1, \dots, k\}} |x_i|$ ensure for every $d \in \mathbb{N}$, $\delta \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^\mathbf{c}$ that

$$\begin{aligned}
& \left(\int_{\mathbb{R}^d} \mathbb{E} \left[\|u_d(x) - \mathcal{V}_{d,\delta,n(\delta)}^0(x)\|^2 \right] \right)^{1/2} \\
& \leq |A|^{1/2} \left(\int_{\mathbb{R}^d} \mathbb{E} \left[\max_{a \in A} |(u_d(x))(a) - (\mathcal{V}_{d,\delta,n(\delta)}^0(x))(a)|^2 \right] \mu(dx) \right)^{1/2} \\
& \leq \delta |A|^{1/2} \mathbf{c}^3 d^{3\mathbf{c}} \left(1 + \frac{1}{1 - \lambda_d(\min\{L_d, \ell_{d,\delta}\})} + \frac{1}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \right) \\
& \leq \delta |A|^{1/2} \mathbf{c}^3 d^{3\mathbf{c}} \left(1 + \frac{1}{1 - \lambda_d L_d} + \frac{1}{(1 - \eta_d L_d)(1 - \lambda_d L_d)} \right).
\end{aligned} \tag{4.148}$$

Throughout the remainder of the proof let $(\delta_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \subseteq (0, \infty)$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\delta_{d,\varepsilon} = \frac{\varepsilon}{|A|^{1/2} \mathbf{c}^3 d^{3\mathbf{c}}} \left(1 + \frac{1}{(1 - \lambda_d L_d)} + \frac{1}{(1 - \lambda_d L_d)(1 - \eta_d L_d)} \right)^{-1}. \tag{4.149}$$

Note that the assumptions that for all $d \in \mathbb{N}$ it holds that $\lambda_d L_d < 1$ and $\eta_d L_d < 1$, the assumption that A is nonempty, and the assumption that $\mathfrak{c} \geq 1$ ensure that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\delta_{d,\varepsilon} \in (0, 1]$. Combining this with (4.148) demonstrates for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathfrak{c} d^{\mathfrak{c}}$ that

$$\begin{aligned} & \int_{\mathbb{R}^d} \mathbb{E} \left[\|u_d(x) - \mathcal{V}_{d,\delta_{d,\varepsilon},n(\delta_{d,\varepsilon})}^0(x)\|^2 \right] \mu(dx) \\ & \leq \left(\delta_{d,\varepsilon} |A|^{1/2} \mathfrak{c}^3 d^{3\mathfrak{c}} \left(1 + \frac{1}{1-\lambda_d L_d} + \frac{1}{(1-\eta_d L_d)(1-\lambda_d L_d)} \right) \right)^2 \leq \varepsilon^2. \end{aligned} \quad (4.150)$$

This and Fubini's theorem establish for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathfrak{c} d^{\mathfrak{c}}$ that

$$\begin{aligned} & \mathbb{E} \left[\int_{\mathbb{R}^d} \|u_d(x) - \mathcal{V}_{d,\delta_{d,\varepsilon},n(\delta_{d,\varepsilon})}^0(x)\|^2 \mu(dx) \right] \\ & = \int_{\mathbb{R}^d} \mathbb{E} \left[\|u_d(x) - \mathcal{V}_{d,\delta_{d,\varepsilon},n(\delta_{d,\varepsilon})}^0(x)\|^2 \right] \mu(dx) \leq \varepsilon^2. \end{aligned} \quad (4.151)$$

Hence, for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathfrak{c} d^{\mathfrak{c}}$ there exists $\omega_{d,\varepsilon,\mu} \in \Omega$, which is assumed to be fixed for the remainder of this proof, such that

$$\int_{\mathbb{R}^d} \|u_d(x) - \mathcal{V}_{d,\delta_{d,\varepsilon},n(\delta_{d,\varepsilon})}(x, \omega_{d,\varepsilon,\mu})\|^2 \mu(dx) \leq \varepsilon^2. \quad (4.152)$$

Next, observe that Lemma 4.2.4 (applied for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $\omega \in \Omega$ with $d \curvearrowright d$, $\mathfrak{d} \curvearrowright \mathfrak{d}$, $M \curvearrowright M$, $\Theta \curvearrowright \Theta$, $A \curvearrowright A$, $\mathfrak{a} \curvearrowright \mathfrak{a}$, $\phi \curvearrowright \phi$, $\mathfrak{J} \curvearrowright \mathfrak{J}$, $\mathbf{F} \curvearrowright \mathbf{F}_{d,\varepsilon}$, $(\mathbf{X}_a^\theta)_{(\theta,a) \in \Theta \times A} \curvearrowright (\mathbf{X}_{d,\varepsilon}^{\theta,a}(\omega))_{(\theta,a) \in \Theta \times A}$, $(\mathcal{U}_n^\theta)_{(\theta,n) \in \Theta \times \mathbb{N}_0} \curvearrowright (\mathcal{V}_{d,\varepsilon,n}^\theta(\cdot, \omega))_{(\theta,n) \in \Theta \times \mathbb{N}_0}$ in the notation of Lemma 4.2.4) establishes that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $\omega \in \Omega$ there exists $\Phi_{d,\varepsilon,n}^\theta(\omega) \in \mathbf{N}$, $\theta \in \Theta$, $n \in \mathbb{N}_0$ such that

- (I) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that $\mathcal{R}_\phi(\Phi_{d,\varepsilon,n}^\theta(\omega)) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$,
- (II) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$, $x \in \mathbb{R}^d$ that $(\mathcal{R}_\phi(\Phi_{d,\varepsilon,n}^\theta(\omega)))(x) = \mathcal{V}_{d,\varepsilon,n}^\theta(x, \omega)$,
- (III) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that $\mathcal{D}(\Phi_{d,\varepsilon,n}^\theta(\omega)) = \mathcal{D}(\Phi_{d,\varepsilon,n}^0(\omega))$,
- (IV) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that $\mathcal{L}(\Phi_{d,\varepsilon,n}^\theta(\omega)) \leq n \left(\mathcal{L}(\mathbf{F}_{d,\varepsilon}) + \mathcal{L}(\mathbf{X}_{d,\varepsilon}^{0,\mathfrak{a}}(\omega)) - 1 \right) + 1$,
- (V) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that

$$|\mathcal{D}(\Phi_{d,\varepsilon,n}^\theta(\omega))| \leq 2 \max\{|\mathcal{D}(\mathbf{F}_{d,\varepsilon})|, |\mathcal{D}(\mathbf{X}_{d,\varepsilon}^{0,\mathfrak{a}}(\omega))|, \mathfrak{d}d + \mathfrak{d}|A|\} (4|A|M)^{2n}, \quad (4.153)$$

and

- (VI) it holds for all $\theta \in \Theta$, $n \in \mathbb{N}_0$ that

$$\begin{aligned} \mathcal{P}(\Phi_{d,\varepsilon,n}^\theta(\omega)) & \leq 2(n(\mathcal{L}(\mathbf{F}_{d,\varepsilon}) + \mathcal{L}(\mathbf{X}_{d,\varepsilon}^{0,\mathfrak{a}}(\omega)) - 1) + 1) \\ & \cdot (2 \max\{|\mathcal{D}(\mathbf{F}_{d,\varepsilon})|, |\mathcal{D}(\mathbf{X}_{d,\varepsilon}^{0,\mathfrak{a}}(\omega))|, \mathfrak{d}d + \mathfrak{d}|A|\})^2 (4|A|M)^{2n}. \end{aligned} \quad (4.154)$$

For every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathfrak{c} d^{\mathfrak{c}}$ let $\Psi_{d,\varepsilon,\mu} \in \mathbf{N}$ satisfy that

$$\Psi_{d,\varepsilon,\mu} = \Phi_{d,\delta_{d,\varepsilon},n(\delta_{d,\varepsilon})}^0(\omega_{d,\varepsilon,\mu}). \quad (4.155)$$

Combining this, (4.152), and items (I) and (II) demonstrates for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $y \in \mathbb{R}^d$, $b \in A$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^{\mathbf{c}}$ that $\mathcal{R}_\phi(\Psi_{d,\varepsilon,\mu}) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, $((\mathcal{R}_\phi(\Psi_{d,\varepsilon,\mu}))(y))(b) = (\mathcal{V}_{d,\delta_{d,\varepsilon},\mathbf{n}(\delta_{d,\varepsilon})}^0(y, \omega_{d,\varepsilon,\mu}))(b)$, and

$$\begin{aligned} & \left(\int_{\mathbb{R}^d} \|u_d(x) - (\mathcal{R}_\phi(\Psi_{d,\varepsilon,\mu}))(x)\|^2 \mu(dx) \right)^{1/2} \\ & \leq \left(\int_{\mathbb{R}^d} \|u_d(x) - \mathcal{V}_{d,\delta_{d,\varepsilon},\mathbf{n}(\delta_{d,\varepsilon})}(x, \omega_{d,\varepsilon,\mu})\|^2 \mu(dx) \right)^{1/2} \leq \varepsilon. \end{aligned} \quad (4.156)$$

This establishes (4.133). Furthermore, note that (4.155), item (VI), the fact that for all $\varepsilon \in (0, 1]$ it holds that $\mathbf{n}(\varepsilon) \in \mathbb{N}$, and the assumption that for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$, $\omega \in \Omega$ it holds that $\mathcal{D}(\mathbf{X}_{d,\varepsilon}^{\theta,a}(\omega)) = \mathcal{D}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega}))$ yield that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^{\mathbf{c}}$ it holds that

$$\begin{aligned} \mathcal{P}(\Psi_{d,\varepsilon,\mu}) & \leq 2(\mathbf{n}(\delta_{d,\varepsilon})(\mathcal{L}(\mathbf{F}_{d,\delta_{d,\varepsilon}}) + \mathcal{L}(\mathbf{X}_{d,\delta_{d,\varepsilon}}^{0,a}(\bar{\omega})) - 1) + 1) \\ & \quad \cdot (2 \max\{|\mathcal{D}(\mathbf{F}_{d,\delta_{d,\varepsilon}})|, |\mathcal{D}(\mathbf{X}_{d,\delta_{d,\varepsilon}}^{0,a}(\bar{\omega}))|, \mathfrak{d}d + \mathfrak{d}|A|\})^2 (4|A|M)^{2\mathbf{n}(\delta_{d,\varepsilon})} \\ & \leq 8(\max\{|\mathcal{D}(\mathbf{F}_{d,\delta_{d,\varepsilon}})|, |\mathcal{D}(\mathbf{X}_{d,\delta_{d,\varepsilon}}^{0,a}(\bar{\omega}))|, \mathfrak{d}d + \mathfrak{d}|A|\})^2 \\ & \quad \cdot (\mathcal{L}(\mathbf{F}_{d,\delta_{d,\varepsilon}}) + \mathcal{L}(\mathbf{X}_{d,\delta_{d,\varepsilon}}^{0,a}(\bar{\omega}))) \mathbf{n}(\delta_{d,\varepsilon}) (4|A|M)^{2\mathbf{n}(\delta_{d,\varepsilon})}. \end{aligned} \quad (4.157)$$

Next, observe that for all $d \in \mathbb{N}$ it holds that $\mathfrak{d}d + \mathfrak{d}|A| \leq (\mathfrak{d} + \mathfrak{d}|A|)d \leq \mathfrak{K}d$. This, the assumption that $\max\{\mathbf{c}, \mathfrak{d}\} \geq 1$, the assumption that A is nonempty, and the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\max\{|\mathcal{L}(\mathbf{F}_{d,\varepsilon})|, |\mathcal{L}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega}))|, |\mathcal{D}(\mathbf{F}_{d,\varepsilon})|, |\mathcal{D}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega}))|\} \leq \mathbf{c}d^{\mathbf{c}}\varepsilon^{-\mathbf{c}}$ assure that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that

$$\begin{aligned} & (\mathcal{L}(\mathbf{F}_{d,\varepsilon}) + \mathcal{L}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega})))(\max\{|\mathcal{D}(\mathbf{F}_{d,\varepsilon})|, |\mathcal{D}(\mathbf{X}_{d,\varepsilon}^{0,a}(\bar{\omega}))|, \mathfrak{d}d + \mathfrak{d}|A|\})^2 \\ & \leq 2\mathbf{c}d^{\mathbf{c}}\varepsilon^{-\mathbf{c}}(\mathfrak{K}d^{\mathbf{c}}\varepsilon^{-\mathbf{c}})^2 \leq 2\mathfrak{K}^3 d^{3\mathbf{c}}\varepsilon^{-3\mathbf{c}}. \end{aligned} \quad (4.158)$$

Moreover, note that [6, Lemma 3.15] (applied with $m \curvearrowright 1$, $\alpha \curvearrowright \alpha$, $\beta \curvearrowright (4|A|M)^2$, $\kappa_1 \curvearrowright \gamma$, $\kappa_2 \curvearrowright 1$, $(c_n)_{n \in \mathbb{N}} \curvearrowright (((4|A|M)^2)^n)_{n \in \mathbb{N}}$, $(e_n)_{n \in \mathbb{N}} \curvearrowright (\gamma\alpha^n)_{n \in \mathbb{N}}$, $N \curvearrowright \mathbf{n}$ in the notation of [6, Lemma 3.15]) and the fact that $\gamma \in [3/2, \infty)$ yield for all $\varepsilon \in (0, 1]$ that

$$\begin{aligned} ((4|A|M)^2)^{\mathbf{n}(\varepsilon)} & \leq (4|A|M)^2 \max\left\{1, \gamma^{\frac{2\ln(4|A|M)}{\ln(1/\alpha)}}\right\} \varepsilon^{-\frac{2\ln(4|A|M)}{\ln(1/\alpha)}} \\ & = (4|A|M)^2 \gamma^{\frac{2\ln(4|A|M)}{\ln(1/\alpha)}} \varepsilon^{-\frac{2\ln(4|A|M)}{\ln(1/\alpha)}}. \end{aligned} \quad (4.159)$$

Furthermore, observe that the fact that for all $\varepsilon \in (0, 1]$ it holds that $\mathbf{n}(\varepsilon) = \min\{n \in \mathbb{N} : \gamma\alpha^n \leq \varepsilon\}$ implies for all $\varepsilon \in (0, 1]$ with $\mathbf{n}(\varepsilon) \in \mathbb{N} \cap [2, \infty)$ that $\gamma\alpha^{\mathbf{n}(\varepsilon)-1} > \varepsilon$. Hence, it holds for all $\varepsilon \in (0, 1]$ with $\mathbf{n}(\varepsilon) \in \mathbb{N} \cap [2, \infty)$ that $\mathbf{n}(\varepsilon) < \frac{\ln(\varepsilon^{-1}) + \ln(\gamma)}{\ln(1/\alpha)} + 1$. This, the fact that $\alpha \in (0, 1)$, and the fact that $\gamma \in [3/2, \infty)$ establish that for all $\varepsilon \in (0, 1]$ it holds that

$$\mathbf{n}(\varepsilon) \leq \left(\frac{\ln(1/\alpha) + \ln(\gamma) + 1}{\ln(1/\alpha)} \right) \varepsilon^{-1}. \quad (4.160)$$

Combining this, (4.157), (4.158), and (4.159) ensures for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^{\mathbf{c}}$ that

$$\begin{aligned} \mathcal{P}(\Psi_{d,\varepsilon,\mu}) & \leq 16\mathfrak{K}^3 d^{3\mathbf{c}} \delta_{d,\varepsilon}^{-3\mathbf{c}} \left(\frac{\ln(1/\alpha) + \ln(\gamma) + 1}{\ln(1/\alpha)} \right) \delta_{d,\varepsilon}^{-1} (4|A|M)^2 \gamma^{\frac{2\ln(4|A|M)}{\ln(1/\alpha)}} \delta_{d,\varepsilon}^{-\frac{2\ln(4|A|M)}{\ln(1/\alpha)}} \\ & = 256\mathfrak{K}^3 (|A|M)^2 \gamma^{\frac{2\ln(4|A|M)}{\ln(1/\alpha)}} \left(\frac{\ln(1/\alpha) + \ln(\gamma) + 1}{\ln(1/\alpha)} \right) d^{3\mathbf{c}} \delta_{d,\varepsilon}^{-\left(1+3\mathbf{c}+\frac{2\ln(4|A|M)}{\ln(1/\alpha)}\right)}. \end{aligned} \quad (4.161)$$

Moreover, observe the assumption that $\sup_{d \in \mathbb{N}} \lambda_d L_d < 1$, the assumption that $\sup_{d \in \mathbb{N}} \eta_d L_d < 1$, and (4.149) establish for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\delta_{d,\varepsilon}^{-1} \leq \frac{3|A|^{1/2} \mathfrak{c}^3}{(1 - \sup_{u \in \mathbb{N}} \lambda_u L_u)(1 - \sup_{u \in \mathbb{N}} \eta_u L_u)} d^{3\mathfrak{c}} \varepsilon^{-1}. \quad (4.162)$$

Combining this with (4.161) demonstrates for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathfrak{c} d^{\mathfrak{c}}$ that

$$\begin{aligned} \mathcal{P}(\Psi_{d,\varepsilon,\mu}) &\leq 256 \mathfrak{K}^3 (|A|M)^2 \gamma^{\frac{2 \ln(4|A|M)}{\ln(1/\alpha)}} \left(\frac{\ln(1/\alpha) + \ln(\gamma) + 1}{\ln(1/\alpha)} \right) \\ &\cdot \left(\frac{3|A|^{1/2} \mathfrak{c}^3}{(1 - \sup_{u \in \mathbb{N}} \lambda_u L_u)(1 - \sup_{u \in \mathbb{N}} \eta_u L_u)} \right)^{\left(1 + 3\mathfrak{c} + \frac{2 \ln(4|A|M)}{\ln(1/\alpha)}\right)} d^{3\mathfrak{c} + 3\mathfrak{c} \left(1 + 3\mathfrak{c} + \frac{2 \ln(4|A|M)}{\ln(1/\alpha)}\right)} \\ &\cdot \varepsilon^{-\left(1 + 3\mathfrak{c} + \frac{2 \ln(4|A|M)}{\ln(1/\alpha)}\right)}. \end{aligned} \quad (4.163)$$

This proves item (ii). The proof of Theorem 4.3.1 is thus complete. $\square$

4.3.2. ANN Approximations for Functional Fixed-Point Equations with Specific Activation Functions and One-Step Transitions

Lemma 4.3.2. *Let $d, m \in \mathbb{N}$, let $\eta \in [0, \infty)$, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\xi: \Omega \rightarrow \mathbb{R}^m$ be measurable, let $\mathfrak{t}_i: \mathbb{R}^d \times \mathbb{R}^m \rightarrow \mathbb{R}^d$, $i \in \{1, 2\}$, be measurable, assume that for all $x, y \in \mathbb{R}^d$, $p \in \mathbb{R}^m$ it holds that $\|\mathfrak{t}_1(x, p) - \mathfrak{t}_1(y, p)\| \leq \eta \|x - y\|$, let $\kappa_i: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $i \in \{1, 2\}$, satisfy for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, $B \in \mathcal{B}(\mathbb{R}^d)$ that $\kappa_i(x, B) = \mathbb{P}[\mathfrak{t}_i(x, \xi) \in B]$, and assume for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$ that $\mathbb{E}[\|\mathfrak{t}_i(x, \xi)\|] < \infty$. Then*

(i) *it holds for all $x \in \mathbb{R}^d$ that*

$$\mathbb{W}_d(\kappa_1(x, \cdot), \kappa_2(x, \cdot)) \leq \int_{\mathbb{R}^m} \|\mathfrak{t}_1(x, p) - \mathfrak{t}_2(x, p)\| (\mathbb{P} \circ \xi^{-1})(dp) \quad (4.164)$$

and

(ii) *it holds for all $x, y \in \mathbb{R}^d$ that*

$$\mathbb{W}_d(\kappa_1(x, \cdot), \kappa_1(y, \cdot)) \leq \eta \|x - y\|. \quad (4.165)$$

Proof of Lemma 4.3.2. Let $X_i: \mathbb{R}^d \times \Omega \rightarrow \mathbb{R}^d$, $i \in \{1, 2\}$, satisfy for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$ that $X_i^x = \mathfrak{t}_i(x, \xi)$. Note that the assumption that for all $x \in \mathbb{R}^d$, $i \in \{1, 2\}$ it holds that $\mathbb{E}[\|\mathfrak{t}_i(x, \xi)\|] < \infty$ ensures for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$ that $\kappa_i(x, \cdot) \in \mathcal{M}_d$. Observe that for all $i \in \{1, 2\}$, $x \in \mathbb{R}^d$, it holds that $X_i^x \sim \kappa_i(x, \cdot)$. Furthermore, it holds for all $x \in \mathbb{R}^d$ that $\mathbb{P} \circ (X_1^x, X_2^x)^{-1} \in \Pi_d(\kappa_1(x, \cdot), \kappa_2(x, \cdot))$. This demonstrates for all $x \in \mathbb{R}^d$ that

$$\begin{aligned} \mathbb{W}_d(\kappa_1(x, \cdot), \kappa_2(x, \cdot)) &= \inf_{\gamma \in \Pi_d(\kappa_1(x, \cdot), \kappa_2(x, \cdot))} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \\ &\leq \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| (\mathbb{P} \circ (X_1^x, X_2^x)^{-1})(d(\mathbf{x}, \mathbf{y})) \\ &= \int_{\Omega} \|X_1^x(\omega) - X_2^x(\omega)\| \mathbb{P}[d\omega] \\ &= \int_{\Omega} \|\mathfrak{t}_1(x, \xi(\omega)) - \mathfrak{t}_2(x, \xi(\omega))\| \mathbb{P}[d\omega] \\ &= \int_{\mathbb{R}^m} \|\mathfrak{t}_1(x, p) - \mathfrak{t}_2(x, p)\| (\mathbb{P} \circ \xi^{-1})(dp). \end{aligned} \quad (4.166)$$

This proves item (i). Moreover, it holds for all $x, y \in \mathbb{R}^d$ that $\mathbb{P} \circ (X_1^x, X_1^y)^{-1} \in \Pi_d(\kappa_1(x, \cdot), \kappa_1(y, \cdot))$. Combining this with the assumption that for all $x, y \in \mathbb{R}^d$, $p \in \mathbb{R}^m$ it holds that $\|\mathbf{t}_1(x, p) - \mathbf{t}_1(y, p)\| \leq \eta\|x - y\|$ yields for all $x, y \in \mathbb{R}^d$ that

$$\begin{aligned} \mathbb{W}_d(\kappa_1(x, \cdot), \kappa_1(y, \cdot)) &= \inf_{\gamma \in \Pi_d(\kappa_1(x, \cdot), \kappa_1(y, \cdot))} \int_{\mathbb{R}^d \times \mathbb{R}^d} \|\mathbf{x} - \mathbf{y}\| \gamma(d(\mathbf{x}, \mathbf{y})) \\ &\leq \int_{\mathbb{R}^m} \|\mathbf{t}_1(x, p) - \mathbf{t}_1(y, p)\| (\mathbb{P} \circ \xi^{-1})(dp) \\ &\leq \eta\|x - y\|. \end{aligned} \quad (4.167)$$

This proves item (ii). The proof of Lemma 4.3.2 is thus complete. $\square$

Corollary 4.3.3. *Let A be a nonempty finite set, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\mathbf{c} \in [1, \infty)$, let $\nu \in \{0, 1\}$, let $\beta \in [0, \infty) \setminus \{1\}$, for every $d \in \mathbb{N}$ let $\lambda_d, \eta_d, K_d, L_d \in (0, \infty)$, for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ let $\ell_{d,\varepsilon} \in (0, \infty)$, assume $\sup_{d \in \mathbb{N}} \lambda_d L_d < 1$, $\sup_{d \in \mathbb{N}} \eta_d L_d < 1$, $\sup_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \lambda_d \ell_{d,\varepsilon} < 1$, for every $d \in \mathbb{N}$ let $\mathbf{w}_d: \mathbb{R}^d \rightarrow (0, \infty)$ be measurable, for every $d \in \mathbb{N}$, $x \in (x_1, \dots, x_d) \in \mathbb{R}^d$ let $\|\mathbf{x}\| \in \mathbb{R}$ satisfy $\|\mathbf{x}\| = \max_{i \in \{1, \dots, d\}} |x_i|$, for every $d \in \mathbb{N}$ let $f_d: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ be measurable, for every $d \in \mathbb{N}$, $a \in A$ let $t_d^{(a)}: \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ be measurable, for every $d \in \mathbb{N}$ let $\xi_d: \Omega \rightarrow \mathbb{R}^d$ be measurable, let $a \in A$, let $\phi: \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}$ that $\phi(x) = \nu \ln(1 + \exp(x)) + (1 - \nu) \max\{x, \beta x\}$, let $(\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \subseteq \mathbf{N}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}) \in C(\mathbb{R}^{d+|A|}, \mathbb{R})$, let $(\mathbf{t}_{d,\varepsilon}^{(a)})_{(d,\varepsilon,a) \in \mathbb{N} \times (0,1] \times A} \subseteq \mathbf{N}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ that $\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(a)}) = \mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(a)})$ and $\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}) \in C(\mathbb{R}^{2d}, \mathbb{R}^d)$, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $\max\{\lambda_d, K_d\} \leq \mathbf{c}d^{\mathbf{c}}$ and*

$$\max\left\{\mathcal{L}(\mathbf{F}_{d,\varepsilon}), \mathcal{L}(\mathbf{t}_{d,\varepsilon}^{(a)}), \|\mathcal{D}(\mathbf{F}_{d,\varepsilon})\|, \|\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(a)})\|\right\} \leq \mathbf{c}d^{\mathbf{c}}\varepsilon^{-\mathbf{c}}, \quad (4.168)$$

assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, y, p \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$, $a \in A$ that

$$\begin{aligned} |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - f_d(x, r)| &\leq \varepsilon \mathbf{c}d^{\mathbf{c}}|\mathbf{w}_d(x)|, \\ \|(\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, p) - t_d^{(a)}(x, p)\| &\leq \varepsilon \mathbf{c}d^{\mathbf{c}}|\mathbf{w}_d(x)|, \end{aligned} \quad (4.169)$$

$$\begin{aligned} |f_d(x, r) - f_d(x, s)| &\leq L_d \max_{b \in A} |r(b) - s(b)|, \quad |f_d(x, r) - f_d(y, r)| \leq K_d \|x - y\|, \\ |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, s)| &\leq \ell_{d,\varepsilon} \max_{b \in A} |r(b) - s(b)|, \quad \text{and} \\ \|t_d^{(a)}(x, p) - t_d^{(a)}(y, p)\| &\leq \eta_d \|x - y\|, \end{aligned} \quad (4.170)$$

assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ that

$$\sup_{x \in \mathbb{R}^d} |\mathbf{w}_d(x)|^{-1} \mathbb{E}[\|t_d^{(a)}(x, \xi_d)\|] < \infty \quad \text{and} \quad \sup_{x \in \mathbb{R}^d} |\mathbf{w}_d(x)|^{-1} \mathbb{E}[\|(\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d)\|] < \infty, \quad (4.171)$$

assume for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $a \in A$ that

$$\left(\mathbb{E}\left[\left|\mathbf{w}_d\left((\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d)\right)\right|^2\right]\right)^{1/2} \leq \lambda_d |\mathbf{w}_d(x)|, \quad (4.172)$$

$$\begin{aligned} \left(\mathbb{E}\left[\left|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))((\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d), 0)\right|^2\right]\right)^{1/2} &\leq \mathbf{c}d^{\mathbf{c}}|\mathbf{w}_d(x)|, \\ \mathbb{E}\left[|\mathbf{w}_d(t_d^{(a)}(x, \xi_d))|\right] &\leq \lambda_d |\mathbf{w}_d(x)|, \quad \text{and} \quad \sup_{\substack{y \in \mathbb{R}^d \\ b \in A}} |\mathbf{w}_d(y)|^{-1} \mathbb{E}\left[|f_d(t_d^{(b)}(y, \xi_d), 0)|\right] < \infty. \end{aligned} \quad (4.173)$$

Then there exists $c \in \mathbb{R}$ such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^{\mathbf{c}}$ it holds that

(i) there exists a unique measurable $u: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that $\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathbf{w}_d(z)|^{-1} |(u(z))(b)| < \infty$, $\mathbb{E} \left[|f_d(t_d^{(a)}(x, \xi_d), u(t_d^{(a)}(x, \xi_d)))| \right] < \infty$, and

$$(u(x))(a) = \mathbb{E} \left[f_d(t_d^{(a)}(x, \xi_d), u(t_d^{(a)}(x, \xi_d))) \right] \quad (4.174)$$

and

(ii) there exists $\Psi \in \mathbf{N}$ which satisfies $\mathcal{H}(\Psi) > 0$, $\mathcal{P}(\Psi) \leq cd^c \varepsilon^{-c}$, $\mathcal{R}_\phi(\Psi) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, and

$$\left(\int_{\mathbb{R}^d} \|u(x) - (\mathcal{R}_\phi(\Psi))(x)\|^2 \mu(dx) \right)^{1/2} \leq \varepsilon. \quad (4.175)$$

Proof of Corollary 4.3.3. Throughout this proof let $\Theta = \cup_{n \in \mathbb{N}} \mathbb{Z}^n$, for every $d \in \mathbb{N}$ let $\xi_d^\theta: \Omega \rightarrow \mathbb{R}^d$, $\theta \in \Theta$, be i.i.d. random variables, assume for all $d \in \mathbb{N}$, $\theta \in \Theta$ that ξ_d and ξ_d^θ are identically distributed, for every $d \in \mathbb{N}$, $\theta \in \Theta$ let $\mathbb{F}_d^\theta = \sigma(\xi_d^\theta)$, let $\bar{\omega} \in \Omega$, let $\mathbf{I}, \mathfrak{J} \in \mathbf{N}$ satisfy

$$\mathbf{I} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), \left((1 \ -1), 0 \right) \right) \in (\mathbb{R}^{2 \times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1 \times 2} \times \mathbb{R}^1) \quad \text{and} \quad (4.176)$$

$$\mathfrak{J} = \begin{cases} \mathbf{I} & : \nu = 1 \\ (1 + \beta)^{-1} \otimes \mathbf{I} & : \nu = 0, \end{cases} \quad (4.177)$$

and let $\Xi_d^\theta: \Omega \rightarrow \mathbf{N}$, $\theta \in \Theta$, $d \in \mathbb{N}$, satisfy for every $\theta \in \Theta$, $d \in \mathbb{N}$, $\omega \in \Omega$ that

$$\Xi_d^\theta(\omega) = \left(\begin{pmatrix} \mathbf{I}_d \\ 0 \end{pmatrix}, \begin{pmatrix} 0 \\ \xi_d^\theta(\omega) \end{pmatrix} \right) \in \mathbb{R}^{2d \times d} \times \mathbb{R}^{2d}. \quad (4.178)$$

First, observe that [1, Lemma 3.5] and [1, Lemma 3.8] ensure for all $x \in \mathbb{R}$ that $\mathcal{D}(\mathfrak{J}) = (1, 2, 1) \in \mathbb{N}^3$ and $(\mathcal{R}_\phi(\mathfrak{J}))(x) = x$. Next, note that (4.178) shows for all $d \in \mathbb{N}$, $\theta \in \Theta$, $x \in \mathbb{R}^d$, $\omega \in \Omega$ that $\mathcal{D}(\Xi_d^\theta(\omega)) = (d, 2d) = \mathcal{D}(\Xi_d^0(\bar{\omega})) \in \mathbb{N}^2$, $\mathcal{R}_\phi(\Xi_d^\theta(\omega)) \in C(\mathbb{R}^d, \mathbb{R}^{2d})$, and $(\mathcal{R}_\phi(\Xi_d^\theta(\omega)))(x) = (x, \xi_d^\theta(\omega))$. This, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ it holds that $\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(a)}) = \mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)})$, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\max\{\mathcal{L}(\mathbf{t}_{d,\varepsilon}^{(a)}), \|\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(a)})\|\} \leq cd^c \varepsilon^{-c}$, and Lemma 3.2.2 ensure for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$, $\omega \in \Omega$ that $\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(a)} \bullet \Xi_d^\theta(\omega)) = \mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)} \bullet \Xi_d^0(\bar{\omega}))$, $\mathcal{L}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)} \bullet \Xi_d^0(\bar{\omega})) = \mathcal{L}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)}) + \mathcal{L}(\Xi_d^0(\bar{\omega})) - 1 = \mathcal{L}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)})$, $\|\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)} \bullet \Xi_d^0(\bar{\omega}))\| \leq \max\{\|\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)})\|, \|\mathcal{D}(\Xi_d^0(\bar{\omega}))\|\} = \|\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)})\|$, $\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)} \bullet \Xi_d^0(\bar{\omega})) \in C(\mathbb{R}^d, \mathbb{R}^d)$, and $(\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)} \bullet \Xi_d^\theta(\omega)))(x) = (\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d^\theta(\omega))$. Hence, it holds for all $\theta \in \Theta$, $d \in \mathbb{N}$, $(0, 1] \ni \varepsilon$ that $\max\{\mathcal{L}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)} \bullet \Xi_d^0(\bar{\omega})), \|\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)} \bullet \Xi_d^0(\bar{\omega}))\|\} \leq cd^c \varepsilon^{-c}$. Moreover, observe that the assumption that for all $\theta \in \Theta$, $d \in \mathbb{N}$ it holds that $\mathbb{F}_d^\theta = \sigma(\xi_d^\theta)$ and the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ it holds that $\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(\varepsilon)}) \in C(\mathbb{R}^{2d}, \mathbb{R}^d)$ demonstrate for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ that $\mathbb{R}^d \times \Omega \ni (x, \omega) \mapsto (\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d^\theta(\omega)) = (\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)} \bullet \Xi_d^\theta(\omega)))(x) \in \mathbb{R}^d$ is $(\mathcal{B}(\mathbb{R}^d) \otimes \mathbb{F}_d^\theta) / \mathcal{B}(\mathbb{R}^d)$ -measurable. This implies for all $\theta \in \Theta$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $(\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)} \bullet \Xi_d^\theta))_{a \in A}: \mathbb{R}^d \times \Omega \rightarrow (\mathbb{R}^d)^{|A|}$ is $(\mathcal{B}(\mathbb{R}^d) \otimes \mathbb{F}_d^\theta) / \mathcal{B}((\mathbb{R}^d)^{|A|})$ -measurable. Next, note that the assumption that for all $d \in \mathbb{N}$ it holds that ξ_d^θ , $\theta \in \Theta$, are i.i.d. random variables assures that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $(\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)} \bullet \Xi_d^\theta))_{a \in A}$, $\theta \in \Theta$, are i.i.d. random fields. Throughout the remainder of this proof let $\kappa_d^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $d \in \mathbb{N}$, $a \in A$, satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$, $B \in \mathcal{B}(\mathbb{R}^d)$ that $\kappa_d^{(a)}(x, B) = \mathbb{P}[t_d^{(a)}(x, \xi_d^0) \in B]$, let $\rho_{d,\varepsilon}^{(a)}: \mathbb{R}^d \times \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$, $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$, satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$, $B \in \mathcal{B}(\mathbb{R}^d)$ that $\rho_{d,\varepsilon}^{(a)}(x, B) = \mathbb{P}[(\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d^0) \in B]$. Note that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ it holds that $\kappa_d^{(a)}$ and $\rho_{d,\varepsilon}^{(a)}$ are stochastic kernels. This and Lemma 4.3.2

(applied for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ with $d \curvearrowright d$, $m \curvearrowright d$, $\eta \curvearrowright \eta_d$, $(\Omega, \mathcal{F}, \mathbb{P}) \curvearrowright (\Omega, \mathcal{F}, \mathbb{P})$, $\xi \curvearrowright \xi_d^0$, $\mathbf{t}_1 \curvearrowright \mathbf{t}_d^{(a)}$, $\mathbf{t}_2 \curvearrowright \mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)})$ in the notation of Lemma 4.3.2) establishes that for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $a \in A$ it holds that $\mathbb{W}_d(\kappa_d^{(a)}(x, \cdot), \kappa_d^{(a)}(y, \cdot)) \leq \eta_d \|x - y\|$ and

$$\mathbb{W}_d(\kappa_d^{(a)}(x, \cdot), \rho_{d,\varepsilon}^{(a)}(x, \cdot)) \leq \int_{\mathbb{R}^d} \|t_d^{(a)}(x, p) - (\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, p)\| (\mathbb{P} \circ (\xi_d^0)^{-1})(dp). \quad (4.179)$$

This and the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, p \in \mathbb{R}^d$, $a \in A$ it holds that $\|t_d^{(a)}(x, p) - (\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, p)\| \leq \varepsilon c d^c |\mathbf{w}_d(x)|$ implies for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $a \in A$ that $\mathbb{W}_d(\kappa_d^{(a)}(x, \cdot), \rho_{d,\varepsilon}^{(a)}(x, \cdot)) \leq \varepsilon c d^c |\mathbf{w}_d(x)|$. Combining this with Theorem 4.3.1 (applied with

$$\begin{aligned} & \Theta \curvearrowright \Theta, \quad A \curvearrowright A, \quad (\Omega, \mathcal{F}, \mathbb{P}) \curvearrowright (\Omega, \mathcal{F}, \mathbb{P}), \quad \mathbf{c} \curvearrowright \mathbf{c}, \quad (\lambda_d, \eta_d, K_d, L_d)_{d \in \mathbb{N}} \curvearrowright (\lambda_d, \eta_d, K_d, L_d)_{d \in \mathbb{N}} \\ & (\ell_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \curvearrowright (\ell_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]}, \quad (\mathbf{w}_d)_{d \in \mathbb{N}} \curvearrowright (\mathbf{w}_d)_{d \in \mathbb{N}}, \quad (f_d)_{d \in \mathbb{N}} \curvearrowright (f_d)_{d \in \mathbb{N}}, \\ & (\kappa_d^{(a)})_{(d,a) \in \mathbb{N} \times A} \curvearrowright (\kappa_d^{(a)})_{(d,a) \in \mathbb{N} \times A}, \quad (\rho_{d,\varepsilon}^{(a)})_{(d,\varepsilon,a) \in \mathbb{N} \times (0,1] \times A} \curvearrowright (\rho_{d,\varepsilon}^{(a)})_{(d,\varepsilon,a) \in \mathbb{N} \times (0,1] \times A}, \\ & \mathbf{a} \curvearrowright \mathbf{a}, \quad \bar{\mathbf{w}} \curvearrowright \bar{\mathbf{w}}, \quad \phi \curvearrowright \phi, \quad \mathfrak{J} \curvearrowright \mathfrak{J}, \quad (\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \curvearrowright (\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]}, \\ & (\mathbf{X}_{d,\varepsilon}^{\theta,a})_{(\theta,d,\varepsilon,a) \in \Theta \times \mathbb{N} \times (0,1] \times A} \curvearrowright (\mathbf{t}_{d,\varepsilon}^{(a)} \bullet \Xi_d^\theta)_{(\theta,d,\varepsilon,a) \in \Theta \times \mathbb{N} \times (0,1] \times A}, \\ & (\mathcal{F}_{d,\varepsilon}^\theta)_{(\theta,d,\varepsilon) \in \Theta \times \mathbb{N} \times (0,1]} \curvearrowright (\mathbb{R}_d^\theta)_{(\theta,d,\varepsilon) \in \Theta \times \mathbb{N} \times (0,1]} \end{aligned} \quad (4.180)$$

in the notation of Theorem 4.3.1) yields that there exists $c \in \mathbb{R}$ such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq c d^c$ it holds that

- (I) there exists a unique measurable $u_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that $\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathbf{w}_d(z)|^{-1} |(u_d(z))(b)| < \infty$, $\int_{\mathbb{R}^d} f_d(y, u_d(y)) |\kappa_d^{(a)}(x, dy)| < \infty$, and

$$(u_d(x))(a) = \int_{\mathbb{R}^d} f_d(y, u_d(y)) \kappa_d^{(a)}(x, dy) \quad (4.181)$$

and

- (II) there exists $\Psi_{d,\varepsilon,\mu} \in \mathbf{N}$ which satisfies $\mathcal{H}(\Psi_{d,\varepsilon,\mu}) > 0$, $\mathcal{P}(\Psi_{d,\varepsilon,\mu}) \leq c d^c \varepsilon^{-c}$, $\mathcal{R}_\phi(\Psi_{d,\varepsilon,\mu}) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, and

$$\left(\int_{\mathbb{R}^d} \|u_d(x) - (\mathcal{R}_\phi(\Psi_{d,\varepsilon,\mu}))(x)\|^2 \mu(dx) \right)^{1/2} \leq \varepsilon. \quad (4.182)$$

Note that item (I) shows for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ that $\mathbb{E} \left[\left| f_d(t_d^{(a)}(x, \xi_d^0), u_d(t_d^{(a)}(x, \xi_d^0)) \right| \right] < \infty$ and

$$(u_d(x))(a) = \mathbb{E} \left[f_d(t_d^{(a)}(x, \xi_d^0), u_d(t_d^{(a)}(x, \xi_d^0)) \right]. \quad (4.183)$$

This and the assumption that for all $d \in \mathbb{N}$ it holds that ξ_d and ξ_d^0 are identically distributed prove item (i). Moreover, observe that item (II) establishes item (ii). The proof of Corollary 4.3.3 is thus complete. $\square$

4.3.3. ANN Approximations for Bellman Equations with Specific Activation Functions and One-Step Transitions

Corollary 4.3.4. *Let A be a nonempty finite set, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $\nu \in \{0, 1\}$, let $\beta \in [0, \infty) \setminus \{1\}$, let $\lambda, \eta, K, L \in (0, \infty)$, assume $L \max\{\lambda, \eta\} < 1$, let $\mathbf{c} \in [\max\{1, \lambda, K\}, \infty)$, for every $d \in \mathbb{N}$, $x \in (x_1, \dots, x_d) \in \mathbb{R}^d$ let $\|x\| \in \mathbb{R}$ satisfy $\|x\| =$*

$\max_{i \in \{1, \dots, d\}} |x_i|$, for every $d \in \mathbb{N}$ let $\mathbf{w}_d: \mathbb{R}^d \rightarrow (0, \infty)$ and $g_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ be measurable, for every $d \in \mathbb{N}$, $a \in A$ let $t_d^{(a)}: \mathbb{R}^{2d} \rightarrow \mathbb{R}^d$ be measurable, for every $d \in \mathbb{N}$ let $\xi_d: \Omega \rightarrow \mathbb{R}^d$ be measurable, let $\alpha \in A$, let $\phi: \mathbb{R} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}$ that $\phi(x) = \nu \ln(1 + \exp(x)) + (1 - \nu) \max\{x, \beta x\}$, let $(\mathbf{G}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \subseteq \mathbf{N}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in \mathbb{N}$ that $\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, let $(\mathbf{t}_{d,\varepsilon}^{(a)})_{(d,\varepsilon,a) \in \mathbb{N} \times (0,1] \times A} \subseteq \mathbf{N}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ that $\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(a)}) = \mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\alpha)})$ and $\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}) \in C(\mathbb{R}^{2d}, \mathbb{R}^d)$, assume that $\inf_{d \in \mathbb{N}, x \in \mathbb{R}^d} \mathbf{w}_d(x) \geq \nu$, assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\max\left\{\mathcal{L}(\mathbf{G}_{d,\varepsilon}), \mathcal{L}(\mathbf{t}_{d,\varepsilon}^{(\alpha)}), |\mathcal{D}(\mathbf{G}_{d,\varepsilon})|, |\mathcal{D}(\mathbf{t}_{d,\varepsilon}^{(\alpha)})|\right\} \leq \mathbf{c}d^c \varepsilon^{-c}, \quad (4.184)$$

assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x, y, p \in \mathbb{R}^d$, $a \in A$ that

$$\left| \left((\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}))(x) \right)(a) - (g_d(x))(a) \right| \leq \varepsilon \mathbf{c}d^c |\mathbf{w}_d(x)|, \quad (4.185)$$

$$\left\| (\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, p) - t_d^{(a)}(x, p) \right\| \leq \varepsilon \mathbf{c}d^c |\mathbf{w}_d(x)|, \quad \text{and}$$

$$\max\left\{K^{-1} |(g_d(x))(a) - (g_d(y))(a)|, \eta^{-1} \|t_d^{(a)}(x, p) - t_d^{(a)}(y, p)\|\right\} \leq \|x - y\|, \quad (4.186)$$

assume for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $a \in A$ that

$$\sup_{x \in \mathbb{R}^d} |\mathbf{w}_d(x)|^{-1} \mathbb{E}[\|t_d^{(a)}(x, \xi_d)\|] < \infty \quad \text{and} \quad \sup_{x \in \mathbb{R}^d} |\mathbf{w}_d(x)|^{-1} \mathbb{E}[\|(\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d)\|] < \infty, \quad (4.187)$$

assume for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $\varepsilon \in (0, 1]$, $a \in A$ that

$$\left(\mathbb{E} \left[\left| \mathbf{w}_d \left((\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d) \right) \right|^2 \right] \right)^{1/2} \leq \lambda |\mathbf{w}_d(x)|, \quad (4.188)$$

$$\left(\mathbb{E} \left[\max_{b \in A} \left| \left((\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}))((\mathcal{R}_\phi(\mathbf{t}_{d,\varepsilon}^{(a)}))(x, \xi_d)) \right)(b) \right|^2 \right] \right)^{1/2} \leq \mathbf{c}d^c |\mathbf{w}_d(x)|, \\ \mathbb{E}[\|\mathbf{w}_d(t_d^{(a)}(x, \xi_d))\|] \leq \lambda |\mathbf{w}_d(x)|, \quad \text{and} \quad \sup_{\substack{y \in \mathbb{R}^d \\ b \in A}} |\mathbf{w}_d(y)|^{-1} |(g_d(y))(b)| < \infty. \quad (4.189)$$

Then there exists $c \in \mathbb{R}$ such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^c$ it holds that

(i) there exists a unique measurable $\mathcal{Q}: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that $\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathbf{w}_d(z)|^{-1} |(\mathcal{Q}(z))(b)| < \infty$, $\mathbb{E}[\max_{b \in A} |(\mathcal{Q}(t_d^{(a)}(x, \xi_d)))(b)|] < \infty$, and

$$(\mathcal{Q}(x))(a) = (g_d(x))(a) + L \mathbb{E} \left[\max_{b \in A} (\mathcal{Q}(t_d^{(a)}(x, \xi_d)))(b) \right] \quad (4.190)$$

and

(ii) there exists $\mathbf{Q} \in \mathbf{N}$ which satisfies $\mathcal{H}(\mathbf{Q}) > 0$, $\mathcal{P}(\mathbf{Q}) \leq \mathbf{c}d^c \varepsilon^{-c}$, $\mathcal{R}_\phi(\mathbf{Q}) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, and

$$\left(\int_{\mathbb{R}^d} \|\mathcal{Q}(x) - (\mathcal{R}_\phi(\mathbf{Q}))(x)\|^2 \mu(dx) \right)^{1/2} \leq \varepsilon. \quad (4.191)$$

Proof of Corollary 4.3.4. Throughout this proof let $c = \max\{L(\mathbf{c} + 3 \ln(2) \lceil \log_2(|A|) \rceil), \mathbf{c}(2|A| + 1)\}$, let $\mathbf{I}, \mathfrak{J} \in \mathbf{N}$ satisfy

$$\mathbf{I} = \left(\left(\begin{pmatrix} 1 \\ -1 \end{pmatrix}, \begin{pmatrix} 0 \\ 0 \end{pmatrix} \right), \left((1 \ -1), 0 \right) \right) \in (\mathbb{R}^{2 \times 1} \times \mathbb{R}^2) \times (\mathbb{R}^{1 \times 2} \times \mathbb{R}^1) \quad \text{and} \quad (4.192)$$

$$\mathfrak{J} = \begin{cases} \mathbf{I} & : \nu = 1 \\ (1 + \beta)^{-1} \oplus \mathbf{I} & : \nu = 0, \end{cases} \quad (4.193)$$

let $\mathbb{I}_k \in \mathbf{N}$, $k \in \mathbb{N}$, satisfy for all $k \in \mathbb{N}$ that $\mathbb{I}_k = \mathbf{P}_k(\mathfrak{J}, \mathfrak{J}, \dots, \mathfrak{J})$, for every $d \in \mathbb{N}$ let $f_d: \mathbb{R}^d \times \mathbb{R}^{|A|} \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$, $r \in \mathbb{R}^{|A|}$ that $f_d(x, r) = L \max_{a \in A} ((g_d(x))(a) + r(a))$. First, observe that [1, Lemma 3.5] and [1, Lemma 3.8] ensure for all $x \in \mathbb{R}$ that $\mathcal{D}(\mathfrak{J}) = (1, 2, 1) \in \mathbb{N}^3$ and $(\mathcal{R}_\phi(\mathfrak{J}))(x) = x$. Next, note that Lemma 4.2.2 and Lemma 4.2.3 establish that there exist $\mathbf{M}_\delta \in \mathbf{N}$, $\delta \in (0, 1]$, such that for all $x = (x_1, \dots, x_{|A|})$, $y \in \mathbb{R}^{|A|}$, $\delta \in (0, 1]$ it holds that $\mathcal{R}_\phi(\mathbf{M}_\delta) \in C(\mathbb{R}^{|A|}, \mathbb{R})$, $\mathcal{L}(\mathbf{M}_\delta) = \lceil \log_2(|A|) \rceil + 1$, $\|\mathcal{D}(\mathbf{M}_\delta)\| \leq 2|A|$,

$$|(\mathcal{R}_\phi(\mathbf{M}_\delta))(x) - \max\{x_1, \dots, x_{|A|}\}| \leq 3 \ln(2) \lceil \log_2(|A|) \rceil \delta \nu, \quad (4.194)$$

and

$$|(\mathcal{R}_\phi(\mathbf{M}_\delta))(x) - (\mathcal{R}_\phi(\mathbf{M}_\delta))(y)| \leq \|x - y\|. \quad (4.195)$$

Throughout the remainder of this proof let $(\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \subseteq \mathbf{N}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that

$$\mathbf{F}_{d,\varepsilon} = L \otimes (\mathbf{M}_\varepsilon \bullet \mathfrak{S}_{|A|,2} \bullet \mathbf{P}_{2,(\mathbb{I}_{|A|}, \mathbb{I}_{|A|})}(\mathbf{G}_{d,\varepsilon}, \mathbb{I}_{|A|})). \quad (4.196)$$

Next, observe that the assumption that A is nonempty, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\max\{\mathcal{L}(\mathbf{G}_{d,\varepsilon}), \|\mathcal{D}(\mathbf{G}_{d,\varepsilon})\|\} \leq cd^\varepsilon \varepsilon^{-c}$, (4.194), (4.195), the fact that for all $k \in \mathbb{N}$ it holds that $\mathcal{D}(\mathbb{I}_k) = (k, 2k, k)$, the fact that for all $n \in \mathbb{N}$ it holds that $\lceil \log_2(n) \rceil + 1 \leq n$, Lemma 3.2.2, Lemma 3.2.8, Corollary 3.3.4, and Lemma 3.4.5 assure for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $r = (r(a))_{a \in A} \in \mathbb{R}^{|A|}$ that

$$\begin{aligned} \mathcal{I}(\mathbf{F}_{d,\varepsilon}) &= \mathcal{I}(\mathbf{P}_{2,(\mathbb{I}_{|A|}, \mathbb{I}_{|A|})}(\mathbf{G}_{d,\varepsilon}, \mathbb{I}_{|A|})) = d + |A|, \\ \mathcal{O}(\mathbf{F}_{d,\varepsilon}) &= \mathcal{O}(\mathbf{M}_\varepsilon) = 1, \quad \mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}) \in C(\mathbb{R}^{d+|A|}, \mathbb{R}), \\ \mathcal{L}(\mathbf{F}_{d,\varepsilon}) &= \mathcal{L}(\mathbf{M}_\varepsilon) + \mathcal{L}(\mathfrak{S}_{|A|,2}) + \mathcal{L}(\mathbf{P}_{2,(\mathbb{I}_{|A|}, \mathbb{I}_{|A|})}(\mathbf{G}_{d,\varepsilon}, \mathbb{I}_{|A|})) - 2 \\ &\leq \lceil \log_2(|A|) \rceil + 1 + \max\{\mathcal{L}(\mathbf{G}_{d,\varepsilon}), \mathcal{L}(\mathbb{I}_{|A|})\} - 1 \leq |A| + \mathcal{L}(\mathbf{G}_{d,\varepsilon}) + 1 \\ &\leq (|A| + 2)cd^\varepsilon \varepsilon^{-c} \leq (2|A| + 1)cd^\varepsilon \varepsilon^{-c} \leq cd^e \varepsilon^{-e}, \\ \|\mathcal{D}(\mathbf{F}_{d,\varepsilon})\| &\leq \max\{\|\mathcal{D}(\mathbf{M}_\varepsilon)\|, \|\mathcal{D}(\mathfrak{S}_{|A|,2})\|, \|\mathcal{D}(\mathbf{P}_{2,(\mathbb{I}_{|A|}, \mathbb{I}_{|A|})}(\mathbf{G}_{d,\varepsilon}, \mathbb{I}_{|A|}))\|\} \\ &\leq \max\{\|\mathcal{D}(\mathbf{M}_\varepsilon)\|, 2|A|, 2|A| + \|\mathcal{D}(\mathbf{G}_{d,\varepsilon})\|\} \\ &= 2|A| + \|\mathcal{D}(\mathbf{G}_{d,\varepsilon})\| \leq (2|A| + 1)cd^\varepsilon \varepsilon^{-c} \leq cd^e \varepsilon^{-e}, \\ (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) &= L \left((\mathcal{R}_\phi(\mathbf{M}_\varepsilon) \circ \mathcal{R}_\phi(\mathfrak{S}_{|A|,2}) \circ \mathcal{R}_\phi(\mathbf{P}_{2,(\mathbb{I}_{|A|}, \mathbb{I}_{|A|})}(\mathbf{G}_{d,\varepsilon}, \mathbb{I}_{|A|}))(x, r) \right) \\ &= L \left((\mathcal{R}_\phi(\mathbf{M}_\varepsilon) \circ \mathcal{R}_\phi(\mathfrak{S}_{d,\varepsilon})) \left(\underset{r}{(\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}))(x)} \right) \right) \\ &= L((\mathcal{R}_\phi(\mathbf{M}_\varepsilon))((\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}))(x) + r)), \end{aligned} \quad (4.197)$$

and

$$|(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - (\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|. \quad (4.199)$$

Next, observe that for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $r, s \in \mathbb{R}^{|A|}$, $\varepsilon \in (0, 1]$ it holds that $|f_d(x, r) - f_d(x, s)| \leq L \max_{a \in A} |r(a) - s(a)|$. Moreover, note that the assumption that for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $a \in A$ it holds that $|(g_d(x))(a) - (g_d(y))(a)| \leq K\|x - y\|$ implies for all $d \in \mathbb{N}$, $x, y \in \mathbb{R}^d$, $r \in \mathbb{R}^{|A|}$ that

$$|f_d(x, r) - f_d(y, r)| \leq L \max_{a \in A} |(g_d(x))(a) - (g_d(y))(a)| \leq KL\|x - y\|. \quad (4.200)$$

In addition, observe that the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $|((\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}))(x))(a) - (g_d(x))(a)| \leq \varepsilon cd^\varepsilon \|\mathbf{w}_d(x)\|$, (4.194), (4.195), the fact that $|A| \geq 1$,

the assumption that $\inf_{d \in \mathbb{N}, x \in \mathbb{R}^d} \mathfrak{w}_d(x) \geq \nu$, and the triangle inequality establish for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $r \in \mathbb{R}^{|A|}$ that

$$\begin{aligned}
& |(\mathcal{R}_\phi(\mathbf{F}_{d,\varepsilon}))(x, r) - f_d(x, r)| \\
& \leq L \left(|(\mathcal{R}_\phi(\mathbf{M}_\varepsilon))((\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}))(x) + r) - (\mathcal{R}_\phi(\mathbf{M}_\varepsilon))(g_d(x) + r)| \right. \\
& \quad \left. + |(\mathcal{R}_\phi(\mathbf{M}_\varepsilon))(g_d(x) + r) - \max_{a \in A}((g_d(x))(a) + r(a))| \right) \\
& \leq L(\|\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon})(x) - g_d(x)\| + 3 \ln(2) \lceil \log_2(|A|) \rceil \varepsilon \nu) \\
& \leq \varepsilon L(\mathfrak{c} d^c |\mathfrak{w}_d(x)| + 3 \ln(2) \lceil \log_2(|A|) \rceil |\mathfrak{w}_d(x)|) \\
& \leq \varepsilon L d^c |\mathfrak{w}_d(x)| (\mathfrak{c} + 3 \ln(2) \lceil \log_2(|A|) \rceil) \leq \varepsilon c d^c |\mathfrak{w}_d(x)|.
\end{aligned} \tag{4.201}$$

This and Corollary 4.3.3 (applied with

$$\begin{aligned}
& A \curvearrowright A, \quad (\Omega, \mathcal{F}, \mathbb{P}) \curvearrowright (\Omega, \mathcal{F}, \mathbb{P}), \quad \mathfrak{c} \curvearrowright c, \quad \nu \curvearrowright \nu, \quad \beta \curvearrowright \beta, \\
& (\lambda_d, \eta_d, K_d, L_d)_{d \in \mathbb{N}} \curvearrowright (\lambda, \eta, KL, L)_{d \in \mathbb{N}}, \quad (\ell_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \curvearrowright (L)_{(d,\varepsilon) \in \mathbb{N} \times (0,1]}, \\
& (\mathfrak{w}_d)_{d \in \mathbb{N}} \curvearrowright (\mathfrak{w}_d)_{d \in \mathbb{N}}, \quad (f_d)_{d \in \mathbb{N}} \curvearrowright (f_d)_{d \in \mathbb{N}}, \\
& (t_d^{(a)})_{(d,a) \in \mathbb{N} \times A} \curvearrowright (t_d^{(a)})_{(d,a) \in \mathbb{N} \times A}, \quad (\xi_d)_{d \in \mathbb{N}} \curvearrowright (\xi_d)_{d \in \mathbb{N}}, \quad a \curvearrowright a, \quad \phi \curvearrowright \phi, \\
& (\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \curvearrowright (\mathbf{F}_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]}, \quad (\mathfrak{t}_{d,\varepsilon}^{(a)})_{(d,\varepsilon,a) \in \mathbb{N} \times (0,1] \times A} \curvearrowright (\mathfrak{t}_{d,\varepsilon}^{(a)})_{(d,\varepsilon,a) \in \mathbb{N} \times (0,1] \times A}
\end{aligned} \tag{4.202}$$

in the notation of Corollary 4.3.3) demonstrate that there exists $c \in \mathbb{R}$ such that for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq c d^c$ it holds that

- (I) there exists a unique measurable $u_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$ which satisfies for all $x \in \mathbb{R}^d$, $a \in A$ that $\sup_{(z,b) \in \mathbb{R}^d \times A} |\mathfrak{w}_d(z)|^{-1} |(u_d(z))(b)| < \infty$,

$$\mathbb{E} \left[\max_{b \in A} |(g_d(t_d^{(a)}(x, \xi_d)))(b) + (u_d(t_d^{(a)}(x, \xi_d)))(b)| \right] < \infty, \tag{4.203}$$

and

$$(u_d(x))(a) = L \mathbb{E} \left[\max_{b \in A} \left((g_d(t_d^{(a)}(x, \xi_d)))(b) + (u_d(t_d^{(a)}(x, \xi_d)))(b) \right) \right] \tag{4.204}$$

and

- (II) there exists $\Psi_{d,\varepsilon,\mu} \in \mathbf{N}$ which satisfies $\mathcal{H}(\Psi_{d,\varepsilon,\mu}) > 0$, $\mathcal{P}(\Psi_{d,\varepsilon,\mu}) \leq c d^c \varepsilon^{-c}$, $\mathcal{R}_\phi(\Psi_{d,\varepsilon,\mu}) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, and

$$\left(\int_{\mathbb{R}^d} \|u_d(x) - (\mathcal{R}_\phi(\Psi_{d,\varepsilon,\mu}))(x)\|^2 \mu(dx) \right)^{1/2} \leq \varepsilon. \tag{4.205}$$

Note that the fact that $\mathfrak{c} \leq c$ implies for every $d \in \mathbb{N}$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq c d^c$ it holds that $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq c d^c$. Throughout the remainder of this proof let $\varphi_d: \mathbb{R}^d \rightarrow \mathbb{R}^{|A|}$, $d \in \mathbb{N}$, satisfy for all $d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ that $(\varphi_d(x))(a) = (g_d(x))(a) + (u_d(x))(a)$, let $(\delta_{d,\varepsilon})_{(d,\varepsilon) \in \mathbb{N} \times (0,1]} \subseteq \mathbb{R}$ satisfy for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $\delta_{d,\varepsilon} = \varepsilon(1 + |A|^{1/2} \mathfrak{c}^2 d^{2c})^{-1}$, let $\mathfrak{K} = \max\{2|A|, \mathfrak{c}, c\}$, and for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathfrak{w}_d(x)|^2 \mu(dx))^{1/2} \leq c d^c$ let $\mathbf{Q}_{d,\varepsilon,\mu} \in \mathbf{N}$ satisfy that $\mathbf{Q}_{d,\varepsilon,\mu} = \mathbf{G}_{d,\delta_{d,\varepsilon}} \boxplus_{\mathbb{I}_{|A|}} \Psi_{d,\delta_{d,\varepsilon},\mu}$. Observe that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\delta_{d,\varepsilon} \in (0, 1]$. Moreover, note that the assumption that for all $d \in \mathbb{N}$ it holds that g_d is measurable, the assumption that for all $d \in \mathbb{N}$ it holds that $\sup_{(x,a) \in \mathbb{R}^d \times A} |\mathfrak{w}_d(x)|^{-1} |(g_d(x))(a)| < \infty$, and item (I) ensure that for all

$d \in \mathbb{N}$, $x \in \mathbb{R}^d$, $a \in A$ it holds that φ_d is measurable, $\sup_{(y,b) \in \mathbb{R}^d \times A} |\mathbf{w}_d(y)|^{-1} |(\varphi_d(y))(b)| < \infty$, $\mathbb{E} \left[\max_{b \in A} |(\varphi_d(t_d^{(a)}(x, \xi_d)))(b)| \right] < \infty$, and

$$\begin{aligned} (\varphi_d(x))(a) &= (g_d(x))(a) + (u_d(x))(x) \\ &= (g_d(x))(a) + L \mathbb{E} \left[\max_{b \in A} ((g_d(t_d^{(a)}(x, \xi_d)))(b) + (u_d(t_d^{(a)}(x, \xi_d)))(b)) \right] \\ &= (g_d(x))(a) + L \mathbb{E} \left[\max_{b \in A} (\varphi_d(t_d^{(a)}(x, \xi_d)))(b) \right]. \end{aligned} \quad (4.206)$$

This proves item (i). Furthermore, note that Lemma 3.4.7 establishes for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^c$ that $\mathcal{I}(\mathbf{Q}_{d,\varepsilon,\mu}) = \mathcal{I}(\mathbf{G}_{d,\delta_{d,\varepsilon}}) = d$, $\mathcal{O}(\mathbf{Q}_{d,\varepsilon,\mu}) = \mathcal{O}(\mathbf{G}_{d,\delta_{d,\varepsilon}}) = |A|$, $\mathcal{R}_\phi(\mathbf{Q}_{d,\varepsilon,\mu}) \in C(\mathbb{R}^d, \mathbb{R}^{|A|})$, and $(\mathcal{R}_\phi(\mathbf{Q}_{d,\varepsilon,\mu}))(x) = (\mathcal{R}_\phi(\mathbf{G}_{d,\delta_{d,\varepsilon}}))(x) + (\mathcal{R}_\phi(\Psi_{d,\delta_{d,\varepsilon},\mu}))(x) \in \mathbb{R}^{|A|}$. This, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in \mathbb{R}^d$, $a \in A$ it holds that $|(g_d(x))(a) - ((\mathcal{R}_\phi(\mathbf{G}_{d,\varepsilon}))(x))(a)| \leq \varepsilon \mathbf{c}d^c |\mathbf{w}_d(x)|$, the fact that for all $k \in \mathbb{N}$, $x = (x_1, \dots, x_k) \in \mathbb{R}^k$ it holds that $\|x\| \leq k^{1/2} \max_{i \in \{1, \dots, k\}} |x_i|$, the triangle inequality, and item (II) yield for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^c$ that

$$\begin{aligned} & \left(\int_{\mathbb{R}^d} \|\varphi_d(x) - (\mathcal{R}_\phi(\mathbf{Q}_{d,\varepsilon,\mu}))(x)\|^2 \mu(dx) \right)^{1/2} \\ &= \left(\int_{\mathbb{R}^d} \|g_d(x) - (\mathcal{R}_\phi(\mathbf{G}_{d,\delta_{d,\varepsilon}}))(x) + u_d(x) - (\mathcal{R}_\phi(\Psi_{d,\delta_{d,\varepsilon},\mu}))(x)\|^2 \mu(dx) \right)^{1/2} \\ &\leq \left(\int_{\mathbb{R}^d} \|g_d(x) - (\mathcal{R}_\phi(\mathbf{G}_{d,\delta_{d,\varepsilon}}))(x)\|^2 \mu(dx) \right)^{1/2} \\ &\quad + \left(\int_{\mathbb{R}^d} \|u_d(x) - (\mathcal{R}_\phi(\Psi_{d,\delta_{d,\varepsilon},\mu}))(x)\|^2 \mu(dx) \right)^{1/2} \\ &\leq \delta_{d,\varepsilon} |A|^{1/2} \mathbf{c}d^c \left(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx) \right)^{1/2} + \delta_{d,\varepsilon} \leq \delta_{d,\varepsilon} (1 + |A|^{1/2} \mathbf{c}^2 d^{2c}) \leq \varepsilon. \end{aligned} \quad (4.207)$$

This proves (4.191). Moreover, observe that the assumption that $\mathfrak{K} = \max\{2|A|, \mathbf{c}, c\}$, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\max\{\mathcal{L}(\mathbf{G}_{d,\varepsilon}), |\mathcal{D}(\mathbf{G}_{d,\varepsilon})|\} \leq \mathbf{c}d^c \varepsilon^{-c}$, and item (II) establish for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^c$ that

$$\mathcal{L}(\mathbf{Q}_{d,\varepsilon,\mu}) \leq \max\{\mathcal{L}(\mathbf{G}_{d,\delta_{d,\varepsilon}}), \mathcal{L}(\Psi_{d,\delta_{d,\varepsilon},\mu})\} \leq \max\{\mathbf{c}d^c \delta_{d,\varepsilon}^{-c}, \mathbf{c}d^c \delta_{d,\varepsilon}^{-c}\} \leq \mathfrak{K} d^{\mathfrak{K}} \delta_{d,\varepsilon}^{-\mathfrak{K}}. \quad (4.208)$$

and

$$\begin{aligned} |\mathcal{D}(\mathbf{Q}_{d,\varepsilon,\mu})| &\leq |\mathcal{D}(\mathcal{E}_{\mathcal{L}(\mathbf{Q}_{d,\varepsilon,\mu}), \mathbb{I}_{|A|}}(\mathbf{G}_{d,\delta_{d,\varepsilon}}))| + |\mathcal{D}(\mathcal{E}_{\mathcal{L}(\mathbf{Q}_{d,\varepsilon,\mu}), \mathbb{I}_{|A|}}(\Psi_{d,\delta_{d,\varepsilon},\mu}))| \\ &\leq \max\{|\mathcal{D}(\mathbb{I}_{|A|})|, |\mathcal{D}(\mathbf{G}_{d,\delta_{d,\varepsilon}})|\} + \max\{|\mathcal{D}(\mathbb{I}_{|A|})|, |\mathcal{D}(\Psi_{d,\delta_{d,\varepsilon},\mu})|\} \\ &\leq \max\{2|A|, \mathbf{c}d^c \delta_{d,\varepsilon}^{-c}\} + \max\{2|A|, \mathbf{c}d^c \delta_{d,\varepsilon}^{-c}\} \leq 2\mathfrak{K} d^{\mathfrak{K}} \delta_{d,\varepsilon}^{-\mathfrak{K}}. \end{aligned} \quad (4.209)$$

This, the assumption that for all $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ it holds that $\delta_{d,\varepsilon} = \varepsilon(1 + |A|^{1/2} \mathbf{c}^2 d^{2c})^{-1}$, and Lemma 3.1.3 demonstrate for every $d \in \mathbb{N}$, $\varepsilon \in (0, 1]$ and every probability measure $\mu: \mathcal{B}(\mathbb{R}^d) \rightarrow [0, 1]$ with $(\int_{\mathbb{R}^d} |\mathbf{w}_d(x)|^2 \mu(dx))^{1/2} \leq \mathbf{c}d^c$ that

$$\begin{aligned} \mathcal{P}(\mathbf{Q}_{d,\varepsilon,\mu}) &\leq 2\mathcal{L}(\mathbf{Q}_{d,\varepsilon,\mu}) |\mathcal{D}(\mathbf{Q}_{d,\varepsilon,\mu})|^2 \leq 2\mathfrak{K} d^{\mathfrak{K}} \delta_{d,\varepsilon}^{-\mathfrak{K}} (2\mathfrak{K} d^{\mathfrak{K}} \delta_{d,\varepsilon}^{-\mathfrak{K}})^2 = 8\mathfrak{K}^3 d^{3\mathfrak{K}} (1 + |A|^{1/2} \mathbf{c}^2 d^{2c})^{3\mathfrak{K}} \varepsilon^{-3\mathfrak{K}} \\ &\leq 8\mathfrak{K}^3 d^{3\mathfrak{K}} 2^{3\mathfrak{K}} |A|^{\frac{3}{2}\mathfrak{K}} \mathbf{c}^{6\mathfrak{K}} d^{6\mathfrak{K}} \varepsilon^{-3\mathfrak{K}} = 8^{\mathfrak{K}+1} \mathfrak{K}^3 |A|^{\frac{3}{2}\mathfrak{K}} \mathbf{c}^{6\mathfrak{K}} d^{3\mathfrak{K}(1+2c)} \varepsilon^{-3\mathfrak{K}}. \end{aligned} \quad (4.210)$$

This proves item (ii). The proof of Corollary 4.3.4 is thus complete. $\square$

Part II.

Monte Carlo and Deep Neural Network Approximations of Elliptic PDEs with Drift and Killing

In this part, we apply the common approximation strategy outlined in the general introduction in Chapter 1 to elliptic PDEs with drift and killing on bounded domains.

While the Monte Carlo methods and the corresponding error analysis for approximating expectations of the form (1.5) are presented in Chapter 5, the combination of these results with the ANN calculus of Chapter 3 is conducted in Chapter 6.

Both Chapter 5 and Chapter 6 are based on the joint work [64] with T. Kruse.

Chapter 5.

Walk-on-Spheres Monte Carlo Methods for Elliptic PDEs with Drift and Killing

This chapter is based on Sections 2 to 4 of the joint work [64] with T. Kruse. While Part II considers the numerical approximation of expectations of the form (1.5) by means of both Monte Carlo and ANN approximation, this chapter addresses the Monte Carlo approximation stage.

More precisely, this chapter considers the approximation of functions $w: D \rightarrow \mathbb{R}$,

$$w(x) = \mathbb{E}[h_0(\tau^x, x + B_{\tau^x})] + \mathbb{E}\left[\int_0^{\tau^x} h_1(s, x + B_s) ds\right], \quad x \in D, \quad (5.1)$$

where $d \in \mathbb{N} = \{1, 2, \dots\}$, $D \subseteq \mathbb{R}^d$ is open, bounded, and convex, $h_0, h_1: [0, \infty) \times \bar{D} \rightarrow \mathbb{R}$ are bounded and Lipschitz continuous, $(\Omega, \mathcal{F}, \mathbb{P})$ is a probability space on which all appearing random variables are defined, $B = (B_t)_{t \in [0, \infty)}: \Omega \times [0, \infty) \rightarrow \mathbb{R}^d$ is a d -dimensional standard Brownian motion starting in 0, and τ^x , $x \in D$, is the first exit time of the process $x + B$ from D , i.e., $\tau^x = \inf\{t \in [0, \infty): x + B_t \in \partial D\}$, $x \in D$. The approximation of expectations of the form (5.1) is particularly relevant in the numerical solution of PDEs. Indeed, such representations appear naturally in stochastic representations of solutions of linear elliptic drift-diffusion PDEs on bounded domains of the following type (cf. (1.4)):

$$\begin{aligned} \frac{1}{2} \operatorname{tr}(\sigma \sigma^\top D^2 u) + \langle b, \nabla u \rangle - ru + f &= 0 & \text{on } D, \\ u|_{\partial D} &= g & \text{on } \partial D, \end{aligned} \quad (5.2)$$

where $u: \bar{D} \rightarrow \mathbb{R}$ is the unknown, $\sigma \in \mathbb{R}^{d \times d}$, $b \in \mathbb{R}^d$, $r \in [0, \infty)$, $f: D \rightarrow \mathbb{R}$, and $g: \partial D \rightarrow \mathbb{R}$ (see Section 5.1 for a precise derivation of the relation between (5.1) and (5.2)). The goal of this chapter is to develop Monte Carlo schemes with error and complexity analyses for expectations of the form (5.1), with a particular focus on high-dimensional situations.

Probabilistic approximation methods for elliptic boundary problems include in particular the so-called Walk-on-Spheres algorithm, originally proposed in [74] to approximate the solution of the Dirichlet problem (i.e., the special case of (5.2) with $\sigma = \sqrt{2} \operatorname{Id}_{\mathbb{R}^d}$, $b = 0$, $r = 0$, $f = 0$) given by

$$\begin{aligned} \Delta u &= 0 & \text{on } D, \\ u|_{\partial D} &= g & \text{on } \partial D, \end{aligned} \quad (5.3)$$

via Monte Carlo simulation. In particular, it allows to efficiently approximate the exit point of Brownian motion from a bounded domain. Conceptually the Walk-on-Spheres algorithm works as follows: At the current position $x \in D$ determine the largest sphere around x contained in D . To obtain the next position, sample uniformly from this sphere. The key observation here is that this new random position is indeed identically distributed to the location of first exit of a Brownian motion started in x from said sphere. To obtain an approximation for the exit point of Brownian motion from the entire domain, this procedure is repeated until the process is within a prescribed distance to the boundary of the domain. In particular, the number

of steps required to reach that location is in general random. In this original algorithm the spheres from which the next point of the process is sampled exhibit maximal radii, i.e., the radius of the sphere to be sampled from is given by the distance of the current position to the boundary of the domain. Subsequent developments based on this algorithm cover methods for more general elliptic boundary value problems (see, e.g., [21, 67, 94]).

Recently, there is also a growing interest in Monte Carlo approximations of elliptic PDEs that include a drift component. Indeed, in the work [91] the Walk-on-Spheres algorithm is extended to elliptic drift-diffusion equations by utilizing a von Mises-Fisher exit distribution on the successive spheres. The article [92] introduces Monte Carlo methods for self-adjoint elliptic PDEs with constant coefficients that estimate the solutions and its derivatives at multiple points from one ensemble of trajectories. In [93], the authors extend this global approach to elliptic drift-diffusion equations.

To obtain global approximations for the drift-free Poisson problem (i.e., the special case of (5.2) with $\sigma = \text{Id}_{\mathbb{R}^d}$, $b = 0$, $r = 0$)

$$\begin{aligned} \frac{1}{2}\Delta u + f &= 0 & \text{on } D, \\ u|_{\partial D} &= g & \text{on } \partial D, \end{aligned} \tag{5.4}$$

the works [39] and [20] combine Walk-on-Spheres based Monte Carlo estimators with suitable neural network representations and approximations. In particular, under suitable assumptions these works construct ANNs that approximate the solution of (5.4) with a complexity that only grows polynomially in both the inverse of the approximation accuracy¹ and the problem dimension, thereby mitigating the curse of dimensionality.

Two features of the framework developed in [20] are particularly important for the developments of this chapter. First, it employs a modified version of the Walk-on-Spheres process utilizing a specific class of modified distance functions as a substitute for the true distance function in order to determine the radii of the successive spheres from which the successive positions are sampled. Second, the framework admits a deterministic stopping criterion, independent of the starting point of the Walk-on-Spheres process, in place of the random number of steps used in the classical Walk-on-Spheres algorithm. Both the deterministic stopping rule and the modified distance functions play an important role in the construction of the ANN approximations, which will be developed for our setting in Chapter 6. Furthermore, the authors of [20] work under an exterior ball condition on D rather than assuming convexity of D .

Building on this framework, we analyze in this chapter Monte Carlo estimators for expectations of the form (5.1) and consequently of solutions of elliptic drift-diffusion equations of the form (5.2). More precisely, we consider Monte Carlo estimators $\bar{w}: D \times \Omega \rightarrow \mathbb{R}$ satisfying for $x \in D$ that

$$\bar{w}(x) = \frac{1}{n} \sum_{i=1}^n \left[h_0(\bar{\tau}_M^{x,i}, \bar{X}_M^{x,i}) + \frac{1}{d} \sum_{k=1}^M \bar{r}(\bar{X}_{k-1}^{x,i})^2 h_1\left(\bar{\tau}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i})^2 V_k^i, \bar{X}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i}) Y_k^i\right) \right]. \tag{5.5}$$

Here, the number $n \in \mathbb{N}$ denotes the number of sample paths of the Walk-on-Spheres process, the number $M \in \mathbb{N}$ denotes the deterministic number of steps the employed Walk-on-Spheres process undertakes, the function $\bar{r}: \bar{D} \rightarrow \mathbb{R}$ determines the radii of the successive spheres in the generation of the Walk-on-Spheres process and is to be understood as a suitable substitute for the true distance to the boundary of D . For the sample with index $i \in \{1, \dots, n\}$ the random variables $\bar{X}_0^{x,i}, \dots, \bar{X}_M^{x,i}$ denote the M steps of the Walk-on-Spheres process with start in $\bar{X}_0^{x,i} = x \in D$. The random variable $\bar{\tau}_k^{x,i}$ corresponds to the physical time a Brownian motion

¹The approximation error is measured in $L^2(D)$ in [39] and in $L^\infty(D)$ in [20].

started in the position $\bar{X}_{k-1}^{x,i}$ needs to hit the boundary of the sphere with radius $\bar{r}(\bar{X}_{k-1}^{x,i})$ around $\bar{X}_{k-1}^{x,i}$. The random variable (V_k^i, Y_k^i) corresponds to the time and space occupation of a Brownian motion with start in 0 inside the unit sphere.

The estimator in (5.5) extends the estimator introduced in [20] by incorporating the sampled random times $\bar{\tau}_M^{x,i}$, $x \in D$, $M \in \mathbb{N}$, $i \in \{1, \dots, n\}$. This additional time dependency appears in stochastic representations of solutions of (5.2) as a compensation for the drift b and killing r components (see Proposition 5.1.2 and Remark 5.1.3 and below).

To characterize appearing complexities (e.g., the sample complexity in this chapter, the number of parameters describing an ANN later in Chapter 6), we adopt the following terminology: A real number $c \in \mathbb{R}$ is called an absolute constant, if c does not depend on any variables or parameters in the problem. Moreover, for quantities $Q, q_1, \dots, q_n \in \mathbb{R}$, where $n \in \mathbb{N}$, we say that Q depends only on $q_1, \dots, q_n$ if Q is an explicit function of the variables $q_1, \dots, q_n$ and Q has no other dependencies. In this case, this dependency is at most polynomial if there exists an absolute constant $c \in [0, \infty)$ such that $Q \leq c(1 + |q_1| + \dots + |q_n|)^c$. The main contribution of this chapter establishes an error analysis for Monte Carlo approximations (5.5) for expectations of the form (5.1) in the L^∞ -sense (see Theorem 5.3.9 and Corollary 5.3.10 below). In particular, we demonstrate that for a prescribed approximation accuracy $\varepsilon \in (0, 1)$ the estimator (5.5) achieves $\mathbb{E}[\sup_{x \in D} |w(x) - \bar{w}(x)|] \leq \varepsilon$ under suitable assumptions on the modified distance function $\bar{r}$ as well as with suitable choices of $n, M \in \mathbb{N}$. More specifically, n and M can be chosen such that $\max\{n, M\} \leq c\varepsilon^{-q}$, where $q \in [0, \infty)$ is an absolute constant and $c \in [0, \infty)$ is a constant that depends only on $\sup_{(t,x) \in [0,\infty) \times \bar{D}} |h_i(t,x)|$, $i \in \{0, 1\}$, the Lipschitz constants of h_i , $i \in \{0, 1\}$, the Lipschitz constant as well as a nondegeneracy constant² of the modified distance $\bar{r}$, the diameter $\text{diam}(D)$ of the domain, and the dimension d and this dependency is at most polynomial. In particular, imposing that these problem features grow at most polynomially with respect to the dimension d the Monte Carlo scheme indeed overcomes the curse of dimensionality.

The remainder of this chapter is structured as follows. In Section 5.1 below we establish the connection between functions of the form (5.1) and solutions of the PDE (5.2). In Section 5.2 we introduce the modified Walk-on-Spheres process and some of its important distributional properties and we present a Walk-on-Spheres-based representation of the expectation (5.1). In Section 5.3 we present the Monte Carlo estimator based on the Walk-on-Spheres process established in Section 5.2 and provide our error analysis.

5.1. A Girsanov-Based Feynman-Kac Representation

In this section, we establish the connection between expectations of the form (5.1) and classical solutions of the PDE (5.2) via suitable stochastic representations.

Unless otherwise specified, let $d \in \mathbb{N}$, $\sigma \in \mathbb{R}^{d \times d}$, $b \in \mathbb{R}^d$, $r \in \mathbb{R}$, $D \subseteq \mathbb{R}^d$, let $f: \bar{D} \rightarrow \mathbb{R}$, $g: \partial D \rightarrow \mathbb{R}$. In the following remark, we recall sufficient conditions to establish existence and uniqueness of a classical solution of the PDE (5.2) (see, e.g., [35, Chapter 6, Theorem 6.13], [34, Chapter 6, Theorem 2.4]).

Remark 5.1.1. Assume σ is invertible, assume $r \geq 0$, assume that $D \subseteq \mathbb{R}^d$ is open, connected, bounded, and satisfies an exterior ball condition at every boundary point, assume that f is uniformly α -Hölder continuous, for some $\alpha \in (0, 1)$, and assume that g is continuous on ∂D . Then there exists a unique $u \in C^0(\bar{D}) \cap C^2(D)$ that satisfies (5.2).

The next Proposition 5.1.2 gives rise to a stochastic representation of the solution of (5.2) that is the base for the Monte Carlo estimates employed in subsequent section. We fix some more notation. Unless otherwise specified, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $B: [0, \infty) \times$

²For details, see Definition 5.2.3 and Corollary 5.3.10 below

$\Omega \rightarrow \mathbb{R}^d$ be a standard Brownian motion, and let $(\mathcal{F}_t)_{t \in [0, \infty)}$ be the completed filtration generated by B . For every $x \in \overline{D}$ let

$$\tau^x = \inf\{t \in [0, \infty) : x + \sigma B_t \in \partial D\}. \quad (5.6)$$

Proposition 5.1.2. *Assume that D is open, bounded, and connected, that f and g are continuous, that σ is invertible, and that $r \geq -\frac{1}{2}\|\sigma^{-1}b\|^2$. Let $u \in C^0(\overline{D}) \cap C^2(D)$ be a solution of (5.2). Then it holds for all $x \in D$ that*

$$u(x) = \mathbb{E} \left[e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)\tau^x + \langle \sigma^{-1}b, B_{\tau^x} \rangle} g(x + \sigma B_{\tau^x}) + \int_0^{\tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, B_s \rangle} f(x + \sigma B_s) ds \right]. \quad (5.7)$$

Proof. Let $x \in D$. Note that Itô's formula implies for all $t \in [0, \infty)$ that

$$de^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)(t \wedge \tau^x) + \langle \sigma^{-1}b, B_{t \wedge \tau^x} \rangle} = \mathbf{1}_{\{t < \tau^x\}} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)t + \langle \sigma^{-1}b, B_t \rangle} (-r dt + \langle \sigma^{-1}b, dB_t \rangle) \quad (5.8)$$

and

$$du(x + \sigma B_{t \wedge \tau^x}) = \mathbf{1}_{\{t < \tau^x\}} (\langle \nabla u(x + \sigma B_t), \sigma dB_t \rangle + \frac{1}{2} \text{tr}(\sigma \sigma^\top D^2 u(x + \sigma B_t)) dt). \quad (5.9)$$

Combining this, the fact that for all $t \in [0, \infty)$ it holds that $B_{t \wedge \tau^x} \mathbf{1}_{\{t < \tau^x\}} = B_t \mathbf{1}_{\{t < \tau^x\}}$, and the fact that u solves (5.2) with the product rule entails for all $t \in [0, \infty)$ that

$$\begin{aligned} & d \left(e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)(t \wedge \tau^x) + \langle \sigma^{-1}b, B_{t \wedge \tau^x} \rangle} u(x + \sigma B_{t \wedge \tau^x}) \right) \\ &= \mathbf{1}_{\{t < \tau^x\}} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)t + \langle \sigma^{-1}b, B_t \rangle} \left(\langle \nabla u(x + \sigma B_t), \sigma dB_t \rangle - f(x + \sigma B_t) dt \right. \\ &\quad \left. - \langle b, \nabla u(x + \sigma B_t) \rangle dt + ru(x + \sigma B_t) dt - ru(x + \sigma B_t) dt + \langle b, \nabla u(x + \sigma B_t) \rangle dt \right. \\ &\quad \left. + u(x + \sigma B_t) \langle \sigma^{-1}b, dB_t \rangle \right) \\ &= \mathbf{1}_{\{t < \tau^x\}} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)t + \langle \sigma^{-1}b, B_t \rangle} \left(\langle \sigma^\top \nabla u(x + \sigma B_t) + \sigma^{-1}bu(x + \sigma B_t), dB_t \rangle \right. \\ &\quad \left. - f(x + \sigma B_t) dt \right). \end{aligned} \quad (5.10)$$

Written as integrals it holds for all $t \in [0, \infty)$ that

$$\begin{aligned} & e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)(t \wedge \tau^x) + \langle \sigma^{-1}b, B_{t \wedge \tau^x} \rangle} u(x + \sigma B_{t \wedge \tau^x}) - u(x) \\ &= - \int_0^{t \wedge \tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, B_s \rangle} f(x + \sigma B_s) ds \\ &\quad + \int_0^{t \wedge \tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, B_s \rangle} \langle u(x + \sigma B_s) \sigma^{-1}b + \sigma^\top \nabla u(x + \sigma B_s), dB_s \rangle. \end{aligned} \quad (5.11)$$

Next, let τ_k , $k \in \mathbb{N}$, be a localizing sequence of stopping times for the local martingale $\int_0^{\cdot \wedge \tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)t + \langle \sigma^{-1}b, B_t \rangle} \langle u(x + \sigma B_t) \sigma^{-1}b + \sigma^\top \nabla u(x + \sigma B_t), dB_t \rangle$. This ensures for all $t \in [0, \infty)$, $k \in \mathbb{N}$ that

$$\begin{aligned} u(x) &= \mathbb{E} \left[e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)(t \wedge \tau_k \wedge \tau^x) + \langle \sigma^{-1}b, B_{t \wedge \tau_k \wedge \tau^x} \rangle} u(x + \sigma B_{t \wedge \tau_k \wedge \tau^x}) \right. \\ &\quad \left. + \int_0^{t \wedge \tau_k \wedge \tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, B_s \rangle} f(x + \sigma B_s) ds \right]. \end{aligned} \quad (5.12)$$

This, the dominated convergence theorem, and the fact that $u|_{\partial D} = g$ on ∂D demonstrate that

$$\begin{aligned}
u(x) &= \lim_{\substack{t \rightarrow \infty \\ k \rightarrow \infty}} \mathbb{E} \left[e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)(t \wedge \tau_k \wedge \tau^x) + \langle \sigma^{-1}b, B_{t \wedge \tau_k \wedge \tau^x} \rangle} u(x + \sigma B_{t \wedge \tau_k \wedge \tau^x}) \right. \\
&\quad \left. + \int_0^{t \wedge \tau_k \wedge \tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, B_s \rangle} f(x + \sigma B_s) ds \right] \\
&= \mathbb{E} \left[e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)\tau^x + \langle \sigma^{-1}b, B_{\tau^x} \rangle} u(x + \sigma B_{\tau^x}) \right. \\
&\quad \left. + \int_0^{\tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, B_s \rangle} f(x + \sigma B_s) ds \right] \\
&= \mathbb{E} \left[e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)\tau^x + \langle \sigma^{-1}b, B_{\tau^x} \rangle} g(x + \sigma B_{\tau^x}) \right. \\
&\quad \left. + \int_0^{\tau^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, B_s \rangle} f(x + \sigma B_s) ds \right].
\end{aligned} \tag{5.13}$$

This completes the proof. $\square$

In the following remark, we informally derive another stochastic representation of the solution u of (5.2) based on the Feynman-Kac representation and establish a connection to the representation (5.7) as well as reason why the representation (5.7) is preferable for our work.

Remark 5.1.3. Let $W: [0, \infty) \times \Omega \rightarrow \mathbb{R}^d$ be a standard Brownian motion on Ω under the probability measure Q . For every $x \in \mathbb{R}^d$ let $X^x: [0, \infty) \times \Omega \rightarrow \mathbb{R}^d$ and $\tilde{\tau}^x: \Omega \rightarrow [0, \infty)$ satisfy $X_t^x = x + bt + \sigma W_t$, $t \in [0, \infty)$, and $\tilde{\tau}^x = \inf\{t \in [0, \infty) | X_t^x \in \partial D\}$. Then the Feynman-Kac formula yields under appropriate assumptions for every $x \in \mathbb{R}^d$ that

$$u(x) = E^Q \left[e^{-r\tilde{\tau}^x} g(X_{\tilde{\tau}^x}^x) + \int_0^{\tilde{\tau}^x} e^{-rs} f(X_s^x) ds \right]. \tag{5.14}$$

In Section 5.2 and following, we employ the Walk-on-Spheres algorithm which is based on the isotropy of Brownian motion. For this reason, representation (5.7), which employs exit times of $x + \sigma B$, is more convenient than (5.14) which is based on the non-isotropic process X^x .

Note that (5.7) can be derived from (5.14) via Girsanov's theorem (see, e.g., [95, Appendix A.1] for a similar derivation). To this end, let $K: [0, \infty) \times \Omega \rightarrow \mathbb{R}^d$, $L: [0, \infty) \times \Omega \rightarrow \mathbb{R}$, and $\tau_K^x: \Omega \rightarrow [0, \infty]$, $x \in \mathbb{R}^d$, satisfy

$$\begin{aligned}
K_t &= W_t + \sigma^{-1}bt, \quad t \in [0, \infty), \\
L_t &= e^{-\langle \sigma^{-1}b, W_t \rangle - \frac{1}{2}\|\sigma^{-1}b\|^2 t}, \quad t \in [0, \infty), \\
\tau_K^x &= \inf\{t \in [0, \infty) : x + \sigma K_t \in \partial D\}, \quad x \in \mathbb{R}^d.
\end{aligned} \tag{5.15}$$

Note that for all $t \in [0, \infty)$, $x \in \mathbb{R}^d$ we have $X_t^x = x + \sigma K_t$, $L_t = e^{-\langle \sigma^{-1}b, K_t \rangle + \frac{1}{2}\|\sigma^{-1}b\|^2 t}$, and $\tilde{\tau}^x = \tau_K^x$. Then, Girsanov's theorem (see, e.g., [61, Section 3.5, Theorem 5.1]) implies for every $T \in [0, \infty)$ that K is a Brownian motion on $(\Omega, \mathcal{F}_T, (\mathcal{F}_t)_{t \in [0, T]}, P_T)$ with $\frac{dP_T}{dQ} = L_T$. By a suitable limiting procedure we obtain a measure P_∞ such that K is a Brownian motion under P_∞ and it follows from (5.14) that for all $x \in \mathbb{R}^d$ it holds that

$$\begin{aligned}
u(x) &= E^{P_\infty} \left[e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)\tau_K^x + \langle \sigma^{-1}b, K_{\tau_K^x} \rangle} g(x + \sigma K_{\tau_K^x}) \right. \\
&\quad \left. + \int_0^{\tau_K^x} e^{-(r+\frac{1}{2}\|\sigma^{-1}b\|^2)s + \langle \sigma^{-1}b, K_s \rangle} f(x + \sigma K_s) ds \right].
\end{aligned} \tag{5.16}$$

This exactly corresponds to (5.7).

Remark 5.1.4. Under the assumption that σ is invertible one can transform the elliptic PDE (5.2) to one with isotropic Laplacian. Indeed, let $v \in C^2(\sigma^{-1}D)$ satisfy

$$\begin{aligned} \frac{1}{2}\Delta v + \langle \sigma^{-1}b, \nabla v \rangle - rv + f(\sigma \cdot) &= 0 \quad \text{on } \sigma^{-1}D \\ v|_{\partial(\sigma^{-1}D)} &= g(\sigma \cdot) \quad \text{on } \partial(\sigma^{-1}D). \end{aligned} \quad (5.17)$$

Then the function $u(x) = v(\sigma^{-1}x)$, $x \in D$, satisfies $u \in C^2(D)$ and (5.2). For this reason we focus attention on the case $\sigma = \text{Id}$ in the remainder of this article. A proof for this transformation in the case $b = 0$, $r = 0$ is given in [20, Lemma 2.1].

Remark 5.1.5. In the situation of Proposition 5.1.2 with $\sigma = \text{Id}$ let $h_0: [0, \infty) \times \partial D \rightarrow \mathbb{R}$, $h_1: [0, \infty) \times \overline{D} \rightarrow \mathbb{R}$ satisfy for all $t \in [0, \infty)$, $x \in \partial D$, $y \in \overline{D}$ that

$$h_0(t, x) = e^{-(r+\frac{1}{2}\|b\|^2)t + \langle b, x \rangle} g(x) \quad \text{and} \quad h_1(t, y) = e^{-(r+\frac{1}{2}\|b\|^2)t + \langle b, y \rangle} f(y). \quad (5.18)$$

Then it holds for all $x \in D$ that

$$u(x) = e^{-\langle b, x \rangle} \left(\mathbb{E}[h_0(\tau^x, x + B_{\tau^x})] + \mathbb{E} \left[\int_0^{\tau^x} h_1(s, x + B_s) ds \right] \right). \quad (5.19)$$

Under the additional assumption that g is extendable to $\overline{D}$ and that both g and f are Lipschitz on $\overline{D}$, it holds that h_0 as in (5.18) is well-defined on $[0, \infty) \times \overline{D}$ and h_0 and h_1 are bounded and Lipschitz continuous. Hence, the relation between the PDE solution u and the expectation w as in (5.1) is established as $u(x) = e^{-\langle b, x \rangle} w(x)$ for $x \in D$. We remark that this transformation is also used in [91, Section 2].

5.2. A Modified Walk-on-Spheres Algorithm

In this section, we introduce the modified Walk-on-Spheres algorithm as proposed in [20] and demonstrate some relevant distributional properties. In particular, we present in Lemma 5.2.7 estimates on the tail probability of the number of steps the Walk-on-Spheres process requires to reach the boundary of the domain within a certain distance. Moreover, we establish in Lemma 5.2.9 a representation of functions of the form (5.1) based on the established Walk-on-Spheres process which subsequently motivates the Monte Carlo estimator considered in Section 5.3.

The following lemma demonstrates a consequence of the rotational invariance of Brownian motion.

Lemma 5.2.1. *Let $\tau = \inf\{t \in [0, \infty): \|B_t\| \geq 1\}$. Then it holds that τ and B_τ are independent.*

Proof. Throughout this proof let $S = \{y \in \mathbb{R}^d: \|y\| = 1\}$ and let $\mu: \mathcal{B}(S) \rightarrow [0, 1]$ denote the uniform distribution on $(S, \mathcal{B}(S))$. First, observe that τ and B_τ are measurable path functionals of the path of B . Next, note that the fact that for all $R \in SO(d)$ it holds that B and RB are identically distributed and the fact that for all $R \in SO(d)$, $x \in \mathbb{R}^d$ it holds that $\|Rx\| = \|x\|$ yield for all $R \in SO(d)$ that (τ, B_τ) and (τ, RB_τ) are identically distributed. Moreover, observe that the fact that $[0, \infty]$ is a Polish space ensures that there exists a regular conditional distribution $\kappa: [0, \infty] \times \mathcal{B}(S) \rightarrow [0, 1]$ of B_τ given τ (cf., [65, Section 8.3]). This implies for all $A \in \mathcal{B}([0, \infty])$, $C \in \mathcal{B}(S)$ that

$$\begin{aligned} \mathbb{P}[\tau \in A, B_\tau \in C] &= \mathbb{E}[\mathbb{1}_{\{\tau \in A\}} \mathbb{1}_{\{B_\tau \in C\}}] = \mathbb{E}[\mathbb{1}_{\{\tau \in A\}} \mathbb{E}[\mathbb{1}_{\{B_\tau \in C\}} | \tau]] \\ &= \int_{\Omega} \mathbb{E}[\mathbb{1}_{\{B_\tau \in C\}} | \tau] \mathbb{1}_{\{\tau \in A\}} d\mathbb{P} = \int_A \kappa(t, C) (\mathbb{P} \circ \tau^{-1})(dt). \end{aligned} \quad (5.20)$$

Combining this with the fact that for all $R \in SO(d)$ it holds that (τ, B_τ) and (τ, RB_τ) are identically distributed demonstrates for all $A \in \mathcal{B}([0, \infty])$, $C \in \mathcal{B}(S)$, $R \in SO(d)$ that

$$\begin{aligned} \int_A \kappa(t, C)(\mathbb{P} \circ \tau^{-1})(dt) &= \mathbb{P}[\tau \in A, B_\tau \in C] = \mathbb{P}[\tau \in A, RB_\tau \in C] \\ &= \int_A \kappa(t, R^\top C)(\mathbb{P} \circ \tau^{-1})(dt). \end{aligned} \quad (5.21)$$

This implies that for $\mathbb{P} \circ \tau^{-1}$ -a.a. $t \in [0, \infty)$ it holds that $\kappa(t, \cdot)$ is a rotational invariant probability measure on S . This and the fact that μ is the unique rotationally invariant probability measure on S (cf., [70, Theorem 3.4]) imply that for $\mathbb{P} \circ \tau^{-1}$ -a.a. $t \in [0, \infty)$ it holds that $\kappa(t, \cdot) = \mu$. This and (5.20) yield for all $A \in \mathcal{B}([0, \infty])$, $C \in \mathcal{B}(S)$ that

$$\mathbb{P}[\tau \in A, B_\tau \in C] = \mathbb{P}[\tau \in A]\mu(C). \quad (5.22)$$

This shows that τ and B_τ are independent. The proof of Lemma 5.2.1 is thus complete. $\square$

Throughout the remainder of this section, assume that D is open and bounded. Let $r: \mathbb{R}^d \rightarrow [0, \infty)$ satisfy for all $x \in \mathbb{R}^d$ that

$$r(x) = \inf\{\|x - y\|: y \in \partial D\}. \quad (5.23)$$

The following lemma is a modification of [39, Lemma 3.2]. The geometric proof given there can be adapted to yield the following statement. Recall for all $x \in \overline{D}$ that $\tau^x = \inf\{t \in [0, \infty): x + B_t \in \partial D\}$.

Lemma 5.2.2. *Assume that D is convex. Then it holds for all $x \in D$ that*

$$\mathbb{E}[\tau^x] \leq r(x) \text{diam}(D). \quad (5.24)$$

The next definition is [20, Definition 2.7] and describes the class of modified distance functions, which subsequently determines the radii of the successive spheres on which the Walk-on-Spheres process is defined.

Definition 5.2.3. Let $\beta \in (0, 1]$, $\varepsilon \in [0, \infty)$, let $\tilde{r}: D \rightarrow [0, \infty)$ be Lipschitz continuous, and assume for all $x, y \in D$ with $r(x) \geq \varepsilon$ that $\tilde{r}(y) \leq r(y)$ and $\tilde{r}(x) \geq \beta r(x)$. Then $\tilde{r}$ is called (β, ε) -distance on D .

With this notion of modified distance functions, we now introduce the Walk-on-Spheres process. Let $\beta \in (0, 1]$, $\delta \in [0, \infty)$, let $\bar{r}$ be a (β, δ) -distance on D . Let $\tau_k^x: \Omega \rightarrow [0, \infty)$, $k \in \mathbb{N}_0$, satisfy for all $x \in \overline{D}$, $k \in \mathbb{N}$ that $\tau_0^x = 0$ and

$$\tau_k^x = \inf\{s \in [\tau_{k-1}^x, \infty): \|B_s - B_{\tau_{k-1}^x}\| \geq \bar{r}(x + B_{\tau_{k-1}^x})\}, \quad (5.25)$$

let $Z_k^j: \Omega \rightarrow \mathbb{R}^d$, $j, k \in \mathbb{N}$, be standard normally distributed, let $\xi_k^j: \Omega \rightarrow \mathbb{R}^d$, $j, k \in \mathbb{N}$, satisfy for all $j, k \in \mathbb{N}$ that $\xi_k^j = \|Z_k^j\|^{-1} Z_k^j$, let $\bar{X}_k^{x,j}: \Omega \rightarrow \overline{D}$, $x \in \overline{D}$, $j \in \mathbb{N}$, $k \in \mathbb{N}_0$, satisfy for all $x \in \overline{D}$, $j, k \in \mathbb{N}$ that $\bar{X}_0^{x,j} = x$ and

$$\bar{X}_k^{x,j} = \bar{X}_{k-1}^{x,j} + \bar{r}(\bar{X}_{k-1}^{x,j})\xi_k^j, \quad (5.26)$$

let $\zeta: \Omega \rightarrow [0, \infty)$ satisfy $\zeta = \inf\{t \in [0, \infty): \|B_t\| \geq 1\}$, let ρ_k^j , $k, j \in \mathbb{N}$, be i.i.d. copies of ζ , assume that ρ_k^j , Z_k^j , $j, k \in \mathbb{N}$ are independent, let $\bar{\tau}_k^{x,j}: \Omega \rightarrow [0, \infty)$, $j \in \mathbb{N}$, $k \in \mathbb{N}_0$ satisfy for all $x \in \overline{D}$, $j, k \in \mathbb{N}$ that $\bar{\tau}_0^{x,j} = 0$ and

$$\bar{\tau}_k^{x,j} = \bar{\tau}_{k-1}^{x,j} + \bar{r}^2(\bar{X}_{k-1}^{x,j})\rho_k^j. \quad (5.27)$$

Lemma 5.2.4. *For all $x \in \overline{D}$, $k \in \mathbb{N}_0$, $j \in \mathbb{N}$ it holds that $(\bar{\tau}_k^{x,j}, \bar{X}_k^{x,j})$ and $(\tau_k^x, x + B_{\tau_k^x})$ are identically distributed.*

Proof. First, note that $(\bar{\tau}_0^{x,j}, \bar{X}_0^{x,j}) = (0, x) = (\tau_0^x, x + B_{\tau_0^x})$. Next, observe that by assumption it holds that $\bar{\tau}_1^{x,j} = \bar{r}^2(x)\rho_1^j$ and $\bar{X}_1^{x,j} = x + \bar{r}(x)\xi_1^j$ are independent. Moreover, note that Lemma 5.2.1 proves that τ_1^x and $x + B_{\tau_1^x}$ are independent. Combining this, the fact that τ_1^x and $\bar{\tau}_1^{x,j}$ are identically distributed, and the fact that $x + B_{\tau_1^x}$ and $\bar{X}_1^{x,j}$ are identically distributed yields that $(\bar{\tau}_1^{x,j}, \bar{X}_1^{x,j})$ and $(\tau_1^x, x + B_{\tau_1^x})$ are identically distributed. This, the strong Markov property (cf., [61, Chapter 2.6]), and induction complete the proof of Lemma 5.2.4. $\square$

For every $\varepsilon \in (0, \infty)$ let

$$D_\varepsilon = \{x \in D : r(x) \geq \varepsilon\}. \quad (5.28)$$

For every $\varepsilon \in (0, \infty)$, $x \in D$, $j \in \mathbb{N}$ let

$$k_\varepsilon^x = \inf\{k \in \mathbb{N}_0 : x + B_{\tau_k^x} \notin D_\varepsilon\}, \quad (5.29)$$

$$\bar{k}_\varepsilon^{x,j} = \inf\{k \in \mathbb{N}_0 : \bar{X}_k^{x,j} \notin D_\varepsilon\}, \quad (5.30)$$

Lemma 5.2.5. *For all $\varepsilon \in (0, \infty)$, $x \in D$ it holds that k_ε^x and $\bar{k}_\varepsilon^{x,1}$ are identically distributed.*

Proof. The strong Markov property (cf., [61, Section 2.6]) and the assumption that ρ_k^j, Z_k^j , $j, k \in \mathbb{N}$ are independent ensure for all $x \in \overline{D}$ that $(\bar{X}_k^{x,1})_{k \in \mathbb{N}_0}$ and $(x + B_{\tau_k^x})_{k \in \mathbb{N}_0}$ are identically distributed. This implies for all $\varepsilon \in (0, \infty)$, $x \in \overline{D}$ that k_ε^x and $\bar{k}_\varepsilon^{x,1}$ are identically distributed. This completes the proof of Lemma 5.2.5. $\square$

The combination of the next two lemmas quantifies the decay of the tail probabilities of $\bar{k}_\delta^{x,1}$ the number of steps the Walk-on-Spheres algorithm needs to reach regions of the domain close to the boundary. To this end Lemma 5.2.6 provides bounds for the moment generating function of a Brownian motions first exit from a sphere and Lemma 5.2.7 establishes the estimates of the tail probabilities of $\bar{k}_\delta^{x,1}$. The statements are modifications of [20, Proposition 2.11]. We provide proofs for the convenience of the reader.

Lemma 5.2.6. *Let $\mathfrak{r} \in (0, \infty)$, $\gamma \in (0, \frac{d}{\mathfrak{r}^2})$ and let $\tau = \inf\{t \in [0, \infty) : \|B_t\| = \mathfrak{r}\}$, Then it holds that*

$$1 + \frac{\gamma \mathfrak{r}^2}{d} \leq \mathbb{E}[e^{\gamma \tau}] \leq \frac{1}{1 - \frac{\gamma \mathfrak{r}^2}{d}}. \quad (5.31)$$

Proof. Throughout this proof let $v, w : \mathbb{R}^d \rightarrow \mathbb{R}$ satisfy for all $x \in \mathbb{R}^d$ that

$$v(x) = \frac{1}{2} - \frac{\gamma \|x\|^2}{2d} \quad \text{and} \quad w(x) = 1 - \frac{\gamma \|x\|^2}{d + \gamma \mathfrak{r}^2}. \quad (5.32)$$

Itô's formula yields for every $h \in \{v, w\}$, $t \in [0, \infty)$ that

$$e^{\gamma t} h(B_t) = h(0) + \int_0^t e^{\gamma s} (\gamma h(B_s) + \frac{1}{2} \Delta h(B_s)) ds + \int_0^t e^{\gamma s} \langle \nabla h(B_s), dB_s \rangle. \quad (5.33)$$

For every $h \in \{v, w\}$ let τ_k^h , $k \in \mathbb{N}$, be a localizing sequence of stopping times for the local martingale $\int_0^\cdot e^{\gamma s} \langle \nabla h(B_s), dB_s \rangle$ and for every $h \in \{v, w\}$, $k \in \mathbb{N}$ let $T_k^h : \Omega \rightarrow \mathbb{R}$ satisfy $T_k^h = k \wedge \tau \wedge \tau_k^h$. Combining the fact that for all $h \in \{v, w\}$, $k \in \mathbb{N}$ it holds that $T_k^h \leq k \wedge \tau$, the fact that for all $h \in \{v, w\}$ it holds that $(\tau_k^h)_{k \in \mathbb{N}}$ localizes $\int_0^\cdot e^{\gamma s} \langle \nabla h(B_s), dB_s \rangle$, and the optional stopping theorem with (5.33) establishes for all $h \in \{v, w\}$, $k \in \mathbb{N}$ that

$$\mathbb{E}[e^{\gamma T_k^h} h(B_{T_k^h})] = h(0) + \mathbb{E}\left[\int_0^{T_k^h} e^{\gamma s} (\gamma h(B_s) + \frac{1}{2} \Delta h(B_s)) ds\right]. \quad (5.34)$$

Furthermore, note that the fact that for every $h \in \{v, w\}$ it holds that $\mathbb{P}(\lim_{k \rightarrow \infty} \tau_k^h = \infty) = 1$ ensures for every $h \in \{v, w\}$ that $\mathbb{P}(\lim_{k \rightarrow \infty} T_k^h = \tau) = 1$. This and the fact that $\mathbb{P}(\tau < \infty) = 1$ (cf. e.g., [61, Section 4.2], [75, Lemma 4.27]) imply for every $h \in \{v, w\}$ that

$$\mathbb{P}\left(\lim_{k \rightarrow \infty} e^{\gamma T_k^h} h(B_{T_k^h}) = e^{\gamma \tau} h(B_\tau)\right) = 1. \quad (5.35)$$

Moreover, observe that the facts that $\gamma \mathfrak{r}^2 < d$ and that for all $k \in \mathbb{N}$ it holds that $T_k^v \leq \tau$ assure for all $k \in \mathbb{N}$ that $e^{\gamma T_k^v} v(B_{T_k^v}) \geq 0$. This, the fact that for all $x \in \mathbb{R}^d$ it holds that $\gamma v(x) + \frac{1}{2} \Delta v(x) = -\frac{\gamma^2 \|x\|^2}{2d} \leq 0$, the fact that $v(B_\tau) = \frac{1}{2} - \frac{\gamma \mathfrak{r}^2}{2d} > 0$, the fact that for all $k \in \mathbb{N}$ it holds that $T_k^v \leq \tau$, (5.34), (5.35), and Fatou's lemma demonstrate that

$$\mathbb{E}[e^{\gamma \tau}] \left(\frac{1}{2} - \frac{\gamma \mathfrak{r}^2}{2d} \right) = \mathbb{E}[e^{\gamma \tau} v(B_\tau)] \leq \liminf_{k \rightarrow \infty} \mathbb{E}[e^{\gamma T_k^v} v(B_{T_k^v})] \leq v(0) = \frac{1}{2}. \quad (5.36)$$

Hence, it holds that

$$\mathbb{E}[e^{\gamma \tau}] \leq \frac{1}{1 - \frac{\gamma \mathfrak{r}^2}{d}}. \quad (5.37)$$

Next, note that the fact that for all $x \in \mathbb{R}^d$ it holds that $w(x) \leq 1$. Combining this with the fact that it holds for all $k \in \mathbb{N}$ that $T_k^w \leq \tau$ assures that for all $k \in \mathbb{N}$ it holds that $e^{\gamma T_k^w} w(B_{T_k^w}) \leq e^{\gamma \tau}$. This, the fact that for all $x \in \mathbb{R}^d$ with $\|x\| \leq \mathfrak{r}$ it holds that $\gamma w(x) + \frac{1}{2} \Delta w(x) = \frac{\gamma^2 (\mathfrak{r}^2 - \|x\|^2)}{d + \gamma \mathfrak{r}^2} \geq 0$, the fact that $w(B_\tau) = 1 - \frac{\gamma \mathfrak{r}^2}{d + \gamma \mathfrak{r}^2} > 0$, (5.34), (5.35), (5.37), and the dominated convergence theorem yield that

$$\mathbb{E}[e^{\gamma \tau}] \left(1 - \frac{\gamma \mathfrak{r}^2}{d + \gamma \mathfrak{r}^2} \right) = \mathbb{E}[e^{\gamma \tau} w(B_\tau)] = \lim_{k \rightarrow \infty} \mathbb{E}[e^{\gamma T_k^w} w(B_{T_k^w})] \geq w(0) = 1. \quad (5.38)$$

This demonstrates

$$\mathbb{E}[e^{\gamma \tau}] \geq \frac{1}{1 - \frac{\gamma \mathfrak{r}^2}{d + \gamma \mathfrak{r}^2}} = \frac{d + \gamma \mathfrak{r}^2}{d} = 1 + \frac{\gamma \mathfrak{r}^2}{d}. \quad (5.39)$$

This and (5.37) establish (5.31). This completes the proof of Lemma 5.2.6. $\square$

Lemma 5.2.7. *Assume $\delta > 0$. Then it holds for all $M \in \mathbb{N}$, $x \in D$ that*

$$\mathbb{P}\left(\bar{k}_\delta^{x,1} \geq M\right) \leq 2 \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)^{-M}. \quad (5.40)$$

Proof. Throughout this proof let $T_c: \Omega \rightarrow [0, \infty]$, $c \in (0, \infty)$, satisfy for every $c \in (0, \infty)$ that $T_c = \inf\{t \in [0, \infty): \|B_t\| = c\}$ and let $\lambda(\gamma) \in \mathbb{R}$, $\gamma \in (0, \frac{d}{\text{diam}(D)^2})$, satisfy for every $\gamma \in (0, \frac{d}{\text{diam}(D)^2})$ that $\lambda(\gamma) = \log(1 + \gamma \frac{\beta^2 \delta^2}{d})$. First, observe that the fact that for all $x \in D$, $y \in D_\delta$ it holds that $\bar{r}(x) \leq r(x)$ and $\bar{r}(y) \geq \beta r(y) \geq \beta \delta$ and Lemma 5.2.6 assure for every $\gamma \in (0, \frac{d}{\text{diam}(D)^2})$, $x \in D_\delta$ that

$$\mathbb{E}[e^{\gamma T_{\text{diam}(D)}}] \leq \frac{d}{d - \gamma \text{diam}(D)^2} \quad (5.41)$$

and

$$\mathbb{E}[e^{\gamma T_{\bar{r}(x)}}] \geq 1 + \frac{\gamma \bar{r}(x)^2}{d} \geq 1 + \gamma \frac{\beta^2 \delta^2}{d} = e^{\lambda(\gamma)}. \quad (5.42)$$

Throughout the remainder of this proof let $K_\delta: D^{\mathbb{N}_0} \rightarrow \mathbb{N}_0 \cup \{+\infty\}$ satisfy for every $x = (x_0, x_1, \dots) \in D^{\mathbb{N}_0}$ that $K_\delta(x) = \inf\{n \in \mathbb{N}_0: r(x_n) < \delta\}$. Observe that K_δ is measurable. Moreover, note that for all $x \in D$, $j \in \mathbb{N}$ it holds that $K_\delta(\bar{X}^{x,j}) = \bar{k}_\delta^{x,j}$. This, Lemma 5.2.5, the fact that $\mathbb{P}(k_\delta^x < \infty) = 1$, (5.42), the fact that for all $x \in D$ it holds that $T_{\text{diam}(D)} \geq \tau_{k_\delta^x}^x$, the

monotone convergence theorem, and the law of total expectation yield for all $\gamma \in (0, \frac{d}{\text{diam}(D)^2})$, $x \in D_\delta$ that

$$\begin{aligned}
\frac{d}{d-\gamma \text{diam}(D)^2} &\geq \mathbb{E}[e^{\gamma T_{\text{diam}(D)}}] \geq \mathbb{E}[e^{\gamma \tau_{\bar{k}_\delta^x}}] = \mathbb{E}\left[e^{\frac{\gamma \tau_{\bar{k}_\delta^{x,1}}}{\bar{k}_\delta^{x,1}}}\right] \\
&= \mathbb{E}\left[\exp\left(\gamma \sum_{i=1}^{\bar{k}_\delta^{x,1}} \bar{r}(\bar{X}_{i-1}^{x,1})^2 \rho_i^1\right)\right] \\
&= \lim_{n \rightarrow \infty} \mathbb{E}\left[\exp\left(\gamma \sum_{i=1}^{\bar{k}_\delta^{x,1} \wedge n} \bar{r}(\bar{X}_{i-1}^{x,1})^2 \rho_i^1\right)\right] \\
&= \lim_{n \rightarrow \infty} \mathbb{E}\left[\mathbb{E}\left[\exp\left(\gamma \sum_{i=1}^{\bar{k}_\delta^{x,1} \wedge n} \bar{r}(\bar{X}_{i-1}^{x,1})^2 \rho_i^1\right) \mid \bar{X}^{x,1}\right]\right].
\end{aligned} \tag{5.43}$$

The fact that $K_\delta(\bar{X}^{x,1})$ is $\sigma(\bar{X}^{x,1})$ -measurable and the fact that $\bar{X}^{x,1}$ and $(\rho_n^1)_{n \in \mathbb{N}}$ are independent ensures for every $x \in D_\delta$, $n \in \mathbb{N}$, $\gamma \in (0, \frac{d}{\text{diam}(D)^2})$ that $\phi_{n,\gamma}(\bar{X}^{x,1})$, where

$$\begin{aligned}
\phi_{n,\gamma}(y) &= \mathbb{E}\left[\exp\left(\gamma \sum_{i=1}^{K_\delta(y) \wedge n} \bar{r}(y_{i-1})^2 \rho_i^1\right)\right] \\
&= \int_{\Omega} \exp\left(\gamma \sum_{i=1}^{K_\delta(y) \wedge n} \bar{r}(y_{i-1})^2 \rho_i^1(\omega)\right) \mathbb{P}(d\omega), \quad y \in D^{\mathbb{N}_0},
\end{aligned} \tag{5.44}$$

is (a version of) the conditional expectation $\mathbb{E}[\exp(\gamma \sum_{i=1}^{\bar{k}_\delta^{x,1} \wedge n} \bar{r}(\bar{X}_{i-1}^{x,1})^2 \rho_i^1) \mid \bar{X}^{x,1}]$. The fact that for all $y \in D^{\mathbb{N}_0}$, $j \in \{0, 1, \dots, K_\delta(y)\}$ it holds that $y_{j-1} \in D_\delta$, (5.42), (5.44), and the assumption that $(\rho_n^1)_{n \in \mathbb{N}}$ are independent imply for all $y \in D^{\mathbb{N}_0}$, $n \in \mathbb{N}$, $\gamma \in (0, \frac{d}{\text{diam}(D)^2})$ that

$$\phi_{n,\gamma}(y) = \prod_{i=1}^{K_\delta(y) \wedge n} \mathbb{E}\left[e^{\gamma \bar{r}(y_{i-1})^2 \rho_i^1}\right] = \prod_{i=1}^{K_\delta(y) \wedge n} \mathbb{E}\left[e^{\gamma T_{\bar{r}(y_{i-1})}}\right] \geq \prod_{i=1}^{K_\delta(y) \wedge n} e^{\lambda(\gamma)} = e^{\lambda(\gamma)(K_\delta(y) \wedge n)}. \tag{5.45}$$

This, (5.43), and the monotone convergence theorem ensure for all $x \in D_\delta$, $\gamma \in (0, \frac{d}{\text{diam}(D)^2})$ that

$$\begin{aligned}
\frac{d}{d-\gamma \text{diam}(D)^2} &\geq \lim_{n \rightarrow \infty} \mathbb{E}\left[\mathbb{E}\left[\exp\left(\gamma \sum_{i=1}^{\bar{k}_\delta^{x,1} \wedge n} \bar{r}(\bar{X}_{i-1}^{x,1})^2 \rho_i^1\right) \mid \bar{X}^{x,1}\right]\right] = \lim_{n \rightarrow \infty} \mathbb{E}[\phi_{n,\gamma}(\bar{X}^{x,1})] \\
&\geq \lim_{n \rightarrow \infty} \mathbb{E}\left[e^{\lambda(\gamma)(K_\delta(\bar{X}^{x,1}) \wedge n)}\right] = \mathbb{E}\left[e^{\lambda(\gamma)K_\delta(\bar{X}^{x,1})}\right] \\
&= \mathbb{E}\left[e^{\lambda(\gamma)\bar{k}_\delta^{x,1}}\right] = \mathbb{E}\left[\left(1 + \gamma \frac{\beta^2 \delta^2}{d}\right)^{\bar{k}_\delta^{x,1}}\right].
\end{aligned} \tag{5.46}$$

Choosing $\gamma = \frac{d}{2 \text{diam}(D)^2}$ demonstrates for all $x \in D_\delta$ that

$$\mathbb{E}\left[\left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)^{\bar{k}_\delta^{x,1}}\right] \leq 2. \tag{5.47}$$

This and Markov's inequality yield for all $x \in D_\delta$, $M \in \mathbb{N}$ that

$$\mathbb{P}(\bar{k}_\delta^{x,1} \geq M) \leq 2 \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)^{-M}. \tag{5.48}$$

This proves (5.40). This completes the proof of Lemma 5.2.7. $\square$

The next lemma provides a probability measure to calculate expectations of integrals of time dependent functions of Brownian motion based on the time-space occupation behavior of Brownian motion.

Lemma 5.2.8. *Let $U = \{x \in \mathbb{R}^d: \|x\| \leq 1\}$, and let $\mu: \mathcal{B}([0, \infty) \times U) \rightarrow [0, \infty)$ satisfy for all $A \in \mathcal{B}([0, \infty) \times U)$ that $\mu(A) = d\mathbb{E}\left[\int_0^\zeta \mathbb{1}_A(s, B_s) ds\right]$. Then*

(i) *it holds that μ is a probability measure and*

(ii) *it holds for all measurable $f: [0, \infty) \times U \rightarrow [-\infty, \infty]$ where f is μ -integrable or satisfies $f \geq 0$ that*

$$\int_{[0, \infty) \times U} f(p) \mu(dp) = d\mathbb{E}\left[\int_0^\zeta f(s, B_s) ds\right]. \quad (5.49)$$

Proof. Throughout this proof let $\nu: \mathcal{B}([0, \infty) \times U) \ni A \mapsto \mathbb{E}\left[\int_0^\zeta \mathbb{1}_A(s, B_s) ds\right] \in [0, \infty]$. First, observe that [83, e.g., Proposition 2.2.21 or 3.1.8] ensures that $\mathbb{E}[\zeta] = \frac{1}{d}$. This proves that ν is a finite measure with $\nu([0, \infty) \times U) = \mathbb{E}[\zeta] = \frac{1}{d}$. This and the fact that for all $A \in \mathcal{B}([0, \infty) \times U)$ it holds that $\mu(A) = \frac{\nu(A)}{\mathbb{E}[\zeta]} = d\nu(A)$ prove item (i). Next, observe that through approximation by simple functions it holds for all measurable $f: [0, \infty) \times U \rightarrow [-\infty, \infty]$ where f is μ -integrable or satisfies $f \geq 0$ that

$$\int_{[0, \infty) \times U} f(p) \mu(dp) = d \int_{[0, \infty) \times U} f(p) \nu(dp) = d\mathbb{E}\left[\int_0^\zeta f(s, B_s) ds\right]. \quad (5.50)$$

This proves item (ii). The proof of Lemma 5.2.8 is thus complete. $\square$

In the following lemma, a representation of functions of the form (5.1) via the Walk-on-Spheres process is provided. In the sense of Remark 5.1.5, this establishes a Walk-on-Spheres based representation of solutions of the PDE (5.2) (see, e.g., [39, Lemma 3.4] and [20, Corollary 2.10] for similar results).

Lemma 5.2.9. *Assume $\delta = 0$, let $h_0: [0, \infty) \times \partial D \rightarrow \mathbb{R}$, $h_1: [0, \infty) \times \overline{D} \rightarrow \mathbb{R}$ be measurable and bounded, let $w: D \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that*

$$w(x) = \mathbb{E}[h_0(\tau^x, x + B_{\tau^x})] + \mathbb{E}\left[\int_0^{\tau^x} h_1(s, x + B_s) ds\right], \quad (5.51)$$

let μ be as in Lemma 5.2.8, let $(V_k, Y_k) \sim \mu$, $k \in \mathbb{N}$, be independent, assume that $((V_k, Y_k))_{k \in \mathbb{N}}$ and B are independent. Then it holds for all $x \in D$ that

$$w(x) = \mathbb{E}\left[h_0(\tau^x, x + B_{\tau^x}) + \frac{1}{d} \sum_{k=1}^{\infty} \bar{r}(x + B_{\tau_{k-1}^x})^2 h_1\left(\tau_{k-1}^x + \bar{r}(x + B_{\tau_{k-1}^x})^2 V_k, x + B_{\tau_{k-1}^x} + \bar{r}(x + B_{\tau_{k-1}^x}) Y_k\right)\right]. \quad (5.52)$$

Proof. Throughout this proof let $T_c: \Omega \rightarrow [0, \infty]$, $c \in (0, \infty)$, satisfy for every $c \in (0, \infty)$ that $T_c = \inf\{t \in [0, \infty): \|B_t\| = c\}$. First, note that the assumption that $\delta = 0$, [20, Remark 2.9], the assumption that h_1 is bounded, and the dominated convergence ensure for all $x \in D$ that

$$\mathbb{E}\left[\int_0^{\tau^x} h_1(s, x + B_s) ds\right] = \lim_{n \rightarrow \infty} \sum_{k=1}^n \mathbb{E}\left[\int_{\tau_{k-1}^x}^{\tau_k^x} h_1(s, x + B_s) ds\right]. \quad (5.53)$$

The strong Markov property (cf., [61, Section 2.6]) and the scaling property imply for all

$x \in D, k \in \mathbb{N}$ that

$$\begin{aligned}
& \mathbb{E} \left[\int_{\tau_{k-1}^x}^{\tau_k^x} h_1(s, x + B_s) ds \right] \\
&= \mathbb{E} \left[\int_0^{\tau_k^x - \tau_{k-1}^x} h_1(\tau_{k-1}^x + s, x + B_{\tau_{k-1}^x + s}) ds \right] \\
&= \mathbb{E} \left[\mathbb{E} \left[\int_0^{\tau_k^x - \tau_{k-1}^x} h_1(\tau_{k-1}^x + s, x + B_{\tau_{k-1}^x + s}) ds \mid \mathcal{F}_{\tau_{k-1}^x}^x \right] \right] \\
&= \mathbb{E} \left[\mathbb{E} \left[\int_0^{T_{\bar{r}(x+y_0)}} h_1(t_0 + s, x + y_0 + B_s) ds \right] \Big|_{(t_0, y_0) = (\tau_{k-1}^x, B_{\tau_{k-1}^x})} \right] \\
&= \mathbb{E} \left[\mathbb{E} \left[\bar{r}(x + y_0)^2 \int_0^{T_1} h_1(t_0 + \bar{r}(x + y_0)^2 s, x + y_0 + \bar{r}(x + y_0) B_s) ds \right] \Big|_{(t_0, y_0) = (\tau_{k-1}^x, B_{\tau_{k-1}^x})} \right]. \tag{5.54}
\end{aligned}$$

Throughout the remainder of this proof let $U = \{x \in \mathbb{R}^d : \|x\| \leq 1\}$. Furthermore, note that Lemma 5.2.8 and the assumption that $(V_1, Y_1) \sim \mu$ demonstrate for all measurable $\varphi : [0, \infty) \times U \rightarrow [-\infty, \infty]$ where φ is μ -integrable or satisfies $\varphi \geq 0$ that that

$$\mathbb{E} \left[\int_0^{T_1} \varphi(s, B_s) ds \right] = \frac{1}{d} \int_{[0, \infty) \times U} \varphi(v, y) \mu(d(v, y)) = \frac{1}{d} \mathbb{E}[\varphi(V, Y)]. \tag{5.55}$$

This, the assumption that $(V_k, Y_k) \sim \mu, k \in \mathbb{N}$, are independent, the assumption that $((V_k, Y_k))_{k \in \mathbb{N}}$ and B are independent, (5.53), and (5.54) establish for all $x \in D$ that

$$\begin{aligned}
& \mathbb{E} \left[\int_0^{\tau^x} h_1(s, x + B_s) ds \right] \\
&= \frac{1}{d} \mathbb{E} \left[\sum_{k=1}^{\infty} \bar{r}(x + B_{\tau_{k-1}^x})^2 h_1 \left(\tau_{k-1}^x + \bar{r}(x + B_{\tau_{k-1}^x})^2 V_k, x + B_{\tau_{k-1}^x} + \bar{r}(x + B_{\tau_{k-1}^x}) Y_k \right) \right]. \tag{5.56}
\end{aligned}$$

This proves (5.52). The proof of Lemma 5.2.9 is thus complete. $\square$

5.3. Error Analysis for the Modified Walk-on-Spheres Algorithm

In this section, we establish an error analysis for the estimator (5.5) as an approximation of expectations of type (5.1) in the L^∞ -sense. To this end, we establish in Subsection 5.3.1 bias estimates for the Monte Carlo estimator (5.5). In Subsection 5.3.2 we first present some preparatory regularity properties of the Walk-on-Spheres process as well as the Monte Carlo estimator (5.5) and conclude with an overall error estimation in the L^∞ -sense. The main results are Theorem 5.3.9 and Corollary 5.3.10. Throughout this section assume $D \subseteq \mathbb{R}^d$ to be open, bounded, and convex, let $r : \mathbb{R}^d \rightarrow [0, \infty)$ satisfy for all $x \in \mathbb{R}^d$ that $r(x) = \inf\{\|x - y\| : y \in \partial D\}$, let $\beta \in (0, 1], \delta \in (0, \infty)$, let $\bar{r} : D \rightarrow [0, \infty)$ be a (β, δ) -distance on D , let $\tau_k^x, \bar{X}_k^{x,j}, \bar{\tau}_k^{x,j}, \zeta, Z_l^j, \xi_l^j, \rho_l^j, x \in \bar{D}, j, l \in \mathbb{N}, k \in \mathbb{N}_0$, be as in Section 5.2, let $h, h_0, h_1 : [0, \infty) \times \bar{D} \rightarrow \mathbb{R}$ be measurable and bounded, let $\mu : \mathcal{B}([0, \infty) \times \{x \in \mathbb{R}^d : \|x\| \leq 1\}) \rightarrow [0, 1]$ be as in Lemma 5.2.8. For $n \in \mathbb{N}, S \subseteq \mathbb{R}^n$, and $F : S \rightarrow \mathbb{R}$ we denote with $\|F\|_\infty$ the supremum norm of F on its domain S , i.e., $\|F\|_\infty = \sup_{s \in S} |F(s)|$.

5.3.1. Bias Estimates

Lemma 5.3.1. *Let $n, M \in \mathbb{N}, L_1, L_2 \in (0, \infty)$, assume for all $t, s \in [0, \infty), x, y \in \bar{D}$ that*

$$|h(t, x) - h(s, y)| \leq L_1 |t - s| + L_2 \|x - y\|, \tag{5.57}$$

let $w: D \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that $w(x) = \mathbb{E}[h(\tau^x, x + B_{\tau^x})]$, and let $\bar{w} = \bar{w}_{\bar{r}, n, M}: D \times \Omega \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that

$$\bar{w}(x) = \frac{1}{n} \sum_{i=1}^n h(\bar{\tau}_M^{x,i}, \bar{X}_M^{x,i}). \quad (5.58)$$

Then it holds for all $x \in D$ that

$$|w(x) - \mathbb{E}[\bar{w}(x)]| \leq L_1 \text{diam}(D)\delta + L_2 d^{1/2} \text{diam}(D)^{1/2} \delta^{1/2} + 4\|h\|_\infty \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)^{-M}. \quad (5.59)$$

Proof. First, observe that Lemma 5.2.4 shows for all $x \in D$ that

$$\begin{aligned} |w(x) - \mathbb{E}[\bar{w}(x)]| &= \left| \mathbb{E}[h(\tau^x, x + B_{\tau^x})] - \frac{1}{n} \sum_{i=1}^n \mathbb{E}[h(\tau_M^{x,i}, \bar{X}_M^{x,i})] \right| \\ &= \left| \mathbb{E}[h(\tau^x, x + B_{\tau^x}) - h(\tau_M^x, x + B_{\tau_M^x})] \right| \\ &\leq \left| \mathbb{E}[(h(\tau^x, x + B_{\tau^x}) - h(\tau_M^x, x + B_{\tau_M^x})) \mathbf{1}_{\{M \geq k_\delta^x\}}] \right| \\ &\quad + \left| \mathbb{E}[(h(\tau^x, x + B_{\tau^x}) - h(\tau_M^x, x + B_{\tau_M^x})) \mathbf{1}_{\{M < k_\delta^x\}}] \right|. \end{aligned} \quad (5.60)$$

Next, note that combining the triangle inequality, the fact that $\{M < k_\delta^x\} \subseteq \{M \leq k_\delta^x\}$, and the assumption that h is bounded with Lemma 5.2.7 yields for all $x \in D$ that

$$\begin{aligned} &\left| \mathbb{E}[(h(\tau^x, x + B_{\tau^x}) - h(\tau_M^x, x + B_{\tau_M^x})) \mathbf{1}_{\{M < k_\delta^x\}}] \right| \\ &\leq 2\|h\|_\infty \mathbb{E}[\mathbf{1}_{\{M < k_\delta^x\}}] \leq 4\|h\|_\infty \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)^{-M}. \end{aligned} \quad (5.61)$$

Additionally, note that the assumption that for all $t, s \in [0, \infty)$, $x, y \in D$ it holds that $|h(t, x) - h(s, y)| \leq L_1|t - s| + L_2\|x - y\|$, the fact that $\tau^x \geq \tau_M^x$, Jensen's inequality, the strong Markov property, the fact that $\{M \geq k_\delta^x\} \subseteq \{\tau_M^x \geq \tau_{k_\delta^x}^x\}$ and Lemma 5.2.2, imply for all $x \in D$ that

$$\begin{aligned} &\left| \mathbb{E}[(h(\tau^x, x + B_{\tau^x}) - h(\tau_M^x, x + B_{\tau_M^x})) \mathbf{1}_{\{M \geq k_\delta^x\}}] \right| \\ &\leq L_1 \mathbb{E}[|\tau^x - \tau_M^x| \mathbf{1}_{\{M \geq k_\delta^x\}}] + L_2 \mathbb{E}[\|B_{\tau^x} - B_{\tau_M^x}\| \mathbf{1}_{\{M \geq k_\delta^x\}}] \\ &\leq L_1 \mathbb{E}[|\tau^x - \tau_M^x| \mathbf{1}_{\{M \geq k_\delta^x\}}] + L_2 \mathbb{E}[\|B_{\tau^x} - B_{\tau_M^x}\| \mathbf{1}_{\{\tau_M^x \geq \tau_{k_\delta^x}^x\}}] \\ &= L_1 \mathbb{E}[|\tau^x - \tau_M^x| \mathbf{1}_{\{M \geq k_\delta^x\}}] + L_2 \mathbb{E}\left[\mathbb{E}[\|B_{\tau^x} - B_{\tau_M^x}\| \mathbf{1}_{\{\tau_M^x \geq \tau_{k_\delta^x}^x\}} \mid \mathcal{F}_{\tau_M^x}^x]\right] \\ &= L_1 \mathbb{E}[|\tau^x - \tau_M^x| \mathbf{1}_{\{M \geq k_\delta^x\}}] + L_2 \mathbb{E}\left[\mathbb{E}[\|B_{\tau^x} - B_{\tau_M^x}\| \mid \mathcal{F}_{\tau_M^x}^x] \mathbf{1}_{\{\tau_M^x \geq \tau_{k_\delta^x}^x\}}\right] \\ &\leq L_1 \mathbb{E}[|\tau^x - \tau_M^x| \mathbf{1}_{\{M \geq k_\delta^x\}}] + L_2 d^{1/2} \left(\mathbb{E}\left[\mathbb{E}[|\tau^x - \tau_M^x| \mid \mathcal{F}_{\tau_M^x}^x] \mathbf{1}_{\{\tau_M^x \geq \tau_{k_\delta^x}^x\}}\right] \right)^{1/2} \\ &\leq L_1 \mathbb{E}[|\tau^x - \tau_{k_\delta^x}^x|] + L_2 d^{1/2} \left(\mathbb{E}[|\tau^x - \tau_{k_\delta^x}^x|] \right)^{1/2} \\ &\leq L_1 \mathbb{E}\left[\mathbb{E}\left[\tau^{x+B_{\tau_{k_\delta^x}^x}} \mid \mathcal{F}_{\tau_{k_\delta^x}^x}^x\right]\right] + L_2 d^{1/2} \left(\mathbb{E}\left[\mathbb{E}\left[\tau^{x+B_{\tau_{k_\delta^x}^x}} \mid \mathcal{F}_{\tau_{k_\delta^x}^x}^x\right]\right] \right)^{1/2} \\ &\leq L_1 \text{diam}(D)\delta + L_2 d^{1/2} \text{diam}(D)^{1/2} \delta^{1/2}. \end{aligned} \quad (5.62)$$

Combining this with (5.61) proves (5.59). $\square$

Lemma 5.3.2. Let $n, M \in \mathbb{N}$, let $w: D \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that

$$w(x) = \mathbb{E} \left[\int_0^{\tau^x} h(s, x + B_s) ds \right], \quad (5.63)$$

let $(V_k^i, Y_k^i) \sim \mu$, $i, k \in \mathbb{N}$, be i.i.d., assume (V_k^i, Y_k^i) , ρ_k^i , Z_k^i , $i, k \in \mathbb{N}$, and B are independent, and let $\bar{w} = \bar{w}_{\bar{r}, n, M}: D \times \Omega \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that

$$\bar{w}(x) = \frac{1}{n} \sum_{i=1}^n \frac{1}{d} \sum_{k=1}^M \bar{r}(\bar{X}_{k-1}^{x,i})^2 h\left(\bar{\tau}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i})^2 V_k^i, \bar{X}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i}) Y_k^i\right). \quad (5.64)$$

Then it holds for all $x \in D$ that

$$|w(x) - \mathbb{E}[\bar{w}(x)]| \leq \|h\|_\infty \text{diam}(D) \left(\frac{2 \text{diam}(D)}{d} \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2} \right)^{-M} + \delta \right). \quad (5.65)$$

Proof. First, observe that Lemmas 5.2.4 and 5.2.8, the techniques presented in the proof of Lemma 5.2.9, and the assumption that (V_k^1, Y_k^1) , $k \in \mathbb{N}$, are i.i.d. and independent from B establish that for all $x \in D$ it holds that

$$\begin{aligned} w(x) &= \mathbb{E} \left[\int_{\tau_M^x}^{\tau^x} h(s, x + B_s) ds \right] + \mathbb{E} \left[\int_0^{\tau_M^x} h(s, x + B_s) ds \right] \\ &= \mathbb{E} \left[\int_{\tau_M^x}^{\tau^x} h(s, x + B_s) ds \right] \\ &\quad + \frac{1}{d} \sum_{k=1}^M \mathbb{E} \left[\bar{r}(x + B_{\tau_{k-1}^x})^2 h\left(\bar{\tau}_{k-1}^x + \bar{r}(x + B_{\tau_{k-1}^x})^2 V_k^1, x + B_{\tau_{k-1}^x} + \bar{r}(x + B_{\tau_{k-1}^x}) Y_k^1\right) \right] \\ &= \mathbb{E} \left[\int_{\tau_M^x}^{\tau^x} h(s, x + B_s) ds \right] \\ &\quad + \frac{1}{d} \sum_{k=1}^M \mathbb{E} \left[\bar{r}(\bar{X}_{k-1}^{x,1})^2 h\left(\bar{\tau}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1})^2 V_k^1, \bar{X}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1}) Y_k^1\right) \right]. \end{aligned} \quad (5.66)$$

Next, note that Lemma 5.2.4 and the assumption that (V_k^i, Y_k^i) , $i, k \in \mathbb{N}$, are i.i.d. and independent from ρ_k^i , Z_k^i , $i, k \in \mathbb{N}$, and B imply for all $x \in D$ that

$$\begin{aligned} \mathbb{E}[\bar{w}(x)] &= \frac{1}{nd} \sum_{i=1}^n \sum_{k=1}^M \mathbb{E} \left[\bar{r}(\bar{X}_{k-1}^{x,i})^2 h\left(\bar{\tau}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i})^2 V_k^i, \bar{X}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i}) Y_k^i\right) \right] \\ &= \frac{1}{d} \sum_{k=1}^M \mathbb{E} \left[\bar{r}(\bar{X}_{k-1}^{x,1})^2 h\left(\bar{\tau}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1})^2 V_k^1, \bar{X}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1}) Y_k^1\right) \right]. \end{aligned} \quad (5.67)$$

Combining (5.66), (5.67), and the triangle inequality with the assumption that h is bounded yields for all $x \in D$ that

$$|w(x) - \mathbb{E}[\bar{w}]| = \left| \mathbb{E} \left[\int_{\tau_M^x}^{\tau^x} h(s, x + B_s) ds \right] \right| \leq \|h\|_\infty \mathbb{E}[\tau^x - \tau_M^x]. \quad (5.68)$$

Next, observe that the strong Markov property, [20, Corollary 2.3], and Lemma 5.2.7 demonstrate for all $x \in D$ that

$$\begin{aligned} \mathbb{E}[(\tau^x - \tau_M^x) \mathbf{1}_{\{M < k_\delta^x\}}] &= \mathbb{E} \left[\mathbb{E}[\tau^{x+B_{\tau_M^x}} \mid \mathcal{F}_{\tau_M^x}] \mathbf{1}_{\{M < k_\delta^x\}} \right] \\ &\leq \frac{\text{diam}(D)^2}{d} \mathbb{P}[k_\delta^x \geq M] \leq \frac{2 \text{diam}(D)^2}{d} \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2} \right)^{-M}. \end{aligned} \quad (5.69)$$

Moreover, note that the strong Markov property, the fact that $\tau^x \geq \tau_M^x$, and Lemma 5.2.2 imply for all $x \in D$ that

$$\begin{aligned} \mathbb{E}[(\tau^x - \tau_M^x) \mathbb{1}_{\{M \geq k_\delta^x\}}] &\leq \mathbb{E}[\tau^x - \tau_{k_\delta^x}^x] = \mathbb{E}\left[\mathbb{E}\left[\tau^{x+B_{\tau_{k_\delta^x}^x}} \mid \mathcal{F}_{\tau_{k_\delta^x}^x}\right]\right] \\ &\leq \mathbb{E}\left[r\left(x + B_{\tau_{k_\delta^x}^x}\right) \text{diam}(D)\right] \leq \delta \text{diam}(D). \end{aligned} \quad (5.70)$$

Combining this and (5.69) with (5.68) proves (5.65). The proof of Lemma 5.3.2 is thus complete. $\square$

5.3.2. Global L^∞ -Error Analysis

Lemma 5.3.3. *Let $n, M \in \mathbb{N}$, let $(V_k^i, Y_k^i) \sim \mu$, $i, k \in \mathbb{N}$, be i.i.d. assume (V_k^i, Y_k^i) , ρ_k^i , Z_k^i , $i, k \in \mathbb{N}$, are independent, and let $\bar{w} = \bar{w}_{\bar{r}, n, M}: D \times \Omega \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that*

$$\bar{w}(x) = \frac{1}{n} \sum_{i=1}^n \left[h_0(\bar{\tau}_M^{x,i}, \bar{X}_M^{x,i}) + \frac{1}{d} \sum_{k=1}^M \bar{r}(\bar{X}_{k-1}^{x,i})^2 h_1\left(\bar{\tau}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i})^2 V_k^i, \bar{X}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i}) Y_k^i\right) \right]. \quad (5.71)$$

Then it holds for all $x \in D$, $t \geq 0$ that

$$\mathbb{P}[|\bar{w}(x) - \mathbb{E}[\bar{w}(x)]| \geq t] \leq 2 \exp\left(-\frac{nt^2}{2(\|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d})^2}\right). \quad (5.72)$$

Proof. Let $\bar{w}_i = \bar{w}_{i, \delta, n, M}: D \times \Omega \rightarrow \mathbb{R}$, $i \in \mathbb{N} \cap [1, n]$, satisfy for all $i \in \mathbb{N} \cap [1, n]$, $x \in D$ that

$$\bar{w}_i(x) = h_0(\bar{\tau}_M^{x,i}, \bar{X}_M^{x,i}) + \frac{1}{d} \sum_{k=1}^M \bar{r}(\bar{X}_{k-1}^{x,i})^2 h_1\left(\bar{\tau}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i})^2 V_k^i, \bar{X}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i}) Y_k^i\right). \quad (5.73)$$

Observe that for all $x \in D$ it holds that $(\bar{w}_i)_{i \in \mathbb{N} \cap [1, n]}$ are independent and $\bar{w}(x) = \frac{1}{n} \sum_{i=1}^n \bar{w}_i(x)$. Moreover, note that the assumption that for all $x \in D$ it holds that $\bar{r}(x) \leq \text{diam}(D)$ and the assumption that h_i , $i \in \{1, 2\}$, is bounded ensure for all $i \in \mathbb{N} \cap [1, n]$, $x \in D$ it holds that

$$|\bar{w}_i(x)| \leq \|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d}. \quad (5.74)$$

This and Hoeffding's inequality show for all $x \in D$, $t \geq 0$ that

$$\begin{aligned} \mathbb{P}[|\bar{w}(x) - \mathbb{E}[\bar{w}(x)]| \geq t] &= \mathbb{P}\left[\left|\sum_{i=1}^n \bar{w}_i(x) - \mathbb{E}\left[\sum_{i=1}^n \bar{w}_i(x)\right]\right| \geq nt\right] \\ &\leq 2 \exp\left(-\frac{nt^2}{2(\|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d})^2}\right). \end{aligned} \quad (5.75)$$

This proves (5.72). The proof of Lemma 5.3.3 is thus complete. $\square$

Lemma 5.3.4. *Let $L \in [0, \infty)$ and assume for all $x, y \in D$ that $|\bar{r}(x) - \bar{r}(y)| \leq L\|x - y\|$. Then*

- (i) *it holds for all $k \in \mathbb{N}_0$, $i \in \mathbb{N}$, $x, y \in D$ that $\|\bar{X}_k^{x,i} - \bar{X}_k^{y,i}\| \leq (1 + L)^k \|x - y\|$ and*
- (ii) *it holds for all $k \in \mathbb{N}_0$, $i \in \mathbb{N}$, $x, y \in D$ that*

$$|\bar{\tau}_k^{x,i} - \bar{\tau}_k^{y,i}| \leq 2 \text{diam}(D) L \left[\sum_{j=1}^k (1 + L)^{j-1} \rho_j^i \right] \|x - y\|. \quad (5.76)$$

Proof. First, we prove item (i) via induction on $k \in \mathbb{N}_0$. Note that for all $i \in \mathbb{N}$, $x, y \in D$ it holds that $\|X_0^{x,i} - X_0^{y,i}\| = \|x - y\|$. This proves item (i) for the base case $k = 0$. Let $k \in \mathbb{N}$ and assume item (i) holds true for all $l \in \mathbb{N}_0 \cap [0, k-1]$. This, the assumption that for all $x, y \in D$ it holds that $|\bar{r}(x) - \bar{r}(y)| \leq L\|x - y\|$, and the triangle inequality show for all $i \in \mathbb{N}$, $x, y \in D$ that

$$\begin{aligned} \|\bar{X}_k^{x,i} - \bar{X}_k^{y,i}\| &\leq \|\bar{X}_{k-1}^{x,i} - \bar{X}_{k-1}^{y,i}\| + \|\xi_k^i\| |\bar{r}(\bar{X}_{k-1}^{x,i}) - \bar{r}(\bar{X}_{k-1}^{y,i})| \\ &\leq (1+L)\|\bar{X}_{k-1}^{x,i} - \bar{X}_{k-1}^{y,i}\| \leq (1+L)^k \|x - y\|. \end{aligned} \quad (5.77)$$

This proves that item (i) holds true for all $k \in \mathbb{N}_0$. Next, we prove item (ii) via induction on $k \in \mathbb{N}_0$. Observe that for all $i \in \mathbb{N}$, $x, y \in D$ it holds that $|\bar{\tau}_0^{x,i} - \bar{\tau}_0^{y,i}| = 0$. This proves item (ii) for the base case $k = 0$. Let $k \in \mathbb{N}$ and assume item (ii) holds true for all $l \in \mathbb{N}_0 \cap [0, k-1]$. This, the assumption that for all $x, y \in D$ it holds that $|\bar{r}(x) - \bar{r}(y)| \leq L\|x - y\|$, the fact that for all $x \in D$ it holds that $\bar{r}(x) \in [0, \text{diam}(D)]$, and item (i) demonstrate for all $i \in \mathbb{N}$, $x, y \in D$ that

$$\begin{aligned} |\bar{\tau}_k^{x,i} - \bar{\tau}_k^{y,i}| &\leq |\bar{\tau}_{k-1}^{x,i} - \bar{\tau}_{k-1}^{y,i}| + |\rho_k^i| |\bar{r}(\bar{X}_{k-1}^{x,i})^2 - \bar{r}(\bar{X}_{k-1}^{y,i})^2| \\ &\leq 2 \text{diam}(D) L \left[\sum_{j=1}^{k-1} (1+L)^{j-1} \rho_j^i \right] \|x - y\| + 2 \text{diam}(D) L \rho_k^i \|\bar{X}_{k-1}^{x,i} - \bar{X}_{k-1}^{y,i}\| \\ &\leq 2 \text{diam}(D) L \left[(1+L)^{k-1} \rho_k^i + \sum_{j=1}^{k-1} (1+L)^{j-1} \rho_j^i \right] \|x - y\|. \end{aligned} \quad (5.78)$$

This proves that item (ii) holds true for all $k \in \mathbb{N}_0$. The proof of Lemma 5.3.4 is thus complete. $\square$

Lemma 5.3.5. *Let $(V, Y) \sim \mu$. Then it holds that $\mathbb{E}[V] = \frac{d+4}{2d(d+2)}$.*

Proof of Lemma 5.3.5. First, note that $\mathbb{E}[\zeta] = \frac{1}{d}$ (cf. [83, e.g., Proposition 2.2.21 or 3.1.8]). Next, observe that Itô's formula yields for all $t \in [0, \infty)$ that

$$\|B_t\|^4 - 2(d+2) \int_0^t \|B_s\|^2 ds = 4 \int_0^t \langle B_s, dB_s \rangle. \quad (5.79)$$

Let $M = (M_t)_{t \in [0, \infty)}$ satisfy for all $t \in [0, \infty)$ that $M_t = 4 \int_0^{t \wedge \zeta} \langle B_s, dB_s \rangle$. This and the fact that for all $t \in [0, \zeta]$ it holds that $\|B_s\| \leq 1$ implies that the local martingale M is indeed a square integrable martingale. Thus it holds for all $t \in [0, \infty)$ that

$$\mathbb{E}[\|B_{t \wedge \zeta}\|^4] - 2(2+d) \mathbb{E} \left[\int_0^{t \wedge \zeta} \|B_s\|^2 ds \right] = 0. \quad (5.80)$$

This, the dominated convergence theorem, and the monotone convergence theorem imply that

$$1 = 2(2+d) \mathbb{E} \left[\int_0^\zeta \|B_s\|^2 ds \right]. \quad (5.81)$$

Similarly, Itô's product rule establishes for all $t \in [0, \infty)$ that

$$t\|B_t\|^2 - \frac{dt^2}{2} - \int_0^t \|B_s\|^2 ds = 2 \int_0^t s \langle B_s, dB_s \rangle. \quad (5.82)$$

Again this defines a local martingale, which stopped at ζ is a square integrable martingale. This, the dominated convergence theorem, and the monotone convergence theorem hence demonstrate that

$$\mathbb{E}[\zeta] - \frac{d}{2} \mathbb{E}[\zeta^2] - \mathbb{E} \left[\int_0^\zeta \|B_s\|^2 ds \right] = 0. \quad (5.83)$$

Combining this, the fact that $\mathbb{E}[\zeta] = \frac{1}{d}$, and (5.81) yields $\mathbb{E}[\zeta^2] = \frac{d+4}{d^2(d+2)}$. This and item (ii) in Lemma 5.2.8 establish that

$$\mathbb{E}[V] = \int_{[0,\infty) \times U} v \mu(d(v, y)) = d \mathbb{E} \left[\int_0^\zeta s \, ds \right] = \frac{d}{2} \mathbb{E}[\zeta^2] = \frac{d+4}{2d(d+2)}. \quad (5.84)$$

The proof of Lemma 5.3.5 is thus complete. $\square$

Lemma 5.3.6. *Let $n, M \in \mathbb{N}$, let $(V_k^i, Y_k^i) \sim \mu$, $i, k \in \mathbb{N}$, be i.i.d., assume (V_k^i, Y_k^i) , ρ_k^i , Z_k^i , $i, k \in \mathbb{N}$, are independent, let $L, L_j^i \in [0, \infty)$, $i \in \{0, 1\}$, $j \in \{1, 2\}$, assume for all $s, t \in [0, \infty)$, $x, y \in D$, $i \in \{0, 1\}$ that*

$$|h_i(s, x) - h_i(t, y)| \leq L_1^i |s - t| + L_2^i \|x - y\| \quad \text{and} \quad |\bar{r}(x) - \bar{r}(y)| \leq L \|x - y\|, \quad (5.85)$$

let $\bar{w} = \bar{w}_{\bar{r}, n, M}: D \times \Omega \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that

$$\bar{w}(x) = \frac{1}{n} \sum_{i=1}^n \left[h_0(\bar{\tau}_M^{x,i}, \bar{X}_M^{x,i}) + \frac{1}{d} \sum_{k=1}^M \bar{r}(\bar{X}_{k-1}^{x,i})^2 h_1 \left(\bar{\tau}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i})^2 V_k^i, \bar{X}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i}) Y_k^i \right) \right]. \quad (5.86)$$

Then

(i) it holds for all $x, y \in D$ that

$$\begin{aligned} |\mathbb{E}[\bar{w}(x)] - \mathbb{E}[\bar{w}(y)]| &\leq \|x - y\| (1 + L)^M M^2 \left(L_2^0 + \frac{2 \operatorname{diam}(D) L_1^0}{d} \right. \\ &\quad \left. + \frac{1}{d} \left(2 \operatorname{diam}(D) \|h_1\|_\infty + \operatorname{diam}(D)^2 \left(L_2^1 + \frac{2 \operatorname{diam}(D) L_1^1}{d} \right) \right) \right) \end{aligned} \quad (5.87)$$

(ii) it holds for all $x, y \in D$ that

$$\begin{aligned} |\bar{w}(x) - \bar{w}(y)| &\leq \|x - y\| (1 + L)^M \left(L_2^0 + \frac{2 L_1^0 \operatorname{diam}(D)}{n} \sum_{i=1}^n \sum_{j=1}^M \rho_j^i + \frac{2 M \operatorname{diam}(D) \|h_1\|_\infty}{d} \right. \\ &\quad \left. + \frac{2 L_1^1 \operatorname{diam}(D)^3}{d n} \sum_{i=1}^n \sum_{k=1}^M \left[V_k^i + \sum_{j=1}^{k-1} \rho_j^i \right] + \frac{L_2^1 M \operatorname{diam}(D)^2}{d} \right). \end{aligned} \quad (5.88)$$

Proof. First, note that Lemma 5.2.4, the assumption that (V_k^i, Y_k^i) , $i, k \in \mathbb{N}$ are identically distributed, and the assumption that (V_k^i, Y_k^i) , ρ_k^i, Z_k^i , $i, k \in \mathbb{N}$, are independent ensure for all $x \in D$ that

$$\begin{aligned} \mathbb{E}[\bar{w}(x)] &= \mathbb{E} \left[h_0(\bar{\tau}_M^{x,1}, \bar{X}_M^{x,1}) \right] \\ &\quad + \frac{1}{d} \sum_{k=1}^M \mathbb{E} \left[\bar{r}(\bar{X}_{k-1}^{x,1})^2 h_1 \left(\bar{\tau}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1})^2 V_k^1, \bar{X}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1}) Y_k^1 \right) \right]. \end{aligned} \quad (5.89)$$

This assures for all $x, y \in D$ that

$$\begin{aligned} |\mathbb{E}[\bar{w}(x)] - \mathbb{E}[\bar{w}(y)]| &\leq \mathbb{E} \left[\left| h_0(\bar{\tau}_M^{x,1}, \bar{X}_M^{x,1}) - h_0(\bar{\tau}_M^{y,1}, \bar{X}_M^{y,1}) \right| \right] \\ &\quad + \frac{1}{d} \sum_{k=1}^M \mathbb{E} \left[\left| \bar{r}(\bar{X}_{k-1}^{x,1})^2 h_1 \left(\bar{\tau}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1})^2 V_k^1, \bar{X}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1}) Y_k^1 \right) \right. \right. \\ &\quad \left. \left. - \bar{r}(\bar{X}_{k-1}^{y,1})^2 h_1 \left(\bar{\tau}_{k-1}^{y,1} + \bar{r}(\bar{X}_{k-1}^{y,1})^2 V_k^1, \bar{X}_{k-1}^{y,1} + \bar{r}(\bar{X}_{k-1}^{y,1}) Y_k^1 \right) \right| \right]. \end{aligned} \quad (5.90)$$

Next, observe that the assumption that for all $i \in \{0, 1\}$, $s, t \in [0, \infty)$, $x, y \in D$ it holds that $|h_i(s, x) - h_i(t, y)| \leq L_1^i |s - t| + L_2^i \|x - y\|$ and the fact that for all $x \in D$ it holds that $\bar{r}(x) \leq \text{diam}(D)$ imply for all $k \in \mathbb{N} \cap [1, M]$, $x, y \in D$ that

$$\begin{aligned}
& \left| \bar{r}(\bar{X}_{k-1}^{x,1})^2 h_1 \left(\bar{\tau}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1})^2 V_k^1, \bar{X}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1}) Y_k^1 \right) \right. \\
& \quad \left. - \bar{r}(\bar{X}_{k-1}^{y,1})^2 h_1 \left(\bar{\tau}_{k-1}^{y,1} + \bar{r}(\bar{X}_{k-1}^{y,1})^2 V_k^1, \bar{X}_{k-1}^{y,1} + \bar{r}(\bar{X}_{k-1}^{y,1}) Y_k^1 \right) \right| \\
& \leq 2 \text{diam}(D) \|h_1\|_\infty |\bar{r}(\bar{X}_{k-1}^{x,1}) - \bar{r}(\bar{X}_{k-1}^{y,1})| \\
& \quad + \text{diam}(D)^2 \left(L_1^1 \left(|\bar{\tau}_{k-1}^{x,1} - \bar{\tau}_{k-1}^{y,1}| + |V_k^1| |\bar{r}(\bar{X}_{k-1}^{x,1})^2 - \bar{r}(\bar{X}_{k-1}^{y,1})^2| \right) \right. \\
& \quad \left. + L_2^1 \left(\|\bar{X}_{k-1}^{x,1} - \bar{X}_{k-1}^{y,1}\| + \|Y_k^1\| |\bar{r}(\bar{X}_{k-1}^{x,1}) - \bar{r}(\bar{X}_{k-1}^{y,1})| \right) \right). \tag{5.91}
\end{aligned}$$

This, the assumption that for all $x, y \in D$ it holds that $|\bar{r}(x) - \bar{r}(y)| \leq L \|x - y\|$, the fact that $\mathbb{E}[\zeta] = \frac{1}{d}$ (cf. [83, e.g., Proposition 2.2.21 or 3.1.8]), the assumption that for all $i, j \in \mathbb{N}$ it holds that ρ_i^j and ζ are identically distributed, item (ii) in Lemma 5.2.8, Lemma 5.3.5, the fact that for all $k \in \mathbb{N}$ it holds that $\frac{k+4}{2k(k+2)} \leq \frac{1}{k}$, the fact that for all $j, k \in \mathbb{N}$ it holds that $\|Y_k^j\| \leq 1$, and Lemma 5.3.4 establish for all $x, y \in D$ that

$$\begin{aligned}
& \frac{1}{d} \sum_{k=1}^M \mathbb{E} \left[\left| \bar{r}(\bar{X}_{k-1}^{x,1})^2 h_1 \left(\bar{\tau}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1})^2 V_k^1, \bar{X}_{k-1}^{x,1} + \bar{r}(\bar{X}_{k-1}^{x,1}) Y_k^1 \right) \right. \right. \\
& \quad \left. \left. - \bar{r}(\bar{X}_{k-1}^{y,1})^2 h_1 \left(\bar{\tau}_{k-1}^{y,1} + \bar{r}(\bar{X}_{k-1}^{y,1})^2 V_k^1, \bar{X}_{k-1}^{y,1} + \bar{r}(\bar{X}_{k-1}^{y,1}) Y_k^1 \right) \right| \right] \\
& \leq \|x - y\| \frac{1}{d} \sum_{k=1}^M \left[(1 + L)^k \left(2 \text{diam}(D) \|h_1\|_\infty \right. \right. \\
& \quad \left. \left. + \text{diam}(D)^2 \left(L_2^1 + 2L_1^1 \text{diam}(D) \left(\mathbb{E}[V_k^1] + \sum_{j=1}^{k-1} \mathbb{E}[\rho_j^1] \right) \right) \right) \right] \\
& \leq \|x - y\| (1 + L)^M \frac{M^2}{d} \left(2 \text{diam}(D) \|h_1\|_\infty + \text{diam}(D)^2 \left(L_2^1 + \frac{2L_1^1 \text{diam}(D)}{d} \right) \right). \tag{5.92}
\end{aligned}$$

Similarly, it holds for all $x, y \in D$ that

$$\mathbb{E} \left[\left| h_0(\bar{\tau}_M^{x,1}, \bar{X}_M^{x,1}) - h_0(\bar{\tau}_M^{y,1}, \bar{X}_M^{y,1}) \right| \right] \leq \|x - y\| (1 + L)^M \left(L_2^0 + \frac{2ML_1^0 \text{diam}(D)}{d} \right). \tag{5.93}$$

Combining this and (5.92) yields for all $x, y \in D$ that

$$\begin{aligned}
|\mathbb{E}[\bar{w}(x)] - \mathbb{E}[\bar{w}(y)]| & \leq \|x - y\| (1 + L)^M M^2 \left(L_2^0 + \frac{2L_1^0 \text{diam}(D)}{d} \right. \\
& \quad \left. + \frac{1}{d} \left(2 \text{diam}(D) \|h_1\|_\infty + \text{diam}(D)^2 \left(L_2^1 + \frac{2L_1^1 \text{diam}(D)}{d} \right) \right) \right) \tag{5.94}
\end{aligned}$$

This proves item (i). Moreover, note that item (ii) is proven analogously. The proof of Lemma 5.3.6 is thus complete. $\square$

Lemma 5.3.7. *Let $(V, Y) \sim \mu$. Then it holds for all $t \in [0, \infty)$ that*

$$\mathbb{P}[\zeta \geq t] \leq 2e^{-\frac{d}{2}t} \quad \text{and} \quad \mathbb{P}[V \geq t] \leq 4e^{-\frac{d}{2}t}. \tag{5.95}$$

Proof of Lemma 5.3.7. First, observe that Lemma 5.2.6 yields that $\mathbb{E}[e^{\frac{d}{2}\zeta}] \leq 2$. Combining this with Markov's inequality demonstrates for all $t \in [0, \infty)$ that

$$\mathbb{P}[\zeta \geq t] \leq \mathbb{E}[e^{\frac{d}{2}\zeta}] e^{-\frac{d}{2}t} \leq 2e^{-\frac{d}{2}t}. \tag{5.96}$$

Next, observe that item (ii) in Lemma 5.2.8 and Fubini's theorem establish for all $t \geq 0$ that

$$\mathbb{P}[V \geq t] = \mathbb{E}[\mathbb{1}_{[t, \infty)}(V)] = d\mathbb{E}\left[\int_0^\zeta \mathbb{1}_{[t, \infty)}(s) ds\right] = d \int_t^\infty \mathbb{P}[\zeta \geq s] ds. \quad (5.97)$$

This and (5.96) show for all $t \in [0, \infty)$ that

$$\mathbb{P}[V \geq t] \leq 2d \int_t^\infty e^{-\frac{d}{2}s} ds = 4e^{-\frac{d}{2}t}. \quad (5.98)$$

The proof of Lemma 5.3.7 is thus complete. $\square$

The next lemma is a modification of [20, Lemma 4.5]. We include a proof for the convenience of the reader.

Lemma 5.3.8. *Let $X: \Omega \rightarrow [0, \infty)$ be a random variable, let $c_1, c_2, c_3, c_4, A \geq 0$, assume $c_2, c_4 > 0$, assume for all $t \in [0, \infty)$ that $\mathbb{P}[X \geq t] \leq 2(e^{c_1 - c_2((t-A)^+)^2} + e^{c_3 - c_4(t-A)^+})$. Then it holds that*

$$\mathbb{E}[X] \leq A + \frac{\sqrt{c_1 + \log(2)} + 1}{\sqrt{c_2}} + \frac{c_3 + \log(2) + 1}{c_4}. \quad (5.99)$$

Proof of Lemma 5.3.8. Throughout this proof let $\lambda \in \mathbb{R}$ satisfy

$$\lambda = A + \frac{\sqrt{c_1 + \log(2)}}{\sqrt{c_2}} + \frac{c_3 + \log(2)}{c_4}. \quad (5.100)$$

Observe that $\lambda > A$. This and the assumption that for all $t \in [0, \infty)$ it holds that $\mathbb{P}[X \geq t] \leq 2(e^{c_1 - c_2((t-A)^+)^2} + e^{c_3 - c_4(t-A)^+})$ ensures that

$$\begin{aligned} \mathbb{E}[X] &= \int_0^\infty \mathbb{P}[X \geq t] dt \leq \lambda + \int_\lambda^\infty \mathbb{P}[X \geq t] dt \\ &\leq \lambda + 2e^{c_1} \int_\lambda^\infty e^{-c_2(t-A)^2} dt + 2e^{c_3} \int_\lambda^\infty e^{-c_4(t-A)} dt. \end{aligned} \quad (5.101)$$

Next, note that the fact that it holds that $\lambda - A \geq \frac{c_3 + \log(2)}{c_4}$ implies that

$$2e^{c_3} \int_\lambda^\infty e^{-c_4(t-A)} dt = \frac{2}{c_4} e^{c_3 - c_4(\lambda - A)} \leq \frac{2}{c_4} e^{-\log(2)} = \frac{1}{c_4}. \quad (5.102)$$

Moreover, observe that the fact that $\lambda - A \geq \frac{\sqrt{c_1 + \log(2)}}{\sqrt{c_2}}$ and the fact that $2\sqrt{\log(2)} \geq 1$ establish that

$$\begin{aligned} 2e^{c_1} \int_\lambda^\infty e^{-c_2(t-A)^2} dt &= 2e^{c_1} \int_{\lambda-A}^\infty e^{-c_2 s^2} ds \leq \frac{2e^{c_1}}{\lambda - A} \int_{\lambda-A}^\infty s e^{-c_2 s^2} ds \\ &\leq \frac{e^{c_1}}{(\lambda - A)c_2} e^{-c_2(\lambda-A)^2} \leq \frac{e^{-\log(2)}}{\sqrt{c_2}\sqrt{c_1 + \log(2)}} \leq \frac{1}{\sqrt{c_2}}. \end{aligned} \quad (5.103)$$

Combining this, (5.101), and (5.102) with (5.100) yields

$$\mathbb{E}[X] \leq \lambda + \frac{1}{\sqrt{c_2}} + \frac{1}{c_4} = A + \frac{\sqrt{c_1 + \log(2)} + 1}{\sqrt{c_2}} + \frac{c_3 + \log(2) + 1}{c_4}. \quad (5.104)$$

The proof of Lemma 5.3.8 is thus complete. $\square$

The following Theorem demonstrates that the estimator (5.5) approximates functions (5.1) in the L^∞ -sense. The proof follows similar arguments as those presented in [20, Proof of Theorem 2.26].

Theorem 5.3.9. Let $n, M \in \mathbb{N}$, let $(V_k^i, Y_k^i) \sim \mu$, $i, k \in \mathbb{N}$, be i.i.d., assume (V_k^i, Y_k^i) , ρ_k^i , Z_k^i , $i, k \in \mathbb{N}$, and B are independent, let $L, L_j^i \in [0, \infty)$, $i \in \{0, 1\}$, $j \in \{1, 2\}$, assume for all $s, t \in [0, \infty)$, $x, y \in D$, $i \in \{0, 1\}$ that

$$|h_i(s, x) - h_i(t, y)| \leq L_1^i |s - t| + L_2^i \|x - y\| \quad \text{and} \quad |\bar{r}(x) - \bar{r}(y)| \leq L \|x - y\|, \quad (5.105)$$

let $w: D \rightarrow \mathbb{R}$ be continuous and satisfy for all $x \in D$ that

$$w(x) = \mathbb{E}[h_0(\tau^x, x + B_{\tau^x})] + \mathbb{E}\left[\int_0^{\tau^x} h_1(s, x + B_s) ds\right], \quad (5.106)$$

let $\bar{w} = \bar{w}_{\bar{r}, n, M}: D \times \Omega \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that

$$\bar{w}(x) = \frac{1}{n} \sum_{i=1}^n \left[h_0(\bar{\tau}_M^{x,i}, \bar{X}_M^{x,i}) + \frac{1}{d} \sum_{k=1}^M \bar{r}(\bar{X}_{k-1}^{x,i})^2 h_1\left(\bar{\tau}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i})^2 V_k^i, \bar{X}_{k-1}^{x,i} + \bar{r}(\bar{X}_{k-1}^{x,i}) Y_k^i\right) \right]. \quad (5.107)$$

Then it holds for all $K \in \mathbb{N}$ that

$$\begin{aligned} & \mathbb{E} \left[\sup_{x \in D} |w(x) - \bar{w}(x)| \right] \\ & \leq (L_1^0 + \|h_1\|_\infty) \text{diam}(D) \delta + L_2^0 d^{1/2} \text{diam}(D)^{1/2} \delta^{1/2} \\ & \quad + \left(4\|h_0\|_\infty + \frac{2 \text{diam}(D)^2 \|h_1\|_\infty}{d} \right) \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2} \right)^{-M} \\ & \quad + \left(L_2^0 + \frac{2L_1^0 \text{diam}(D)}{d} + \frac{1}{d} \left(2 \text{diam}(D) \|h_1\|_\infty + \text{diam}(D)^2 \left(L_2^1 + \frac{2L_1^1 \text{diam}(D)}{d} \right) \right) \right) \frac{d^{1/2} \text{diam}(D)}{K} \\ & \quad + \left(L_2^0 + \frac{2M \text{diam}(D) \|h_1\|_\infty + L_2^1 M \text{diam}(D)^2}{d} \right) \frac{d^{1/2} \text{diam}(D)}{KM^2} \\ & \quad + \frac{3\sqrt{2} \left(\sqrt{d(\log(K) + 2\log(M) + M \log(1+L)) + \log(2)} + 1 \right) \left(\|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d} \right)}{\sqrt{n}} \\ & \quad + \frac{12 \text{diam}(D)^2 (\log(n) + \log(M) + \log(6) + 1) (L_1^0 d + L_1^1 \text{diam}(D)^2 M)}{d^{3/2} KM}. \end{aligned} \quad (5.108)$$

Proof. We first note that the supremum $\sup_{x \in D} |w(x) - \bar{w}(x)|$ is measurable because of the continuity of w and $\bar{w}$. Throughout this proof, let $K \in \mathbb{N}$ be fixed and assume without loss of generality that $D \subseteq [0, \text{diam}(D)]^d$. Let $\mathfrak{D} \subseteq \mathbb{R}^d$ satisfy that

$$\mathfrak{D} = \left\{ \frac{i \text{diam}(D)}{KM^2(1+L)^M} : i \in \mathbb{N}_0 \cap [0, KM^2(1+L)^M] \right\}^d \cap D. \quad (5.109)$$

Note that for all $x \in D$ there exist $i_1, \dots, i_d \in \mathbb{N}_0 \cap [0, KM^2(1+L)^M]$ such that

$$x \in \left[\frac{i_1 \text{diam}(D)}{KM^2(1+L)^M}, \frac{(i_1+1) \text{diam}(D)}{KM^2(1+L)^M} \right) \times \dots \times \left[\frac{i_d \text{diam}(D)}{KM^2(1+L)^M}, \frac{(i_d+1) \text{diam}(D)}{KM^2(1+L)^M} \right). \quad (5.110)$$

This implies for all $x \in D$ that there exists $x_{\mathfrak{D}} \in \mathfrak{D}$ such that $\|x - x_{\mathfrak{D}}\| \leq d^{1/2} \frac{\text{diam}(D)}{KM^2(1+L)^M}$.

This, Lemma 5.3.1, Lemma 5.3.2, and Lemma 5.3.6 demonstrate

$$\begin{aligned}
& \sup_{x \in D} |w(x) - \bar{w}(x)| \\
& \leq \sup_{x \in D} |w(x) - \mathbb{E}[\bar{w}(x)]| + \sup_{x \in D} |\bar{w}(x) - \bar{w}(x_{\mathfrak{D}})| + \sup_{x \in D} |\mathbb{E}[\bar{w}(x)] - \mathbb{E}[\bar{w}(x_{\mathfrak{D}})]| \\
& \quad + \sup_{y \in \mathfrak{D}} |\bar{w}(y) - \mathbb{E}[\bar{w}(y)]| \\
& \leq (L_1^0 + \|h_1\|_{\infty}) \text{diam}(D) \delta + L_2^0 d^{1/2} \text{diam}(D)^{1/2} \delta^{1/2} \\
& \quad + \left(4\|h_0\|_{\infty} + \frac{2 \text{diam}(D)^2 \|h_1\|_{\infty}}{d}\right) \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)^{-M} \\
& \quad + \left(L_2^0 + \frac{2L_1^0 \text{diam}(D)}{d} + \frac{1}{d} \left(2 \text{diam}(D) \|h_1\|_{\infty} + \text{diam}(D)^2 \left(L_2^1 + \frac{2L_1^1 \text{diam}(D)}{d}\right)\right)\right) \frac{d^{1/2} \text{diam}(D)}{K} \\
& \quad + \left(L_2^0 + \frac{2M \text{diam}(D) \|h_1\|_{\infty} + L_2^1 M \text{diam}(D)^2}{d}\right) \frac{d^{1/2} \text{diam}(D)}{KM^2} \\
& \quad + \left(\frac{2L_1^0 \text{diam}(D)}{n} \sum_{i=1}^n \sum_{j=1}^M \rho_j^i + \frac{2L_1^1 \text{diam}(D)^3}{dn} \sum_{i=1}^n \sum_{k=1}^M \left[V_k^i + \sum_{j=1}^{k-1} \rho_j^i\right]\right) \frac{d^{1/2} \text{diam}(D)}{KM^2} \\
& \quad + \sup_{y \in \mathfrak{D}} |\bar{w}(y) - \mathbb{E}[\bar{w}(y)]| \\
& = C(h_0, h_1, \delta, d, D, K, M) + Z,
\end{aligned} \tag{5.111}$$

where $C(h_0, h_1, \delta, d, D, K, M) \in [0, \infty)$ and $Z: \Omega \rightarrow [0, \infty)$ satisfy

$$\begin{aligned}
& C(h_0, h_1, \delta, d, D, K, M) \\
& = (L_1^0 + \|h_1\|_{\infty}) \text{diam}(D) \delta + L_2^0 d^{1/2} \text{diam}(D)^{1/2} \delta^{1/2} \\
& \quad + \left(4\|h_0\|_{\infty} + \frac{2 \text{diam}(D)^2 \|h_1\|_{\infty}}{d}\right) \left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)^{-M} \\
& \quad + \left(L_2^0 + \frac{2L_1^0 \text{diam}(D)}{d} + \frac{1}{d} \left(2 \text{diam}(D) \|h_1\|_{\infty} + \text{diam}(D)^2 \left(L_2^1 + \frac{2L_1^1 \text{diam}(D)}{d}\right)\right)\right) \frac{d^{1/2} \text{diam}(D)}{K} \\
& \quad + \left(L_2^0 + \frac{2M \text{diam}(D) \|h_1\|_{\infty} + L_2^1 M \text{diam}(D)^2}{d}\right) \frac{d^{1/2} \text{diam}(D)}{KM^2}
\end{aligned} \tag{5.112}$$

and

$$\begin{aligned}
Z & = \left(\frac{2L_1^0 \text{diam}(D)}{n} \sum_{i=1}^n \sum_{j=1}^M \rho_j^i + \frac{2L_1^1 \text{diam}(D)^3}{dn} \sum_{i=1}^n \sum_{k=1}^M \left[V_k^i + \sum_{j=1}^{k-1} \rho_j^i\right]\right) \frac{d^{1/2} \text{diam}(D)}{KM^2} \\
& \quad + \sup_{y \in \mathfrak{D}} |\bar{w}(y) - \mathbb{E}[\bar{w}(y)]|.
\end{aligned} \tag{5.113}$$

Moreover, note that the fact that for all $i \in \mathbb{N}$ it holds that $\sum_{k=1}^M \sum_{j=1}^{k-1} \rho_k^i \leq M \sum_{k=1}^M \rho_j^i$ ensures that

$$Z \leq \frac{2 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M)}{d^{1/2} n K M^2} \sum_{i=1}^n \sum_{k=1}^M \rho_k^i + \frac{2L_1^1 \text{diam}(D)^4}{d^{1/2} n K M^2} \sum_{i=1}^n \sum_{k=1}^M V_k^i + \sup_{y \in \mathfrak{D}} |\bar{w}(y) - \mathbb{E}[\bar{w}(y)]|. \tag{5.114}$$

Next, observe that the union bound inequality, Lemma 5.3.3, and (5.109) assure for all $t \in [0, \infty)$ that

$$\begin{aligned}
& \mathbb{P} \left[\sup_{y \in \mathfrak{D}} |\bar{w}(y) - \mathbb{E}[\bar{w}(y)]| \geq \frac{t}{3} \right] \leq \sum_{y \in \mathfrak{D}} 2 \exp \left(- \frac{nt^2}{18 \left(\|h_0\|_{\infty} + \frac{M \text{diam}(D)^2 \|h_1\|_{\infty}}{d} \right)^2} \right) \\
& \leq 2 \exp \left(d(\log(K) + 2 \log(M) + M \log(1 + L)) - \frac{nt^2}{18 \left(\|h_0\|_{\infty} + \frac{M \text{diam}(D)^2 \|h_1\|_{\infty}}{d} \right)^2} \right).
\end{aligned} \tag{5.115}$$

Moreover, note that Lemma 5.3.7 and the assumption that V_k^i , $i, k \in \mathbb{N}$, are identically distributed demonstrate for all $t \in [0, \infty)$ that

$$\mathbb{P}\left[\frac{2L_1^1 \text{diam}(D)^4}{d^{1/2}nKM^2} \sum_{i=1}^n \sum_{k=1}^M V_k^i \geq \frac{t}{3}\right] \leq nM\mathbb{P}\left[V_1^1 \geq \frac{td^{1/2}KM}{6L_1^1 \text{diam}(D)^4}\right] \leq 4nM \exp\left(-\frac{td^{3/2}KM}{12L_1^1 \text{diam}(D)^4}\right). \quad (5.116)$$

Furthermore, observe that Lemma 5.3.7 and the assumption that ρ_k^i , $i, k \in \mathbb{N}$, are identically distributed imply for all $t \in [0, \infty)$ that

$$\begin{aligned} \mathbb{P}\left[\frac{2 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M)}{d^{1/2}nKM^2} \sum_{i=1}^n \sum_{k=1}^M \rho_k^i \geq \frac{t}{3}\right] &\leq nM\mathbb{P}\left[\rho_1^1 \geq \frac{td^{1/2}KM}{6 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M)}\right] \\ &\leq 2nM \exp\left(-\frac{td^{3/2}KM}{12 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M)}\right). \end{aligned} \quad (5.117)$$

Combining (5.114), (5.115), (5.116), and (5.117) with the union bound inequality establishes for all $t \in [0, \infty)$ that

$$\begin{aligned} &\mathbb{P}(Z \geq t) \\ &\leq 2 \exp\left(d(\log(K) + 2\log(M) + M\log(1+L)) - \frac{nt^2}{18\left(\|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d}\right)^2}\right) \\ &\quad + 4nM \exp\left(-\frac{td^{3/2}KM}{12L_1^1 \text{diam}(D)^4}\right) + 2nM \exp\left(-\frac{td^{3/2}KM}{12 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M)}\right) \\ &\leq 2 \left(\exp\left(d(\log(K) + 2\log(M) + M\log(1+L)) - \frac{nt^2}{18\left(\|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d}\right)^2}\right) \right. \\ &\quad \left. + \exp\left(\log(3) + \log(n) + \log(M) - \frac{td^{3/2}KM}{12 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M)}\right) \right). \end{aligned} \quad (5.118)$$

This, (5.111), and Lemma 5.3.8 (applied with $X \curvearrowright Z$, $A \curvearrowright 0$,

$$\begin{aligned} c_1 \curvearrowright d(\log(K) + 2\log(M) + M\log(1+L)), \quad c_2 \curvearrowright \frac{n}{18\left(\|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d}\right)^2}, \\ c_3 \curvearrowright \log(3) + \log(n) + \log(M), \quad c_4 \curvearrowright \frac{d^{3/2}KM}{12 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M)} \end{aligned} \quad (5.119)$$

in the notation of Lemma 5.3.8) prove (5.108). The proof of Theorem 5.3.9 is thus complete. $\square$

Corollary 5.3.10. *Assume the situation of Theorem 5.3.9, let $\varepsilon \in (0, 1)$, and assume*

$$\begin{aligned} \delta &= \min\left\{1, \frac{\varepsilon^2}{9((L_1^0 + \|h_1\|_\infty) \text{diam}(D) + L_2^0 d^{1/2} \text{diam}(D)^{1/2})^2}\right\}, \\ M &= \max\left\{1, \left\lceil \frac{\log(6(2\|h_0\|_\infty d + \text{diam}(D)\|h_1\|_\infty)) - \log(d) + \log(\varepsilon^{-1})}{\log\left(1 + \frac{\beta^2 \delta^2}{2 \text{diam}(D)^2}\right)} \right\rceil \right\}, \quad \text{and} \\ n &= K^2 = \max\{1, \lceil 9C(h_0, h_1, d, D, M)^2 \varepsilon^{-2} \rceil^2\}, \end{aligned} \quad (5.120)$$

where $C(h_0, h_1, d, D, M) \in [0, \infty)$ satisfies

$$\begin{aligned} &C(h_0, h_1, d, D, M) \\ &= \left(L_2^0 + \frac{2L_1^0 \text{diam}(D)}{d} + \frac{1}{d} \left(2 \text{diam}(D)\|h_1\|_\infty + \text{diam}(D)^2 \left(L_2^1 + \frac{2L_1^1 \text{diam}(D)}{d}\right)\right)\right) d^{1/2} \text{diam}(D) \\ &\quad + \left(L_2^0 + \frac{2M \text{diam}(D)\|h_1\|_\infty + L_2^1 M \text{diam}(D)^2}{d}\right) \frac{d^{1/2} \text{diam}(D)}{M^2} \\ &\quad + 3\sqrt{2} \left(1 + \sqrt{d} + \sqrt{d(2\log(M) + M\log(1+L)) + \log(2)}\right) \left(\|h_0\|_\infty + \frac{M \text{diam}(D)^2 \|h_1\|_\infty}{d}\right) \\ &\quad + \frac{12 \text{diam}(D)^2 (L_1^0 d + L_1^1 \text{diam}(D)^2 M) (\log(6) + 3 + \log(M))}{d^{3/2} M}. \end{aligned} \quad (5.121)$$

Then it holds that $\mathbb{E}[\sup_{x \in D} |w(x) - \bar{w}(x)|] \leq \varepsilon$ and there exist $c, q \in (0, \infty)$ such that $\max\{n, M\} \leq cd^q \varepsilon^{-q}$, where q is an absolute constant and c depends only on $L, L_1^0, L_2^0, L_1^1, L_2^1, \|h_0\|_\infty, \|h_1\|_\infty, \beta^{-1}, \text{diam}(D)$, and d and this dependency is at most polynomial.

In this chapter, we developed Walk-on-Spheres Monte Carlo approximations for expectations of the form (5.1) arising as stochastic representations of solutions of elliptic PDEs of the type (5.2). In particular, we conducted an L^∞ -type error analysis and established bounds on the complexity of the resulting estimators. Combining these results with the ANN framework in Chapter 3, we develop next in Chapter 6 ANN representations and approximations for the occurring processes and Monte Carlo estimators to achieve ANN approximations for expectations of the form (5.1) for the special case $h_1 \equiv 0$.

Chapter 6.

Deep Neural Network Approximation of Elliptic PDEs with Drift and Killing

This chapter is based on Section 5 of the joint work [64] with T. Kruse.

In this chapter, we transfer the Monte Carlo results from Chapter 5 to approximation via ANNs. For this purpose, we employ the ANN framework presented in Chapter 3 in combination with the Monte Carlo estimators introduced in Chapter 5 to obtain ANN approximations. While the Monte Carlo analysis in Chapter 5 was conducted for expectations (5.1) involving both h_0 and h_1 , we consider in this chapter ANN approximations of functions $w: D \rightarrow \mathbb{R}$,

$$w(x) = \mathbb{E}[h(\tau^x, x + B_{\tau^x})], \quad x \in D. \quad (6.1)$$

This corresponds to the special case of (5.1) with $h_0 = h$ and $h_1 \equiv 0$. We refer to Remark 6.2.2 for a discussion on how to include nonzero h_1 as well.

The main contribution of this chapter, Theorem 6.2.1, develops an ANN approximation result for functions of the form (6.1). More precisely, under the assumption that h can be represented by an ANN and the distance to the boundary $\text{dist}(\cdot, \partial D)$ can be suitably approximated by ANNs, w can be approximated in the L^∞ -sense with accuracy $\varepsilon \in (0, 1)$ by an ANN Ψ such that $\mathcal{P}(\Psi) \leq \bar{c}\varepsilon^{-\bar{q}}$, where $\bar{q} \in [0, \infty)$ is an absolute constant and $\bar{c} \in [0, \infty)$ depends only on the number of parameters, the Lipschitz constant, and the least upper bound on $[0, \infty) \times \bar{D}$ of the realization of the ANN representing h , the number of parameters and a uniform Lipschitz constant of the ANNs approximating the distance to the boundary, $\text{diam}(D)$, and d . Moreover, this dependency is at most polynomial. In particular, under similar polynomial bounds on the problem quantities as in Chapter 5, the number of parameters $\mathcal{P}(\Psi)$ of the approximating ANN Ψ grows at most polynomially in the dimension d and the reciprocal ε^{-1} of the approximation accuracy (see Remark 6.2.2 below).

Through the stochastic representation established in Chapter 5, these approximation results yield corresponding approximation results for associated elliptic PDEs.

For the remainder of this chapter, we recall the notation and definitions presented in Chapter 3. We specialize the activation function to the ReLU activation function given by $\phi: \mathbb{R} \ni x \mapsto \max\{x, 0\} \in \mathbb{R}$ and subsequently write for every $\Phi \in \mathbf{N}$ the realization function $\mathcal{R}_\phi(\Phi)$ of the ANN Φ as $\mathcal{R}(\Phi)$.

The remainder of this chapter is structured as follows. In Section 6.1, we provide ANN representation and approximation results for the Walk-on-Spheres process introduced in Section 5.2 and for the Monte Carlo estimator studied in Section 5.3. In Section 6.2, we combine the ANN representations and approximations from Section 6.1 with the error analysis established in Section 5.3 to obtain an ANN approximation of functions that satisfy (6.1) in the L^∞ -sense.

6.1. ANN Approximations of Modified Walk-on-Spheres Algorithms

The following lemma is a version of [106, Proposition 3], providing an ANN approximation of the product function $\mathbb{R}^2 \ni (x, y) \rightarrow xy \in \mathbb{R}$ on compact intervals.

Lemma 6.1.1. *There exist absolute constants $\kappa_1, \kappa_2, \kappa_3 \in [0, \infty)$ such that for all $c \in (0, \infty)$, $\varepsilon \in (0, 1]$ there exists $\Pi \in \mathbf{N}$ such that*

- (i) *it holds that $\mathcal{R}(\Pi) \in C(\mathbb{R}^2, \mathbb{R})$,*
- (ii) *it holds for all $x, y \in [-c, c]$ that $|(\mathcal{R}(\Pi))(x, y) - xy| \leq \varepsilon$, and*
- (iii) *it holds¹ that $\max\{\mathcal{L}(\Pi), \|\mathcal{D}(\Pi)\|\} \leq \kappa_1 + \kappa_2 \log_2(\frac{1}{\varepsilon}) + \kappa_3(\log_2(c))^+$.*

The next two lemmas establish ANN representations and approximations of all processes appearing in the estimators of Section 5.3. More precisely, Lemma 6.1.2 establishes an ANN architecture that is used to represent the Walk-on-Spheres process and Lemma 6.1.3 provides an ANN architecture that is used to approximate the corresponding exit times. The proofs are consequences of induction as well as composition (see, e.g., Lemma 3.2.2), parallelization (see, e.g., Lemma 3.3.2 and Definition 3.3.3), and linear combination (see, e.g., Lemmas 3.4.6 and 3.4.7) of ANNs. Additionally, in the proof of Lemma 6.1.3 the ANN approximation of the product function Lemma 6.1.1 is used. A similar statement for the Walk-on-Spheres process itself (cf. Lemma 6.1.2) is formulated and proved in [20, Lemma 3.6, Corollary 3.7].

Lemma 6.1.2. *Let $d \in \mathbb{N}$, let $\mathbf{r} \in \mathbf{N}$ satisfy that $\mathcal{R}(\mathbf{r}) \in C(\mathbb{R}^d, \mathbb{R})$ and $\min\{\mathcal{L}(\mathbf{r}), \|\mathcal{D}(\mathbf{r})\|\} \geq 2$, let $(p_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}^d$. Then there exist $(\mathbf{X}_k)_{k \in \mathbb{N}_0} \subseteq \mathbf{N}$ such that*

- (i) *it holds for all $k \in \mathbb{N}$, $x \in D$ that $\mathcal{R}(\mathbf{X}_0), \mathcal{R}(\mathbf{X}_k) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $(\mathcal{R}(\mathbf{X}_0))(x) = x$, and*

$$(\mathcal{R}(\mathbf{X}_k))(x) = (\mathcal{R}(\mathbf{X}_{k-1}))(x) + (\mathcal{R}(\mathbf{r}))((\mathcal{R}(\mathbf{X}_{k-1}))(x))p_k, \quad (6.2)$$

and

- (ii) *it holds for all $k \in \mathbb{N}$ that $\mathcal{L}(\mathbf{X}_0) = 1$, $\|\mathcal{D}(\mathbf{X}_0)\| = d$,*

$$\mathcal{L}(\mathbf{X}_k) = k(\mathcal{L}(\mathbf{r}) - 1) + 1, \quad \text{and} \quad \|\mathcal{D}(\mathbf{X}_k)\| \leq 2d + \|\mathcal{D}(\mathbf{r})\|. \quad (6.3)$$

Lemma 6.1.3. *Let $d \in \mathbb{N}$, let $D \subseteq \mathbb{R}^d$ be open, bounded, let $\mathbf{r} \in \mathbf{N}$ satisfy for all $x \in D$ that $\mathcal{R}(\mathbf{r}) \in C(\mathbb{R}^d, \mathbb{R})$, $\min\{\mathcal{L}(\mathbf{r}), \|\mathcal{D}(\mathbf{r})\|\} \geq 2$ and $0 \leq (\mathcal{R}(\mathbf{r}))(x) \leq r(x)$, let $(p_k)_{k \in \mathbb{N}} \subseteq \mathbb{R}^d$ satisfy for all $k \in \mathbb{N}$ that $\|p_k\| = 1$, let $(v_k)_{k \in \mathbb{N}} \subseteq [0, \infty)$, and for every $x \in D$ let $(P_k^x)_{k \in \mathbb{N}_0} \subseteq \mathbb{R}^d$, $(T_k^x)_{k \in \mathbb{N}_0} \subseteq [0, \infty)$ satisfy for every $k \in \mathbb{N}$ that $P_0^x = x$, $P_k^x = P_{k-1}^x + (\mathcal{R}(\mathbf{r}))(P_{k-1}^x)p_k$, $T_0^x = 0$ and $T_k^x = T_{k-1}^x + ((\mathcal{R}(\mathbf{r}))(P_{k-1}^x))^2 v_k$. Then there exist $(\Theta_k^\varepsilon)_{(k, \varepsilon) \in \mathbb{N}_0 \times (0, 1]} \subseteq \mathbf{N}$*

- (i) *it holds for all $k \in \mathbb{N}$, $\varepsilon \in (0, 1]$, $x \in D$ that $\mathcal{R}(\Theta_0^\varepsilon), \mathcal{R}(\Theta_k^\varepsilon) \in C(\mathbb{R}^d, \mathbb{R})$, $(\mathcal{R}(\Theta_0^\varepsilon))(x) = 0$,*

$$|(\mathcal{R}(\Theta_k^\varepsilon))(x) - T_k^x| \leq \varepsilon \sum_{j=1}^k v_j, \quad (6.4)$$

and

- (ii) *it holds for all $k \in \mathbb{N}$, $\varepsilon \in (0, 1]$ that $\mathcal{L}(\Theta_0^\varepsilon) = 1$, $\|\mathcal{D}(\Theta_0^\varepsilon)\| = d$,*

$$\begin{aligned} \mathcal{L}(\Theta_k^\varepsilon) &\leq \kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)) + k(\mathcal{L}(\mathbf{r}) - 1), \quad \text{and} \\ \|\mathcal{D}(\Theta_k^\varepsilon)\| &\leq \max\{2, d\} + k \max\left\{\kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)), 2d + \|\mathcal{D}(\mathbf{r})\|\right\}, \end{aligned} \quad (6.5)$$

where $\kappa_1, \kappa_2, \kappa_3 \in [0, \infty)$ are the absolute constants from Lemma 6.1.1.

¹For every $n \in \mathbb{N}$, $x \in (x_1, \dots, x_n) \in \mathbb{R}^n$ let $\|\mathbf{x}\| \in \mathbb{R}$ satisfy $\|\mathbf{x}\| = \max_{i \in \{1, \dots, n\}} |x_i|$ throughout the remainder of this chapter.

The following lemma provides an ANN approximation of the Monte Carlo estimator (5.5) in the case of $h_1 \equiv 0$. Before stating Lemma 6.1.4, we briefly explain some of the mathematical objects appearing in Lemma 6.1.4. The ANN $\mathbf{r} \in \mathbf{N}$ in Lemma 6.1.4 denotes a function which determines the radii of the successive spheres on which the Walk-on-Spheres process takes place. For $x \in D$ and index $i \in \{1, \dots, n\}$ the sequence $(P_k^{x,i})_{k \in \mathbb{N}_0} \subseteq \mathbb{R}^d$ denotes the (deterministic) Walk-on-Spheres process, with start in $x \in D$, unit directions $(p_k^i)_{k \in \mathbb{N}} \subseteq \mathbb{R}^d$, and means of determining the radii of the successive spheres through the ANN $\mathbf{r}$. The sequence $(T_k^{x,i})_{k \in \mathbb{N}_0} \subseteq [0, \infty)$ denotes the corresponding exit times of successive spheres of the associated process $(P_k^{x,i})_{k \in \mathbb{N}_0}$, where the real number $v_k^i \in [0, \infty)$ takes on the role of the sampled exit time of a Brownian motion from the unit sphere.

Lemma 6.1.4. *Let $d, n, M \in \mathbb{N}$, $\varepsilon \in (0, 1]$, let $D \subseteq \mathbb{R}^d$ be open, bounded, let $L \in [0, \infty)$, let $\mathbf{h}, \mathbf{r} \in \mathbf{N}$ satisfy for all $x \in \bar{D}$, $s, t \in \mathbb{R}$ that $\mathcal{R}(\mathbf{h}) \in C(\mathbb{R}^{d+1}, \mathbb{R})$, $\mathcal{R}(\mathbf{r}) \in C(\mathbb{R}^d, \mathbb{R})$, $|(\mathcal{R}(\mathbf{h}))(s, x) - (\mathcal{R}(\mathbf{h}))(t, x)| \leq L|s - t|$, $\min\{\mathcal{L}(\mathbf{r}), \|\mathcal{D}(\mathbf{r})\|\} \geq 2$, and $0 \leq (\mathcal{R}(\mathbf{r}))(x) \leq r(x)$, for every $i \in \mathbb{N} \cap [1, n]$ let $(p_k^i)_{k \in \mathbb{N}} \subseteq \mathbb{R}^d$ satisfy for all $k \in \mathbb{N}$ that $\|p_k^i\| = 1$, for every $i \in \mathbb{N} \cap [1, n]$ let $(v_k^i)_{k \in \mathbb{N}} \subseteq [0, \infty)$, and for every $x \in D$, $i \in \mathbb{N} \cap [1, n]$ let $(P_k^{x,i})_{k \in \mathbb{N}_0} \subseteq \mathbb{R}^d$, $(T_k^{x,i})_{k \in \mathbb{N}_0} \subseteq [0, \infty)$ satisfy for every $k \in \mathbb{N}$ that $P_0^{x,i} = x$, $P_k^{x,i} = P_{k-1}^{x,i} + (\mathcal{R}(\mathbf{r}))(P_{k-1}^{x,i})p_k^i$, $T_0^{x,i} = 0$ and $T_k^{x,i} = T_{k-1}^{x,i} + ((\mathcal{R}(\mathbf{r}))(P_{k-1}^{x,i}))^2 v_k^i$. Then there exists $\Phi = \Phi_{n,M}^{\mathbf{r},\varepsilon} \in \mathbf{N}$ such that*

(i) *it holds for all $x \in D$ that $\mathcal{R}(\Phi) \in C(\mathbb{R}^d, \mathbb{R})$ and*

$$\left| (\mathcal{R}(\Phi))(x) - \frac{1}{n} \sum_{i=1}^n (\mathcal{R}(\mathbf{h}))(T_M^{x,i}, P_M^{x,i}) \right| \leq \frac{\varepsilon L}{n} \sum_{i=1}^n \sum_{k=1}^M v_k^i \quad (6.6)$$

and

(ii) *it holds that*

$$\begin{aligned} \mathcal{L}(\Phi) &\leq \mathcal{L}(\mathbf{h}) + \kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)) + M(\mathcal{L}(\mathbf{r}) - 1) \quad \text{and} \\ \|\mathcal{D}(\Phi)\| &\leq n \max\left\{\|\mathcal{D}(\mathbf{h})\|, \max\{2, d\} + 2d + \|\mathcal{D}(\mathbf{r})\|\right. \\ &\quad \left.+ M \max\left\{\kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)), 2d + \|\mathcal{D}(\mathbf{r})\|\right\}\right\}, \end{aligned} \quad (6.7)$$

where $\kappa_1, \kappa_2, \kappa_3 \in [0, \infty)$ are the absolute constants from Lemma 6.1.1.

Proof. First, observe that Lemma 6.1.2 (applied for every $i \in \mathbb{N} \cap [1, n]$ with $d \curvearrowright d$, $\mathbf{r} \curvearrowright \mathbf{r}$, $(p_k)_{k \in \mathbb{N}} \curvearrowright (p_k^i)_{k \in \mathbb{N}}$ in the notation of Lemma 6.1.2) and Lemma 6.1.3 (applied for every $i \in \mathbb{N} \cap [1, n]$ with $d \curvearrowright d$, $D \curvearrowright D$, $\mathbf{r} \curvearrowright \mathbf{r}$, $(p_k)_{k \in \mathbb{N}} \curvearrowright (p_k^i)_{k \in \mathbb{N}}$, $(v_k)_{k \in \mathbb{N}} \curvearrowright (v_k^i)_{k \in \mathbb{N}}$, $(P_k^x)_{(x,k) \in D \times \mathbb{N}_0} \curvearrowright (P_k^{x,i})_{(x,k) \in D \times \mathbb{N}_0}$, $(T_k^x)_{(x,k) \in D \times \mathbb{N}_0} \curvearrowright (T_k^{x,i})_{(x,k) \in D \times \mathbb{N}_0}$ in the notation of Lemma 6.1.3) demonstrate that there exist $(\mathbf{X}_k^i)_{(i,k) \in (\mathbb{N} \cap [1, n] \times \mathbb{N}_0)}$, $(\Theta_k^{\varepsilon,i})_{(i,k) \in (\mathbb{N} \cap [1, n] \times \mathbb{N}_0)} \subseteq \mathbf{N}$ such that

(I) it holds for all $i \in \mathbb{N} \cap [1, n]$, $k \in \mathbb{N}_0$, $x \in D$ that $\mathcal{R}(\mathbf{X}_k^i) \in C(\mathbb{R}^d, \mathbb{R}^d)$, $\mathcal{R}(\Theta_k^{\varepsilon,i}) \in C(\mathbb{R}^d, \mathbb{R})$,

$$(\mathcal{R}(\mathbf{X}_k^i))(x) = P_k^{x,i} \quad \text{and} \quad \left| (\mathcal{R}(\Theta_k^{\varepsilon,i}))(x) - T_k^x \right| \leq \varepsilon \sum_{j=1}^k v_j, \quad (6.8)$$

and

(II) it holds for all $i \in \mathbb{N} \cap [1, n]$, $k \in \mathbb{N}$ that $\mathcal{D}(\mathbf{X}_k^i) = \mathcal{D}(\mathbf{X}_k^1)$, $\mathcal{D}(\Theta_k^{\varepsilon,i}) = \mathcal{D}(\Theta_k^{\varepsilon,1})$, $\mathcal{L}(\mathbf{X}_0^i) = \mathcal{L}(\Theta_0^{\varepsilon,i}) = 1$, $\|\mathcal{D}(\mathbf{X}_0^i)\| = \|\mathcal{D}(\Theta_0^{\varepsilon,i})\| = d$,

$$\mathcal{L}(\mathbf{X}_k^i) = k(\mathcal{L}(\mathbf{r}) - 1) + 1, \quad \|\mathcal{D}(\mathbf{X}_k^i)\| \leq 2d + \|\mathcal{D}(\mathbf{r})\| \quad (6.9)$$

$$\mathcal{L}(\Theta_k^{\varepsilon,i}) \leq \kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)) + k(\mathcal{L}(\mathbf{r}) - 1), \quad \text{and}$$

$$\|\mathcal{D}(\Theta_k^{\varepsilon,i})\| \leq \max\{2, d\} + k \max\left\{\kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)), 2d + \|\mathcal{D}(\mathbf{r})\|\right\},$$

where $\kappa_1, \kappa_2, \kappa_3 \in [0, \infty)$ are the absolute constants from Lemma 6.1.1.

This Lemma 3.2.2, Corollary 3.3.4, and Lemma 3.4.6 establish that there exists $\Phi \in \mathbf{N}$ such that

(A) it holds for all $x \in D$ that $\mathcal{R}(\Phi) \in C(\mathbb{R}^d, \mathbb{R})$ and

$$(\mathcal{R}(\Phi))(x) = \frac{1}{n} \sum_{i=1}^n (\mathcal{R}(\mathbf{h})) \left((\mathcal{R}(\Theta_M^{\varepsilon,i}))(x), (\mathcal{R}(\mathbf{X}_M^i))(x) \right) \quad (6.10)$$

and

(B) it holds that

$$\begin{aligned} \mathcal{L}(\Phi) &\leq \mathcal{L}(\mathbf{h}) + \kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)) + M(\mathcal{L}(\mathbf{r}) - 1) \quad \text{and} \\ \|\mathcal{D}(\Phi)\| &\leq n \max \left\{ \|\mathcal{D}(\mathbf{h})\|, \max\{2, d\} + 2d + \|\mathcal{D}(\mathbf{r})\| \right. \\ &\quad \left. + M \max\left\{ \kappa_1 + \kappa_2 \log_2\left(\frac{1}{\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)), 2d + \|\mathcal{D}(\mathbf{r})\| \right\} \right\}. \end{aligned} \quad (6.11)$$

Observe that item (A), item (I), the assumption that for all $x \in \overline{D}$, $s, t \in \mathbb{R}$ it holds that $|(\mathcal{R}(\mathbf{h}))(s, x) - (\mathcal{R}(\mathbf{h}))(t, x)| \leq L|s - t|$, and the triangle inequality show for all $x \in D$ that

$$\left| (\mathcal{R}(\Phi))(x) - \frac{1}{n} \sum_{i=1}^n (\mathcal{R}(\mathbf{h}))(T_M^{x,i}, P_M^{x,i}) \right| \leq \frac{L}{n} \sum_{i=1}^n \left| (\mathcal{R}(\Theta_M^{\varepsilon,i}))(x) - T_M^{x,i} \right| \leq \frac{\varepsilon L}{n} \sum_{i=1}^n \sum_{k=1}^M v_k^i. \quad (6.12)$$

Combining this with item (B) proves items (i) and (ii). The proof of Lemma 6.1.4 is thus complete. $\square$

6.2. ANN Approximations for Solutions of Elliptic PDEs

The next theorem is the main result of this section. We establish that under the assumptions stated below expectations of the form (6.1) can be approximated by ANNs in the L^∞ -sense, where the size of the approximating ANNs is bounded by a polynomial quantity depending on the problem inputs. In Remark 6.2.2, we discuss the implications of Theorem 6.2.1 for overcoming the curse of dimensionality in the approximation of expectations of the form (6.1) by ANNs. Moreover, we comment on further generalizations of Theorem 6.2.1.

Theorem 6.2.1. *Let $d \in \mathbb{N}$, let $D \subseteq \mathbb{R}^d$ be open, bounded, and convex, let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space, let $B: [0, \infty) \times \Omega \rightarrow \mathbb{R}^d$ be a standard Brownian motion, for every $x \in \overline{D}$ let $\tau^x = \inf\{t \in [0, \infty): x + B_t \in \partial D\}$, let $L_1, L_2 \in (0, \infty)$, let $\mathbf{h} \in \mathbf{N}$ satisfy for all $s, t \in [0, \infty)$, $x, y \in \overline{D}$ that $\mathcal{R}(\mathbf{h}) \in C(\mathbb{R}^{d+1}, \mathbb{R})$, $\sup_{(u,z) \in [0, \infty) \times \overline{D}} |(\mathcal{R}(\mathbf{h}))(u, z)| < \infty$, and*

$$|(\mathcal{R}(\mathbf{h}))(s, x) - (\mathcal{R}(\mathbf{h}))(t, y)| \leq L_1|s - t| + L_2\|x - y\|, \quad (6.13)$$

let $w: D \rightarrow \mathbb{R}$ be continuous and satisfy for all $x \in D$ that $w(x) = \mathbb{E}[(\mathcal{R}(\mathbf{h}))(\tau^x, x + B_{\tau^x})]$, let $\eta \in [0, \infty)$ be an absolute constant, let $\mathbf{C} \in [0, \infty)$, and let $(\mathbf{r}_\varepsilon)_{\varepsilon \in (0,1)} \subseteq \mathbf{N}$ satisfy for all $x, y \in D$, $\varepsilon \in (0, 1)$ that $\mathcal{R}(\mathbf{r}_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{P}(\mathbf{r}_\varepsilon) \leq \mathbf{C}\varepsilon^{-\eta}$,

$$|(\mathcal{R}(\mathbf{r}_\varepsilon))(x) - r(x)| \leq \varepsilon, \quad \text{and} \quad |(\mathcal{R}(\mathbf{r}_\varepsilon))(x) - (\mathcal{R}(\mathbf{r}_\varepsilon))(y)| \leq \mathbf{C}\|x - y\|. \quad (6.14)$$

Then there exist an absolute constant $q \in [0, \infty)$ and a constant $c \in [0, \infty)$ that depends only on $\text{diam}(D)$, $\sup_{(u,z) \in [0, \infty) \times \overline{D}} |(\mathcal{R}(\mathbf{h}))(u, z)|$, L_1 , L_2 , $\mathcal{P}(\mathbf{h})$, $\mathbf{C}$, and d and this dependency is at most polynomial such that for every $\varepsilon \in (0, 1)$ there exists $\Psi_\varepsilon \in \mathbf{N}$ with $\mathcal{R}(\Psi_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R})$,

$$\sup_{x \in D} |w(x) - (\mathcal{R}(\Psi_\varepsilon))(x)| \leq \varepsilon, \quad \text{and} \quad \mathcal{P}(\Psi_\varepsilon) \leq c\varepsilon^{-q}. \quad (6.15)$$

Proof. Throughout this proof let $\mathbf{B} = \sup_{(u,z) \in [0,\infty) \times \overline{D}} |(\mathcal{R}(\mathbf{h}))(u, z)|$, let $Z_k^j: \Omega \rightarrow \mathbb{R}^d$, $j, k \in \mathbb{N}$, be standard normally distributed, let $\xi_k^j: \Omega \rightarrow \mathbb{R}^d$, $j, k \in \mathbb{N}$, satisfy for all $j, k \in \mathbb{N}$ that $\xi_k^j = \|Z_k^j\|^{-1} Z_k^j$, let $\zeta: \Omega \rightarrow [0, \infty)$ satisfy $\zeta = \inf\{t \in [0, \infty) \mid \|B_t\| \geq 1\}$, let ρ_k^j , $k, j \in \mathbb{N}$, be i.i.d. copies of ζ , assume that ρ_k^j , Z_k^j , $j, k \in \mathbb{N}$ are independent, let $(\delta_\varepsilon)_{\varepsilon \in (0,1)} \subseteq (0, 1]$ satisfy for all $\varepsilon \in (0, 1)$ that

$$\delta_\varepsilon = \min\left\{1, \frac{\varepsilon^2}{36(L_1 \text{diam}(D) + L_2 \text{diam}(D)^{1/2} d^{1/2})^2}\right\}, \quad (6.16)$$

let $(M_\varepsilon)_{\varepsilon \in (0,1)} \subseteq \mathbb{N}$ satisfy for all $\varepsilon \in (0, 1)$ that

$$M_\varepsilon = \max\left\{1, \left\lceil \frac{\log(24\mathbf{B}) + \log(\varepsilon^{-1})}{\log\left(1 + \frac{\delta_\varepsilon^2}{18 \text{diam}(D)^2}\right)} \right\rceil \right\}, \quad (6.17)$$

let $(\gamma_\varepsilon)_{\varepsilon \in (0,1)} \subseteq (0, 1]$ satisfy for all $\varepsilon \in (0, 1)$ that

$$\gamma_\varepsilon = \min\left\{1, \frac{d\varepsilon}{2L_1 M_\varepsilon}\right\}, \quad (6.18)$$

and let $(n_\varepsilon)_{\varepsilon \in (0,1)} \subseteq \mathbb{N}$ satisfy for all $\varepsilon \in (0, 1)$ that

$$n_\varepsilon = \max\left\{1, \lceil 36c(d, M_\varepsilon, \mathbf{h}, D, \mathbf{C})^2 \varepsilon^{-2} \rceil^2\right\}, \quad (6.19)$$

where

$$\begin{aligned} c(d, M_\varepsilon, \mathbf{h}, D, \mathbf{C}) &= \left(L_2 + \frac{2L_1 \text{diam}(D)}{d}\right) d^{1/2} \text{diam}(D) + L_2 \frac{d^{1/2} \text{diam}(D)}{M_\varepsilon^2} \\ &\quad + 3\sqrt{2}\mathbf{B} \left(1 + d^{1/2} + \sqrt{d(2\log(M_\varepsilon) + M_\varepsilon \log(1 + \mathbf{C})) + \log(2)}\right) \\ &\quad + \frac{12 \text{diam}(D)^2 L_1 (\log(6) + 3 + \log(M_\varepsilon))}{d^{1/2} M_\varepsilon}. \end{aligned} \quad (6.20)$$

Combining [1, Lemma 3.2], Lemma 3.2.2, and (6.14) with [20, Remark 2.8] demonstrates for all $\varepsilon \in (0, 1)$ that there exists $\mathbf{b}_\varepsilon \in \mathbf{N}$ such that it holds for all $x \in \mathbb{R}^d$ that $\mathcal{R}(\mathbf{b}_\varepsilon) \in C(\mathbb{R}^d, \mathbb{R})$, $\mathcal{L}(\mathbf{b}_\varepsilon) = \mathcal{L}(\mathbf{r}_{\frac{\delta_\varepsilon}{3}}) + 1$, $\|\mathcal{D}(\mathbf{b}_\varepsilon)\| \leq \|\mathcal{D}(\mathbf{r}_{\frac{\delta_\varepsilon}{3}})\|$,

$$(\mathcal{R}(\mathbf{b}_\varepsilon))(x) = ((\mathcal{R}(\mathbf{r}_{\frac{\delta_\varepsilon}{3}}))(x) - \varepsilon)^+, \quad (6.21)$$

and that $\mathcal{R}(\mathbf{b}_\varepsilon)$ is a $(\frac{1}{3}, \delta_\varepsilon)$ -distance (cf., Definition 5.2.3). Subsequently, for $\varepsilon \in (0, 1)$ we denote by $(\bar{X}_k^{x,j,\varepsilon})_{(x,k,j) \in D \times \mathbb{N}_0 \times \mathbb{N}}$ the (family of) modified Walk-on-Spheres process and by $(\bar{\tau}_k^{x,j,\varepsilon})_{(x,k,j) \in D \times \mathbb{N}_0 \times \mathbb{N}}$ the (family of) associated (physical) times generated by the $(\frac{1}{3}, \delta_\varepsilon)$ -distance $\mathcal{R}(\mathbf{b}_\varepsilon)$ as well as the random variables ξ_k^i, ρ_k^i , $i, k \in \mathbb{N}$ (cf., Section 5.2). Moreover, for $\varepsilon \in (0, 1)$ let $\bar{w}_\varepsilon: D \times \Omega \rightarrow \mathbb{R}$ satisfy for all $x \in D$ that

$$\bar{w}_\varepsilon(x) = \frac{1}{n_\varepsilon} \sum_{i=1}^{n_\varepsilon} (\mathcal{R}(\mathbf{h}))(\bar{\tau}_{M_\varepsilon}^{x,i,\varepsilon}, \bar{X}_{M_\varepsilon}^{x,i,\varepsilon}). \quad (6.22)$$

Next, observe that Lemma 6.1.4 (applied for every $\omega \in \Omega$, $\varepsilon \in (0, 1)$ with $d \curvearrowright d$, $n \curvearrowright n_\varepsilon$, $M \curvearrowright M_\varepsilon$, $\varepsilon \curvearrowright \gamma_\varepsilon$, $D \curvearrowright D$, $L \curvearrowright L_1$, $\mathbf{h} \curvearrowright \mathbf{h}$, $\mathbf{r} \curvearrowright \mathbf{b}_\varepsilon$, $(p_k^i)_{(i,k) \in (\mathbb{N} \cap [1,n]) \times \mathbb{N}} \curvearrowright (\xi_k^i(\omega))_{(i,k) \in (\mathbb{N} \cap [1,n]) \times \mathbb{N}}$, $(v_k^i)_{(i,k) \in (\mathbb{N} \cap [1,n]) \times \mathbb{N}} \curvearrowright (\rho_k^i(\omega))_{(i,k) \in (\mathbb{N} \cap [1,n]) \times \mathbb{N}}$ in the notation of Lemma 6.1.4) demonstrates that for every $\varepsilon \in (0, 1)$ there exists $\Phi_\varepsilon: \Omega \rightarrow \mathbf{N}$ such that

(I) it holds for all $\omega \in \Omega$, $x \in D$ that $\mathcal{R}(\Phi_\varepsilon(\omega)) \in C(\mathbb{R}^d, \mathbb{R})$ and

$$|(\mathcal{R}(\Phi_\varepsilon(\omega)))(x) - \bar{w}_\varepsilon(x, \omega)| \leq \frac{\gamma_\varepsilon L_1}{n_\varepsilon} \sum_{i=1}^{n_\varepsilon} \sum_{k=1}^{M_\varepsilon} \rho_k^i(\omega) \quad (6.23)$$

and

(II) it holds for all $\omega \in \Omega$ that

$$\begin{aligned} \mathcal{L}(\Phi_\varepsilon(\omega)) &\leq \mathcal{L}(\mathfrak{h}) + \kappa_1 + \kappa_2 \log_2\left(\frac{1}{\gamma_\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)) + M_\varepsilon(\mathcal{L}(\mathfrak{b}_\varepsilon) - 1) \quad \text{and} \\ \|\mathcal{D}(\Phi_\varepsilon(\omega))\| &\leq n_\varepsilon \max \left\{ \|\mathcal{D}(\mathfrak{h})\|, \max\{2, d\} + 2d + \|\mathcal{D}(\mathfrak{b}_\varepsilon)\| \right. \\ &\quad \left. + M_\varepsilon \max \left\{ \kappa_1 + \kappa_2 \log_2\left(\frac{1}{\gamma_\varepsilon}\right) + \kappa_3 \log_2(\text{diam}(D)), 2d + \|\mathcal{D}(\mathfrak{b}_\varepsilon)\| \right\} \right\}, \end{aligned} \quad (6.24)$$

where $\kappa_1, \kappa_2, \kappa_3 \in [0, \infty)$ are the absolute constants from Lemma 6.1.1.

This, (6.18), the assumption that (ρ_k^i) , $i, k \in \mathbb{N}$, are identically distributed, and the fact that $\mathbb{E}[\rho_1^1] = \frac{1}{d}$ (see [83, e.g., Proposition 2.2.21 or 3.1.8]) imply that for all $\varepsilon \in (0, 1)$ it holds that

$$\mathbb{E} \left[\sup_{x \in D} |(\mathcal{R}(\Phi_\varepsilon))(x) - \bar{w}_\varepsilon(x)| \right] \leq \frac{\gamma_\varepsilon L_1 M_\varepsilon}{d} = \frac{\varepsilon}{2}. \quad (6.25)$$

Next, note that Corollary 5.3.10 (applied for every $\varepsilon \in (0, 1)$ with $\varepsilon \curvearrowright \frac{\varepsilon}{2}$, $h_0 \curvearrowright \mathcal{R}(\mathfrak{h})$, $L_1^0 \curvearrowright L_1$, $L_2^0 \curvearrowright L_2$, $h_1 \curvearrowright 0$, $L_1^1 \curvearrowright 0$, $L_2^1 \curvearrowright 0$, $L \curvearrowright L$, $\bar{r} \curvearrowright \mathcal{R}(\mathfrak{b}_\varepsilon)$ in the notation of Corollary 5.3.10) yields for all $\varepsilon \in (0, 1)$ that

$$\mathbb{E} \left[\sup_{x \in D} |w(x) - \bar{w}_\varepsilon(x)| \right] \leq \frac{\varepsilon}{2}. \quad (6.26)$$

This, (6.25), and the triangle inequality yield for all $\varepsilon \in (0, 1)$ that

$$\mathbb{E} \left[\sup_{x \in D} |w(x) - (\mathcal{R}(\Phi_\varepsilon))(x)| \right] \leq \varepsilon. \quad (6.27)$$

This implies that there exists $\omega_\varepsilon \in \Omega$, $\varepsilon \in (0, 1)$, such that for all $\varepsilon \in (0, 1)$ it holds that

$$\sup_{x \in D} |w(x) - (\mathcal{R}(\Phi_\varepsilon(\omega_\varepsilon)))(x)| \leq \varepsilon. \quad (6.28)$$

Next, observe that (6.16) and Jensen's inequality ensure for all $\varepsilon \in (0, 1)$ that

$$\begin{aligned} \delta_\varepsilon^{-1} &= \max \left\{ 1, 36(L_1 \text{diam}(D) + L_2 \text{diam}(D)^{1/2} d^{1/2})^2 \varepsilon^{-2} \right\} \\ &\leq (1 + 72 \text{diam}(D)(L_1^2 \text{diam}(D) + L_2^2)) d \varepsilon^{-2}. \end{aligned} \quad (6.29)$$

This, (6.17), the fact that for all $x \in (0, \infty)$ it holds that $\log(x) \leq x$, and the fact that for all $x \in (-1, \infty)$ it holds that $\log(1+x) \geq \frac{x}{1+x}$ demonstrate for all $\varepsilon \in (0, 1)$ that

$$\begin{aligned} M_\varepsilon &= \max \left\{ 1, \left\lceil \frac{\log(24\mathbf{B}) + \log(\varepsilon^{-1})}{\log\left(1 + \frac{\delta_\varepsilon^2}{18 \text{diam}(D)^2}\right)} \right\rceil \right\}, \\ &\leq 2 + 18(\log(24\mathbf{B}) + \log(\varepsilon^{-1})) \text{diam}(D)^2 \delta_\varepsilon^{-2} + \log(24\mathbf{B}) + \log(\varepsilon^{-1}) \\ &\leq c_1(\mathfrak{h}, D) d^2 \varepsilon^{-5}, \end{aligned} \quad (6.30)$$

where

$$\begin{aligned} c_1(\mathfrak{h}, D) &= 3 + \log(24\mathbf{B}) \\ &\quad + 18(\log(24\mathbf{B}) + 1) \text{diam}(D)^2 (1 + 72 \text{diam}(D)(L_1^2 \text{diam}(D) + L_2^2))^2 \end{aligned} \quad (6.31)$$

This, the fact that for all $x \in (0, \infty)$ it holds that $\log_2(x) \leq x$, and (6.18) imply for all $\varepsilon \in (0, 1)$ that

$$\log_2(\gamma_\varepsilon^{-1}) \leq \gamma_\varepsilon^{-1} = \max\{1, \frac{2L_1 M_\varepsilon}{d\varepsilon}\} \leq (1 + 2L_1 c_1(\mathfrak{h}, D))d\varepsilon^{-6}. \quad (6.32)$$

Furthermore, observe that (6.20) and the fact that for all $\varepsilon \in (0, 1)$ it holds that $d, M_\varepsilon \in \mathbb{N}$ assure for all $\varepsilon \in (0, 1)$ that

$$c(d, M_\varepsilon, \mathfrak{h}, D) \leq c_2(\mathfrak{h}, D, \mathbf{C})M_\varepsilon^{1/2}d^{1/2}, \quad (6.33)$$

where

$$\begin{aligned} c_2(\mathfrak{h}, D, \mathbf{C}) &= 2 \operatorname{diam}(D)(L_2 + 6 \operatorname{diam}(D)L_1(1 + \log(6))) \\ &\quad + 3\sqrt{2}\mathbf{B}(2 + \sqrt{2 + \log(2) + \log(1 + \mathbf{C})}). \end{aligned} \quad (6.34)$$

This, (6.19), and Jensen's inequality establish for all $\varepsilon \in (0, 1)$ that

$$n_\varepsilon \leq 1 + (1 + 36c(d, M_\varepsilon, \mathfrak{h}, D, \mathbf{C})^2\varepsilon^{-2})^2 \leq 3 + 2592c_2(\mathfrak{h}, D, \mathbf{C})^4M_\varepsilon^2d^2\varepsilon^{-4}. \quad (6.35)$$

This and (6.30) show for all $\varepsilon \in (0, 1)$ that

$$n_\varepsilon \leq c_3(\mathfrak{h}, D, \mathbf{C})d^6\varepsilon^{-14}, \quad (6.36)$$

where $c_3(\mathfrak{h}, D, \mathbf{C}) = 3 + 2592c_1(\mathfrak{h}, D)^2c_2(\mathfrak{h}, D, \mathbf{C})^4$. Combining this, (6.28), the assumption that for all $\varepsilon \in (0, 1)$ it holds that $\mathcal{P}(\mathfrak{r}_\varepsilon) \leq \mathbf{C}\varepsilon^{-\eta}$, (6.29), (6.30), (6.32), and Lemma 3.1.3 with item (II) proves (6.15). The proof of Theorem 6.2.1 is thus complete. $\square$

Remark 6.2.2. We conclude this chapter with a few remarks concerning the conclusion of Theorem 6.2.1 as well as ways to extend the result.

- (i) Observe that Theorem 6.2.1 states an explicit polynomial dependency of the constant c in its conclusion on the problem dimension d . Moreover, additional implicit dependencies of c on d may arise through the remaining problem parameters appearing in the theorem such as for example $\operatorname{diam}(D)$. Theorem 6.2.1 implies that under the assumptions that these terms grow at most polynomially in the dimension $d \in \mathbb{N}$, i.e., there exists an absolute constant $\mathbf{K} \in [0, \infty)$ such that for all

$$T \in \left\{ \operatorname{diam}(D), \sup_{(u,z) \in [0,\infty) \times \overline{D}} |(\mathcal{R}(\mathfrak{h}))(u,z)|, L_1, L_2, \mathcal{P}(\mathfrak{h}), \mathbf{C} \right\} \quad (6.37)$$

it holds that $T \leq \mathbf{K}d^{\mathbf{K}}$, the overall number of parameters of the approximating ANNs grows at most polynomially in the dimension d and the reciprocal ε^{-1} of the approximation accuracy ε . In this sense, the ANN approximation presented in Theorem 6.2.1 overcomes the curse of dimensionality.

- (ii) The assumption that h can be represented directly as an ANN can be weakened to the assumption that h can be approximated uniformly on $[0, \infty) \times \overline{D}$, where the size of these approximating ANNs appears in the dependency of the constant c in the conclusion of Theorem 6.2.1.
- (iii) Note that Lemmas 6.1.2 to 6.1.4 can be extended to include nonzero h_1 under suitable assumptions as imposed on h (i.e., $\mathfrak{h}$) in the approximation of the Monte Carlo scheme (5.5). The statement and proof of Theorem 6.2.1 can be extended to this scenario as well.

Chapter 7.

Conclusion and Outlook

This thesis investigated two classes of high-dimensional approximation problems, one arising from MDPs and the other from elliptic PDEs. For both problems, we conducted the common strategy of establishing problem specific Monte Carlo approximations which are subsequently utilized as structural bases for ANN approximations.

We developed the MLFP approximation methods for functional fixed-point equations that are motivated by Bellman equations of MDPs and under suitable assumptions, these methods achieve prescribed approximation accuracies with computational complexities that grow at most polynomially in the relevant state space dimension and in the reciprocal of the accuracy. Hence, in these situations they overcome the curse of dimensionality. Subsequently, we combined the resulting Monte Carlo methods with ANN representation and approximation to establish ANN approximations, for which the size of the approximating ANNs grows at most polynomially in the dimension of the underlying space as well as in the reciprocal of the prescribed approximation accuracy.

For the approximation problem motivated by linear elliptic PDEs with drift and killing, we developed Walk-on-Spheres Monte Carlo methods for stochastic representations of such PDEs together with error estimates in the L^∞ -sense and quantitative bounds on the complexity parameters depending at most polynomially on relevant problem features. In the case in which the integral in the stochastic representation vanishes, the resulting Monte Carlo methods were transferred to ANN approximations with a corresponding analysis of the network complexity of the approximating ANNs.

The results of this thesis lead naturally to several possible research directions. For the first problem class, these are for instance the consideration of MDPs with more general action spaces to relax the assumption of a common and finite action set. Furthermore, it would be interesting to extend and strengthen the error analysis of the MLFP scheme to uniform bounds or estimates for the tail probabilities of the error. Another direction is to extract strategies from the MLFP approximations or the corresponding ANN approximations to obtain approximations of optimal strategies for control problems.

For the second problem class, a natural extension is to consider elliptic PDEs with state-dependent coefficients. Another direction, which is motivated by the inherent time dependency in the expectations of the form (1.5), which appeared as a result of the drift and killing compensation (cf. Proposition 5.1.2 and Remark 5.1.5), is to extend the Walk-on-Spheres methodology to time-dependent problems, in particular, parabolic PDEs.

Bibliography

- [1] Julia Ackermann, Arnulf Jentzen, Thomas Kruse, Benno Kuckuck, and Joshua Lee Padgett. Deep neural networks with ReLU, leaky ReLU, and softplus activation provably overcome the curse of dimensionality for Kolmogorov partial differential equations with Lipschitz nonlinearities in the l^p -sense. *arXiv:2309.13722*, 2023.
- [2] Julia Ackermann, Arnulf Jentzen, Benno Kuckuck, and Joshua Lee Padgett. Deep neural networks with ReLU, leaky ReLU, and softplus activation provably overcome the curse of dimensionality for space-time solutions of semilinear partial differential equations. *arXiv:2406.10876*, 2024.
- [3] Alekh Agarwal, Sham Kakade, and Lin F Yang. Model-based reinforcement learning with a generative model is minimax optimal. In *Conference on Learning Theory*, pages 67–83. PMLR, 2020.
- [4] Leif Andersen and Mark Broadie. Primal-dual simulation algorithm for pricing multidimensional American options. *Management Science*, 50(9):1222–1234, 2004.
- [5] Mohammad Gheshlaghi Azar, Rémi Munos, and Hilbert J Kappen. Minimax PAC bounds on the sample complexity of reinforcement learning with a generative model. *Mach. Learn.*, 91:325–349, 2013.
- [6] Christian Beck, Lukas Gonon, and Arnulf Jentzen. Overcoming the curse of dimensionality in the numerical approximation of high-dimensional semilinear elliptic partial differential equations. *Partial Differ. Equ. Appl.*, 5(31), 2024.
- [7] Christian Beck, Fabian Hornung, Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse. Overcoming the curse of dimensionality in the numerical approximation of Allen-Cahn partial differential equations via truncated full-history recursive multilevel Picard approximations. *J. Numer. Math.*, 28(4):197–222, 2020.
- [8] Christian Beck, Martin Hutzenthaler, Arnulf Jentzen, and Benno Kuckuck. An overview on deep learning-based approximation methods for partial differential equations. *Discrete Contin. Dyn. Syst. Ser. B*, 28(6):3697–3746, 2023.
- [9] Christian Beck, Arnulf Jentzen, Konrad Kleinberg, and Thomas Kruse. Nonlinear Monte Carlo methods with polynomial runtime for Bellman equations of discrete time high-dimensional stochastic optimal control problems. *Appl. Math. Optim.*, 91(1):1–42, 2025.
- [10] Christian Beck, Arnulf Jentzen, and Thomas Kruse. Nonlinear Monte Carlo methods with polynomial runtime for high-dimensional iterated nested expectations. *arXiv:2009.13989*, 2020.
- [11] Sebastian Becker, Ramon Braunwarth, Martin Hutzenthaler, Arnulf Jentzen, and Philippe von Wurstemberger. Numerical simulations for full history recursive multilevel Picard approximations for systems of high-dimensional partial differential equations. *Commun. Comput. Phys.*, 28(5):2109–2138, 2020.
- [12] Sebastian Becker, Patrick Cheridito, and Arnulf Jentzen. Deep optimal stopping. *J. Mach. Learn. Res.*, 20:74, 2019.

- [13] Sebastian Becker, Patrick Cheridito, Arnulf Jentzen, and Timo Welti. Solving high-dimensional optimal stopping problems using deep learning. *European J. Appl. Math.*, 32(3):470–514, 2021.
- [14] Richard Bellman. *Dynamic programming*. Princeton Landmarks in Mathematics. Princeton University Press, Princeton, NJ, 2010. Reprint of the 1957 edition, With a new introduction by Stuart Dreyfus.
- [15] Denis Belomestny, Christian Bender, and John Schoenmakers. True upper bounds for Bermudan products via non-nested Monte Carlo. *Math. Finance*, 19(1):53–71, 2009.
- [16] Julius Berner, Philipp Grohs, and Arnulf Jentzen. Analysis of the generalization error: empirical risk minimization over deep artificial neural networks overcomes the curse of dimensionality in the numerical approximation of Black-Scholes partial differential equations. *SIAM J. Math. Data Sci.*, 2(3):631–657, 2020.
- [17] Dimitri Bertsekas. *Reinforcement learning and optimal control*. Athena Scientific, 2019.
- [18] Dimitri Bertsekas and Steven E Shreve. *Stochastic optimal control: the discrete-time case*, volume 5. Athena Scientific, 1996.
- [19] Dimitri Bertsekas and John N Tsitsiklis. *Neuro-dynamic programming*. Athena Scientific, 1996.
- [20] Lucian Beznea, Iulian Cimpian, Oana Lupascu-Stamate, Ionel Popescu, and Arghir Zarnescu. From Monte Carlo to neural networks approximations of boundary value problems. *arXiv:2209.01432*, 2022.
- [21] Thomas E. Booth. Exact Monte Carlo solution of elliptic partial differential equations. *J. Comput. Phys.*, 39(2):396–404, 1981.
- [22] Robert L Bray. A comment on “Using randomization to break the curse of dimensionality”. *Econometrica*, 90(4):1915–1929, 2022.
- [23] Mark Broadie and Paul Glasserman. Pricing American-style securities using simulation. *J. Econom. Dynam. Control*, 21(8-9):1323–1352, 1997.
- [24] Mark Broadie and Paul Glasserman. A stochastic mesh method for pricing high-dimensional American options. *J. Comput. Finance*, 7:35–72, 2004.
- [25] Qi Cai, Zhuoran Yang, Jason D Lee, and Zhaoran Wang. Neural temporal-difference learning converges to global optima. *Advances in Neural Information Processing Systems*, 32, 2019.
- [26] Chef-Seng Chow and John N Tsitsiklis. The complexity of dynamic programming. *J. Complexity*, 5(4):466–488, 1989.
- [27] Chef-Seng Chow and John N Tsitsiklis. An optimal one-way multigrid algorithm for discrete-time stochastic control. *IEEE Trans. Automat. Control*, 36(8):898–914, 1991.
- [28] Petru Cioica-Licht, Martin Hutzenthaler, and P Tobias Werner. Deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear partial differential equations. *arXiv:2205.14398*, 2022.
- [29] G. Cybenko. Approximation by superpositions of a sigmoidal function. *Math. Control Signals Systems*, 2(4):303–314, 1989.

- [30] Vijay V Desai, Vivek F Farias, and Ciamac C Moallemi. Pathwise optimization for optimal stopping problems. *Management Science*, 58(12):2292–2308, 2012.
- [31] Weinan E, Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse. On multilevel Picard numerical approximations for high-dimensional nonlinear parabolic partial differential equations and high-dimensional nonlinear backward stochastic differential equations. *J. Sci. Comput.*, 79(3):1534–1571, 2019.
- [32] Weinan E, Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse. Multilevel Picard iterations for solving smooth semilinear parabolic heat equations. *Partial Differ. Equ. Appl.*, 2(6):80, 2021.
- [33] Jianqing Fan, Zhaoran Wang, Yuchen Xie, and Zhuoran Yang. A theoretical analysis of deep Q-learning. In *Learning for dynamics and control*, pages 486–489. PMLR, 2020.
- [34] Avner Friedman. *Stochastic differential equations and applications. Vol. 1*, volume Vol. 28 of *Probability and Mathematical Statistics*. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1975.
- [35] David Gilbarg and Neil S. Trudinger. *Elliptic partial differential equations of second order*. Classics in Mathematics. Springer-Verlag, Berlin, Heidelberg, 2001.
- [36] Michael B. Giles, Arnulf Jentzen, and Timo Welti. Generalised multilevel Picard approximations. *arXiv:1911.03188*, 2019.
- [37] Mike Giles. Improved multilevel Monte Carlo convergence using the Milstein scheme. In *Monte Carlo and Quasi-Monte Carlo Methods 2006*, pages 343–358. Springer, 2008.
- [38] Lukas Gonon. Deep neural network expressivity for optimal stopping problems. *arXiv:2210.10443*, 2022.
- [39] Philipp Grohs and Lukas Herrmann. Deep neural network approximation for high-dimensional elliptic PDEs with boundary conditions. *IMA J. Numer. Anal.*, 42(3):2055–2082, 2022.
- [40] Philipp Grohs, Fabian Hornung, Arnulf Jentzen, and Philippe Von Wurstemberger. A proof that artificial neural networks overcome the curse of dimensionality in the numerical approximation of Black–Scholes partial differential equations. *Mem. Amer. Math. Soc.*, 284(1410), 2023.
- [41] Philipp Grohs, Fabian Hornung, Arnulf Jentzen, and Philipp Zimmermann. Space-time error estimates for deep neural network approximations for differential equations. *Adv. Comput. Math.*, 49(1):4, 2023.
- [42] Philipp Grohs, Arnulf Jentzen, and Diyora Salimova. Deep neural network approximations for solutions of PDEs based on Monte Carlo algorithms. *Partial Differ. Equ. Appl.*, 3(4):45, 2022.
- [43] Lars Grüne, Konrad Kleinberg, Thomas Kruse, and Mario Sperl. Convexity and strict convexity for compositional neural networks in high-dimensional optimal control. *arXiv:2511.05339*, 2025.
- [44] Martin B Haugh and Leonid Kogan. Pricing American options: a duality approach. *Oper. Res.*, 52(2):258–270, 2004.
- [45] Stefan Heinrich. Monte Carlo complexity of global solution of integral equations. *J. Complexity*, 14(2):151–175, 1998.

- [46] Stefan Heinrich. Multilevel Monte Carlo methods. In *International Conference on Large-Scale Scientific Computing*, pages 58–67. Springer, 2001.
- [47] Kurt Hornik, Maxwell Stinchcombe, and Halbert White. Multilayer feedforward networks are universal approximators. *Neural networks*, 2(5):359–366, 1989.
- [48] Martin Hutzenthaler, Arnulf Jentzen, and Thomas Kruse. Overcoming the curse of dimensionality in the numerical approximation of parabolic partial differential equations with gradient-dependent nonlinearities. *Found. Comput. Math.*, pages 1–62, 2021.
- [49] Martin Hutzenthaler, Arnulf Jentzen, Thomas Kruse, and Tuan Anh Nguyen. Multilevel Picard approximations for high-dimensional semilinear second-order PDEs with Lipschitz nonlinearities. *arXiv:2009.02484*, 2020.
- [50] Martin Hutzenthaler, Arnulf Jentzen, Thomas Kruse, and Tuan Anh Nguyen. A proof that rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of semilinear heat equations. *Partial Differ. Equ. Appl.*, 1:1–34, 2020.
- [51] Martin Hutzenthaler, Arnulf Jentzen, Thomas Kruse, Tuan Anh Nguyen, and Philippe von Wurstemberger. Overcoming the curse of dimensionality in the numerical approximation of semilinear parabolic partial differential equations. *Proc. of the Royal Soc. A.*, 476(2244):20190630, 2020.
- [52] Martin Hutzenthaler, Arnulf Jentzen, Benno Kuckuck, and Joshua Lee Padgett. Strong l^p -error analysis of nonlinear Monte Carlo approximations for high-dimensional semilinear partial differential equations. *arXiv:2110.08297*, 2021.
- [53] Martin Hutzenthaler, Arnulf Jentzen, and Philippe von Wurstemberger. Overcoming the curse of dimensionality in the approximative pricing of financial derivatives with default risks. *Electron. J. Probab.*, 25:Paper No. 101, 73, 2020.
- [54] Martin Hutzenthaler and Thomas Kruse. Multilevel Picard approximations of high-dimensional semilinear parabolic differential equations with gradient-dependent nonlinearities. *SIAM J. Numer. Anal.*, 58(2):929–961, 2020.
- [55] Martin Hutzenthaler, Thomas Kruse, and Tuan Anh Nguyen. Multilevel Picard approximations for McKean-Vlasov stochastic differential equations. *J. Math. Anal. Appl.*, 507(1):125761, 2022.
- [56] Arnulf Jentzen, Konrad Kleinberg, and Thomas Kruse. Deep neural networks can provably solve Bellman equations for Markov decision processes without the curse of dimensionality. *arXiv:2506.22851*, 2025.
- [57] Arnulf Jentzen, Benno Kuckuck, and Philippe von Wurstemberger. Mathematical introduction to deep learning: Methods, implementations, and theory. *arXiv:2310.20360*, 2023.
- [58] Arnulf Jentzen, Diyora Salimova, and Timo Welti. A proof that deep artificial neural networks overcome the curse of dimensionality in the numerical approximation of Kolmogorov partial differential equations with constant diffusion and nonlinear drift coefficients. *Commun. Math. Sci.*, 19(5):1167–1205, 2021.
- [59] Sham Kakade. *On the sample complexity of reinforcement learning*. University of London, University College London (United Kingdom), 2003.
- [60] Wei Kang and Qi Gong. Feedforward neural networks and compositional functions with applications to dynamical systems. *SIAM J. Control Optim.*, 60(2):786–813, 2022.

- [61] Ioannis Karatzas and Steven E. Shreve. *Brownian motion and stochastic calculus*, volume 113 of *Graduate Texts in Mathematics*. Springer-Verlag, New York, second edition, 1991.
- [62] Elia Kaufmann, Leonard Bauersfeld, Antonio Loquercio, Matthias Müller, Vladlen Koltun, and Davide Scaramuzza. Champion-level drone racing using deep reinforcement learning. *Nature*, 620(7976):982–987, 2023.
- [63] Michael Kearns, Yishay Mansour, and Andrew Y Ng. A sparse sampling algorithm for near-optimal planning in large Markov decision processes. *Mach. Learn.*, 49:193–208, 2002.
- [64] Konrad Kleinberg and Thomas Kruse. Walk-on-spheres Monte Carlo and deep neural network approximations of elliptic PDEs with drift and killing. *arXiv:2608.09494*, 2026.
- [65] Achim Klenke. *Probability theory*. Universitext. Springer, London, second edition, 2014.
- [66] Dennis Kristensen, Patrick K Mogensen, Jong Myun Moon, and Bertel Schjerning. Solving dynamic discrete choice models using smoothing and sieve methods. *J. Econometrics*, 223(2):328–360, 2021.
- [67] Andreas E. Kyprianou, Ana Osojnik, and Tony Shardlow. Unbiased ‘walk-on-spheres’ Monte Carlo methods for the fractional Laplacian. *IMA J. Numer. Anal.*, 38(3):1550–1578, 2018.
- [68] Timothy P Lillicrap, Jonathan J Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning. *4th International Conference on Learning Representations, ICLR 2016*, 2016.
- [69] Francis A Longstaff and Eduardo S Schwartz. Valuing American options by simulation: a simple least-squares approach. *Rev. Financ. Stud.*, 14(1):113–147, 2001.
- [70] Pertti Mattila. *Geometry of sets and measures in Euclidean spaces*, volume 44 of *Cambridge Studies in Advanced Mathematics*. Cambridge University Press, Cambridge, 1995. Fractals and rectifiability.
- [71] Matthieu Meunier, Christoph Reisinger, and Yufei Zhang. Efficient learning for entropy-regularized Markov decision processes via multilevel Monte Carlo. *arXiv:2503.21224*, 2025.
- [72] H. N. Mhaskar. Neural networks for optimal approximation of smooth and analytic functions. *Neural Computation*, 8(1):164–177, 01 1996.
- [73] Volodymyr Mnih, Koray Kavukcuoglu, David Silver, Andrei A Rusu, Joel Veness, Marc G Bellemare, Alex Graves, Martin Riedmiller, Andreas K Fidjeland, Georg Ostrovski, et al. Human-level control through deep reinforcement learning. *Nature*, 518(7540):529–533, 2015.
- [74] Mervin E Muller. Some continuous Monte Carlo methods for the Dirichlet problem. *The Annals of Mathematical Statistics*, pages 569–589, 1956.
- [75] Thomas Müller-Gronbach, Erich Novak, and Klaus Ritter. *Monte Carlo-Algorithmen*. Springer-Verlag, 2012.
- [76] Ariel Neufeld and Tuan Anh Nguyen. Rectified deep neural networks overcome the curse of dimensionality in the numerical approximation of gradient-dependent semilinear heat equations. *Commun. Math. Sci.*, 23(4):883–912, 2025.

- [77] Ariel Neufeld and Tuan Anh Nguyen. Multilevel Picard approximations and deep neural networks with ReLU, leaky ReLU, and softplus activation overcome the curse of dimensionality when approximating semilinear parabolic partial differential equations in L^p -sense. *J. Comput. Appl. Math.*, 487:Paper No. 117736, 32, 2026.
- [78] Ariel Neufeld, Tuan Anh Nguyen, and Sizhou Wu. Deep ReLU neural networks overcome the curse of dimensionality when approximating semilinear partial integro-differential equations. *Anal. Appl. (Singap.)*, 23(7):1127–1178, 2025.
- [79] Ariel Neufeld, Tuan Anh Nguyen, and Sizhou Wu. Multilevel Picard approximations overcome the curse of dimensionality in the numerical approximation of general semilinear PDEs with gradient-dependent nonlinearities. *J. Complexity*, page 101946, 2025.
- [80] Ariel Neufeld and Sizhou Wu. Multilevel Picard approximation algorithm for semilinear partial integro-differential equations and its complexity analysis. *Stoch. Partial Differ. Equ. Anal. Comput.*, pages 1–59, 2025.
- [81] Erich Novak and Klaus Ritter. The curse of dimension and a universal method for numerical integration. In *Multivariate approximation and splines (Mannheim, 1996)*, volume 125 of *Internat. Ser. Numer. Math.*, pages 177–187. Birkhäuser, Basel, 1997.
- [82] Erich Novak and Henryk Woźniakowski. *Tractability of multivariate problems. Vol. 1: Linear information*, volume 6 of *EMS Tracts in Mathematics*. European Mathematical Society (EMS), Zürich, 2008.
- [83] Sidney C. Port and Charles J. Stone. *Brownian motion and classical potential theory*. Probability and Mathematical Statistics. Academic Press [Harcourt Brace Jovanovich, Publishers], New York-London, 1978.
- [84] Warren B Powell. *Approximate Dynamic Programming: Solving the curses of dimensionality*, volume 703. John Wiley & Sons, 2007.
- [85] Martin L Puterman. *Markov decision processes: discrete stochastic dynamic programming*. John Wiley & Sons, 2014.
- [86] Christoph Reisinger and Yufei Zhang. Rectified deep neural networks overcome the curse of dimensionality for nonsmooth value functions in zero-sum games of nonlinear stiff systems. *Anal. Appl. (Singap.)*, 18(06):951–999, 2020.
- [87] Leonard C. G. Rogers. Monte Carlo valuation of American options. *Math. Finance*, 12(3):271–286, 2002.
- [88] Gavin A. Rummery and Mahesan Niranjana. *On-line Q-learning using connectionist systems*, volume 37. University of Cambridge, Department of Engineering Cambridge, UK, 1994.
- [89] John Rust. A comparison of policy iteration methods for solving continuous-state, infinite-horizon Markovian decision problems using random, quasi-random, and deterministic discretizations. Available at SSRN: <http://dx.doi.org/10.2139/ssrn.37768>, 1997.
- [90] John Rust. Using randomization to break the curse of dimensionality. *Econometrica*, 65(3):487–516, 1997.
- [91] Karl K. Sabelfeld. Random walk on spheres method for solving drift-diffusion problems. *Monte Carlo Methods Appl.*, 22(4):265–275, 2016.

- [92] Karl K. Sabelfeld and Anastasya Kireeva. A new global random walk algorithm for calculation of the solution and its derivatives of elliptic equations with constant coefficients in an arbitrary set of points. *Appl. Math. Lett.*, 107:106466, 9, 2020.
- [93] Karl K. Sabelfeld and Anastasya Kireeva. A global random walk on spheres algorithm for calculating the solution and its derivatives of the drift-diffusion-reaction equations. *Math. Methods Appl. Sci.*, 45(3):1420–1431, 2022.
- [94] Karl K. Sabelfeld and Denis Talay. Integral formulation of the boundary value problems and the method of random walk on spheres. *Monte Carlo Methods Appl.*, 1(1):1–34, 1995.
- [95] Rohan Sawhney, Dario Seyb, Wojciech Jarosz, and Keenan Crane. Grid-free Monte Carlo for PDEs with spatially varying coefficients. *ACM Trans. Graph.*, 41(4), July 2022.
- [96] Aaron Sidford, Mengdi Wang, Xian Wu, Lin Yang, and Yinyu Ye. Near-optimal time and sample complexities for solving Markov decision processes with a generative model. *Advances in Neural Information Processing Systems*, 31, 2018.
- [97] Aaron Sidford, Mengdi Wang, Xian Wu, and Yinyu Ye. Variance reduced value iteration and faster algorithms for solving Markov decision processes. *Naval Res. Logist.*, 70(5):423–442, 2023.
- [98] David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of Go with deep neural networks and tree search. *Nature*, 529(7587):484–489, 2016.
- [99] Richard S Sutton. Learning to predict by the methods of temporal differences. *Mach. Learn.*, 3:9–44, 1988.
- [100] Richard S Sutton and Andrew G Barto. *Reinforcement learning: An introduction*. MIT press, 2018.
- [101] John N Tsitsiklis and Benjamin Van Roy. Optimal stopping of Markov processes: Hilbert space theory, approximation algorithms, and an application to pricing high-dimensional financial derivatives. *IEEE Trans. Automat. Control*, 44(10):1840–1851, 1999.
- [102] John N Tsitsiklis and Benjamin Van Roy. Regression methods for pricing complex American-style options. *IEEE Trans. Neural Networks*, 12(4):694–703, 2001.
- [103] Christopher Watkins. Learning from delayed rewards. *Ph. D. thesis. University of Cambridge*, 1989.
- [104] Christopher Watkins and Peter Dayan. Q-learning. *Mach. Learn.*, 8(3-4):279–292, 1992.
- [105] Pan Xu and Quanquan Gu. A finite-time analysis of Q-learning with neural network function approximation. In *International conference on machine learning*, pages 10555–10565. PMLR, 2020.
- [106] Dmitry Yarotsky. Error bounds for approximations with deep relu networks. *Neural Networks*, 94:103–114, 2017.