



Article

Deep Learning and Mean-Field Games: A Stochastic Optimal Control Perspective

Luca Di Persio ^{1,*}  and Matteo Garbelli ² ¹ Department of Computer Science, University of Verona, 37134 Verona, Italy;² Department of Mathematics, University of Trento, 38123 Trento, Italy; matteo.garbelli@unitn.it* Correspondence: luca.dipersio@univr.it

Abstract: We provide a rigorous mathematical formulation of Deep Learning (DL) methodologies through an in-depth analysis of the learning procedures characterizing Neural Network (NN) models within the theoretical frameworks of Stochastic Optimal Control (SOC) and Mean-Field Games (MFGs). In particular, we show how the supervised learning approach can be translated in terms of a (stochastic) mean-field optimal control problem by applying the Hamilton–Jacobi–Bellman (HJB) approach and the mean-field Pontryagin maximum principle. Our contribution sheds new light on a possible theoretical connection between mean-field problems and DL, melting heterogeneous approaches and reporting the state-of-the-art within such fields to show how the latter different perspectives can be indeed fruitfully unified.

Keywords: deep learning; neural networks; stochastic optimal control; mean-field games; Hamilton–Jacobi–Bellman equation; Pontryagin maximum principle



Citation: Di Persio, L.; Garbelli, M. Deep Learning and Mean-Field Games: A Stochastic Optimal Control Perspective. *Symmetry* **2021**, *13*, 14. <http://dx.doi.org/10.3390/sym13010014>

Received: 18 November 2020

Accepted: 17 December 2020

Published: 23 December 2020

Publisher’s Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2020 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

1. Introduction

Controlled stochastic processes, which naturally arise in a plethora of heterogeneous fields, spanning, e.g., from mathematical finance to industry, can be solved in the setting of continuous time stochastic control theory. In particular, when we have to analyse complex dynamics produced by the mutual interaction of a large set of indistinguishable players, an efficient approach to infer knowledge about the resulting behaviour, typical for example of a neuronal ensemble, is provided by Mean-Field Game (MFG) methods, as described in [1]. MFG theory generalizes classical models of interacting particle systems characterizing statistical mechanics. Intuitively, each particle is replaced by rational agents whose dynamics are represented by a Stochastic Differential Equation (SDE). The term mean-field refers to the highly symmetric form of interaction: the dynamics and the objective of each particle depend on an empirical measure capturing the global behaviour of the population. The solution of an MFG is analogous to a Nash equilibrium for a non-cooperative game [2]. The key idea is that the population limit can be effectively approximated by statistical features of the system corresponding to the behaviour of a typical group of agents, in a Wasserstein space sense [3]. On the other hand, Deep Learning (DL) is frequently used in several Machine Learning (ML) based applications, spanning from image classification and speech recognition to predictive maintenance and clustering. Therefore, it has become essential to provide a strong mathematical formulation and to analyse both the setting and the associated algorithms [4,5]. Commonly, Neural Networks (NNs) are trained through the Stochastic Gradient Descent (SGD) method. It updates the trainable parameters using gradient information computed randomly via a back-propagation algorithm with the disadvantage of being slow in the first steps of training. An alternative consists of expressing the learning procedure of an NN as a dynamical system (see [6]), which can be then analysed as an optimal control problem [7].

The present paper is structured as follows. In Section 2, we introduce the fundamentals about the Wasserstein space, Stochastic Optimal Control (SOC) and MFGs. In Section 3,

the link between NNs and MFGs is deeply analysed in order to reflect the probabilistic nature of the learning process. Conclusions and prospects are outlined in Section 4.

2. Problem Formulation and Preliminaries

2.1. Wasserstein Metrics

Many of the main results contained in this paper will be stated in terms of convergence in the distribution of random variables, vectors, processes and measures. We refer to [8–10] concerning the basics about the theory of the weak convergence of probability measures in metric spaces equipped with the natural Borel σ -algebra; see also [11,12] and the references therein for further details. According to [8] (pp. 7, 8), let us recall the following two definitions, useful to highlight the strict connection between measure theory and probability. Given a probability measure $\mu \in \mathcal{P}(\mathcal{X})$ over the metric space \mathcal{X} and a sequence $(\mu_n) \subset \mathcal{P}(\mathcal{X})$ with $n \in \mathbb{N}$, we say that μ_n converges weakly to μ or $\mu_n \rightarrow \mu$ if:

$$\lim_{n \rightarrow \infty} \int_{\mathcal{X}} f d\mu_n = \int_{\mathcal{X}} f d\mu, \forall f \in \mathcal{C}_b(\mathcal{X}).$$

Given a sequence of \mathcal{X} -valued random variables $\{X_n\}_{n \geq 1}$, we say that X_n converges weakly (or in distribution) to a \mathcal{X} -valued random variable X , if:

$$\lim_{n \rightarrow \infty} \mathbb{E}[f(X_n)] = \mathbb{E}[f(X)], \forall f \in \mathcal{C}_b(\mathcal{X}),$$

denoting this convergence by $X_n \Rightarrow X$. Focusing on the convergence of empirical measures, let (X_i) be a sequence of independent and identically distributed (i.i.d.) \mathcal{X} -random variables and define the $\mathcal{P}(\mathcal{X})$ -valued random variable:

$$\mu_n = \frac{1}{n} \sum_{i=1}^n \delta_{X_i}, \quad (1)$$

then we have a random probability measure, usually indicated as the empirical measure. According to [8] (pp. 12–16), $\mathcal{P}(\mathcal{X})$ is endowed with a metric (compatible with the notion of weak convergence) in order to consider $\mathcal{P}(\mathcal{X})$ as a metric space itself. Let us recall the Wasserstein metric, defined on $\mathcal{P}(\mathcal{X})$, based on the idea of coupling. In particular, given $\mu, \nu \in \mathcal{P}(\mathcal{X})$, $\Pi(\mu, \nu)$ represents the set of Borel probability measures π on $\mathcal{X} \times \mathcal{X}$, with first, resp. second, marginal μ , resp. ν . Namely, $\pi(A \times \mathcal{X}) = \mu(A)$ and $\pi(\mathcal{X} \times A) = \nu(A)$ for every Borel set $A \subset \mathcal{X}$. Then, we define $\mathcal{P}^p(\mathcal{X})$ for $p \geq 1$, as the set of probability measures $\mu \in \mathcal{P}(\mathcal{X})$ satisfying:

$$\int_{\mathcal{X}} d(x, x_0)^p \mu(dx) < \infty,$$

where $x_0 \in \mathcal{X}$ is an arbitrary reference point. Consequently, the p -Wasserstein metric on $\mathcal{P}^p(\mathcal{X})$ is defined as:

$$\mathcal{W}_p(\mu, \nu) = \left(\inf_{\pi \in \Pi(\mu, \nu)} \int_{\mathcal{X} \times \mathcal{X}} d(x, y)^p \pi(dx, dy) \right)^{1/p} = \left(\inf_{X \sim \mu, Y \sim \nu} \mathbb{E}[d(X, Y)^p] \right)^{1/p}, \quad (2)$$

where the infimum is taken over all pairs of \mathcal{X} -valued random variables X, Y with, respectively, given marginals μ and ν .

2.2. Stochastic Optimal Control Problem

Following [13] (see also [14] and the references therein), let $\{\Omega, \mathcal{F}, (\mathcal{F}_t), \mathbb{P}\}$ be a filtered probability space, with filtration $\mathcal{F}_t = \{\mathcal{F}_t, t \in [0, T]\}$, $T > 0$, supporting:

- a controlled state variable $(X_t^\alpha)_{t \in [0, T]}$, where X_t is an i.i.d. sequence of \mathbb{R}^n -valued \mathcal{F}_0 -measurable random variables;
- a sequence $\{W_t^i\}_{i \geq 1}$ of independent and \mathcal{F}_t -adapted Brownian motions.

We consider a control as enforced on the state variable by $\alpha = (\alpha_t)_{t \in [0, T]}$, that is an adapted process defined on $\mathcal{A} = \{(\alpha_t)_{t \in [0, T]}\}$, containing \mathcal{F} -progressively measurable controls, which take values in A identified as a closed subset of \mathbb{R}^d . The objective function J is then defined as:

$$J(\alpha, X) = \mathbb{E} \left[\int_0^T f(\alpha_s, X_s) ds + g(X_T) \right],$$

where $f : A \times \mathbb{R}^n \rightarrow \mathbb{R}$ denotes the running objective function, while $g : \mathbb{R}^n \rightarrow \mathbb{R}$ denotes the terminal objective function, and the expectation is taken with respect to (w.r.t.) the probability measure \mathbb{P} .

The goal corresponds to identifying the control α solving the maximization problem:

$$\max_{\alpha \in \mathcal{A}} J(\alpha, X_t),$$

according to the n -dimensional stochastic controlled process X_s s.t.

$$\begin{cases} dX_s = b(\alpha_s, X_s) ds + \sigma(\alpha_s, X_s) dW_s \\ X_t = x, s \in [t, T]. \end{cases} \tag{3}$$

The drift function $b : A \times \mathbb{R}^n \rightarrow \mathbb{R}^n$ and the volatility function $\sigma : A \times \mathbb{R}^n \rightarrow \mathbb{R}^{d+n}$ are measurable, and they satisfy a uniform Lipschitz condition in x , i.e., there exists $K > 0$ such that, for all $x, y \in \mathbb{R}^n$ and $\alpha_s \in A$, it holds that:

$$|b(x, \alpha) - b(y, \alpha)| + |\sigma(x, \alpha) - \sigma(y, \alpha)| \leq K|x - y|.$$

Previous assumptions guarantee that the SDE (3) has a unique solution, denoted by $(X_s^{t,x})_{s \in [t, T]}$. Therefore, the objective function can be explicitly expressed in terms of x, t , namely:

$$J(\alpha, x, t) = \mathbb{E} \left[\int_0^T f(\alpha_s, X_s^{t,x}) ds + g(X_T^{t,x}) \right],$$

with $(t, x) \in [0, T] \times \mathbb{R}^n$, and the process $\alpha \in \mathcal{A}$ takes values $\alpha_s \in A$. Let v be the value function:

$$v(t, x) = \sup_{\alpha \in \mathcal{A}} J(t, x, \alpha), \tag{4}$$

then the corresponding optimal control $\hat{\alpha} : [0, T] \times \mathbb{R}^n \rightarrow \mathcal{A}$ is defined by:

$$J(t, x, \hat{\alpha}) = v(t, x), \quad \forall (t, x) \in [0, T] \times \mathbb{R}^n,$$

whose solution can be found by exploiting two different and interconnected approaches respectively based on the Hamilton–Jacobi–Bellman (HJB) equation and on the stochastic Pontryagin Maximum Principle (PMP).

The first one moves from the Dynamic Programming Principle (DPP), then leading to the nonlinear second order Partial Differential Equation (PDE) known as the HJB equation (see [3]), namely:

$$\begin{cases} \frac{\partial v}{\partial t}(t, x) + \sup_{a \in A} [\mathcal{L}^a v(t, x) + f(x, a)] = 0 \\ v(T, x) = g(x), \end{cases} \tag{5}$$

that holds $\forall (x, t) \in [0, T] \times \mathbb{R}^n$, where \mathcal{L} defines a second order operator called the infinitesimal generator of the controlled diffusion:

$$\mathcal{L}^a v = b(x, a) \cdot \nabla_x v(x) + \frac{1}{2} \text{Tr} [(\sigma(x, a)\sigma(x, a))^T \nabla_x^2 v(x)].$$

It is possible to define the Hamiltonian for the SOC problem [15,16] as $H : \mathbb{R}^d \times \mathbb{R}^d \times \mathbb{S}_d \rightarrow \mathbb{R} \cup \{\infty\}$, written as follows:

$$H(x, y, z) = \sup_{a \in A} \left(b(x, a) \cdot y + \frac{1}{2} \text{Tr}[\sigma \sigma^T(x, a)z] + f(x, a) \right),$$

where \mathbb{S}_d denotes the set of symmetric $d \times d$ matrices and the variables y and z are called the adjoint variables. The previously defined Hamiltonian allows encapsulating the nonlinearity of the HJB equation, which can be rearranged as:

$$\partial_t v(t, x) + H(x, \nabla v(t, x), \nabla^2 v(t, x)) = 0.$$

On the other hand, the stochastic PMP leads to a system of coupled Forward-Backward SDEs (FBSDEs) plus an external optimality condition in terms of the Hamiltonian function; see, e.g., [2,17] and the references therein. We define the local Hamiltonian as:

$$\mathcal{H}_t(x, u, p, q) = f_t(x, u) + b_t(x, u)p + \sigma_t(x)q, \quad (6)$$

where (p, q) are the adjoint variables.

By assuming that $f_t(x, u)$, $b_t(x, u)$ and $\sigma(x, u)$ are progressively measurable, bounded in $\mathcal{C}^1(x, u)$ and Lipschitz $\forall (x, u) \in \mathbb{R}^2$, we have that if an arbitrary control maximizes the Hamiltonian, then (necessary condition for optimality) it is the optimal one:

$$\mathcal{H}_\tau(\hat{x}_\tau, \hat{v}_\tau, \hat{P}_\tau, \hat{Q}_\tau) = \max_{v \in \mathbb{R}} \mathcal{H}_t(\hat{x}_\tau, v, \hat{P}_\tau, \hat{Q}_\tau).$$

Moreover, by requiring that $x \mapsto g(x)$ and $x \mapsto \hat{\mathcal{H}}_t(x, \hat{P}_\tau, \hat{Q}_\tau) = \sup_v \mathcal{H}_t(\hat{x}_\tau, v, \hat{P}_\tau, \hat{Q}_\tau)$ are both concave functions, the PMP becomes also a sufficient condition to characterize the optimal control. In addition, by the envelope theorem, it follows that the optimal control \hat{v} maximizing \mathcal{H} also maximizes its derivative $\partial_x \mathcal{H}$:

$$\partial_x \mathcal{H}_\tau(\hat{x}_\tau, \hat{v}_\tau, \hat{P}_\tau, \hat{Q}_\tau) = \partial_x \hat{\mathcal{H}}_t(\hat{x}_\tau, \hat{P}_\tau, \hat{Q}_\tau).$$

2.3. Mean-Field Games

In what follows, according to [1,3,5,18], we will exploit the theory of SOC to analyse Mean-Field Games (MFGs), to gain insights into the symmetric stochastic differential games characterized by a number of players that tends to infinity. In particular, consider n players $i = 1, \dots, n$ each of which controls a private state X^i whose dynamics read as follows:

$$dX_t^i = \frac{1}{N} \sum_{j=1}^N b(t, X_t^i, X_t^j, \alpha_t^i) + \sigma dW_t^i, \quad (7)$$

$\sigma \in \mathbb{R}$ being a (common) constant multiplying a random noise component, i.e., dW_t^i , steered by the i -th copy of a standard Brownian motion.

Each control α_t^i represents the strategy played by the i th player, while X_t^j models the influence of the state of the other j -players. The dynamics in Equation (7) can be reformulated as:

$$\frac{1}{N} \sum_{j=1}^N b(t, X_t^i, X_t^j, \alpha_t^i) = \int b(t, X_t^i, X_t^j, \alpha_t^i) \bar{\mu}_t^N dx,$$

where the empirical distribution $\bar{\mu}_t^N$ of private states is defined by:

$$\bar{\mu}_t^N = \frac{1}{N} \sum_{j=1}^N \delta_{X_t^j}.$$

Hence, the same dynamics for each player are obtained, since each player depends only on the global behaviour of the entire system and not on the state of the single player:

$$dX_t^i = \frac{1}{N} \sum_{j=1}^N b(t, X_t^i, \bar{\mu}_t^N, \alpha_t^i) + \sigma dW_t^i.$$

Each individual is minimizing a cost composed by a running cost and a terminal one. Therefore, the objective function $J^i(\alpha)$, $\forall i$, is defined by:

$$J^i(\alpha^i) = \mathbb{E} \left[\int_0^T f(t, X_t^i, \bar{\mu}_t^N, \alpha_t^i) dt + g(X_T^i, \bar{\mu}_T^N) \right],$$

the goal being to select ϵ -Nash equilibria, the vector $(\alpha^{1*}, \dots, \alpha^{N*})$ being a ϵ -Nash equilibrium if:

$$\forall i = 1, \dots, N \wedge \forall \alpha^i \quad J(\alpha^*) - \epsilon \leq J(\alpha^{*-i}, \alpha^i),$$

meaning that if the i th player is changing his/her behaviour for a small value ϵ , while other players follow a fixed strategy, there is no change in the profit of the game in terms of the cost J . Let us note that, if $\epsilon = 0$, we are dealing with the standard Nash equilibria.

In this scenario, the solution of an MFG (see [3]) corresponds to a couple (α^*, μ^*) , where $\mu_t^* = \mathcal{L}(X_t^{\alpha^*})$ models the optimal empirical distribution and $\alpha^* = \phi^*(t, X_t^i)$ the optimal strategies.

Due to the symmetry of the system, all the agents play the same strategy profile:

$$\phi^{1*}(t, X_t^1) = \dots = \phi^{N*}(t, X_t^N) = \phi^*(t), \quad \forall i = 1, \dots, N;$$

hence, by fixing the flow of the probability measure $(\mu_t)_{t \in [0, T]}$, the solution ϕ^* solves the SOC problem parametrized by the choice of the family (μ_t) as:

$$\phi^* = \underset{\phi^{(t)}_{t \in [0, T]}}{\operatorname{argmin}} \mathbb{E} \left[\int_0^T f(t, X_t, \mu_t, \phi(t, X_t)) dt + g(X_T, \mu_T) \right],$$

subject to:

$$dX_t = b(t, X_t, \mu_t, \phi(t, X_t)) dt + \sigma dW_t. \tag{8}$$

The solution of the latter optimization problem returns the best response of a representative player to the empirical measure (μ_t) in a scenario where no player is profitable in changing his/her own strategy.

Concerning the choice of the flow of the probability measure $(\mu_t)_{t \in [0, T]}$ (see Equation (1)), if $N \rightarrow \infty$, the asymptotic independence implied by the law of large numbers ensures that the empirical measure $\bar{\mu}_t^N$ tends to the statistical distribution of X_t in the form of a fixed point equation, that is $\mathcal{L}(X^{\alpha^\mu}) = \mu$. Once the optimal feedback ϕ^* has been found, for each choice of the parameter $(\mu_t)_{0 \leq t \leq T}$, it is necessary to check that the optimal empirical measure μ_t^* can be recovered from the statistical distribution of the optimal states $X_t^{\alpha^*}$, i.e., $\forall t, \mathcal{L}(X_t^{\alpha^*}) = \mu_t^*$ must hold.

By freezing the family μ of probability, the Hamiltonian becomes:

$$H^{\mu_t}(t, x, y, \alpha) = yb(t, x, \mu_t, \alpha) + f(t, x, \mu_t, \alpha). \tag{9}$$

Let us assume that there exists a regular function:

$$\hat{\alpha}(t, x, y) : [0, T] \times \mathbb{R} \times \mathbb{R} \rightarrow \underset{\alpha \in A}{\operatorname{argmin}} H^{\mu_t}(t, x, y, \alpha),$$

then we denote the infimum over controls α of $H^{\mu_t}(t, x, y, \alpha)$ by:

$$\mathcal{H}^{\mu_t}(t, x, y) := \inf_{\alpha \in A} H^{\mu_t}(t, x, y, \alpha),$$

ending up with a stochastic control problem that can be solved (see, e.g., [3] (pp. 7–9)) by applying one of the following two methods:

1. The PDE approach through HJB Equation (21) and the Kolmogorov equation;
2. The Backward Stochastic Differential Equation (BSDE) approach based on the PMP.

We introduce the HJB value equation following [3], (p. 7), i.e.:

$$v(t, x) = \inf_{\alpha \in A_t} \mathbb{E} \left[\int_t^T f(s, X_s, \mu_s, \alpha_s) ds + g(X_T, \mu_T) \mid X_t = x \right], \quad (10)$$

where A_t denotes the set of admissible controls over the interval $[t, T]$. It is expected that v is the solution, in the viscosity sense, of the HJB equation; see, e.g., [19,20] and the references therein. Hence:

$$\partial_t v + \frac{\sigma^2}{2} \partial_{xx}^2 v + \mathcal{H}^{\mu_t}(t, x, \partial_x v(t, x)) = 0, \quad (t, x) \in [0, T] \times \mathbb{R}, \quad (11)$$

with terminal condition $v(T, x) = g(x, \mu_T)$, $x \in \mathbb{R}$. Imposing $\bar{\mu} = (\mu_t)_{0 \leq t \leq T}$ as the flow of the optimally controlled state, the flow of statistical distributions should satisfy Kolmogorov's equation, so that, introducing the notation $\beta(t, x) = b(t, x, \mu_t, \phi(t, x))$ with ϕ the optimal feedback, the flow $(\nu_t)_{0 \leq t \leq T}$ of measures given by $\nu_t = \mathcal{L}(X_t)$ satisfies the Kolmogorov's equation:

$$\partial_t \nu - \frac{\sigma^2}{2} \partial_{xx}^2 \nu - \text{div}(\beta(t, x)\nu) = 0, \quad (t, x) \in [0, T] \times \mathbb{R}, \quad (12)$$

with initial condition $\nu_0 = \mu_0$. Such a PDE holds in the sense of distributions, since ν_t represents a density and the derivatives involved must be considered in the Wasserstein sense. Setting $\nu_t = \mathcal{L}(X_t) = \mu_t$, we end up with a system of coupled non-linear forward-backward PDEs (11) and (12), constituting the so-called MFG-PDE system. On the other hand, we can approximate the solution of the aforementioned stochastic control problem via the stochastic Pontryagin principle. In particular, for each open-loop adapted control $\alpha = (\alpha_t)_{0 \leq t \leq T}$, we denote by $X^\alpha = (X_t^\alpha)_{0 \leq t \leq T}$ the associated state, and we introduce the adjoint equation:

$$\begin{cases} dY_t = -\partial_x H^{\mu_t}(t, X_t, Y_t, \alpha_t) dt + Z_t dW_t, & t \in [0, T] \\ Y_T = \partial_x g(X_T, \mu_T). \end{cases} \quad (13)$$

H^{μ_t} corresponds to the Hamiltonian defined in Equation (9), and the solution $(Y_t, Z_t)_{0 \leq t \leq T}$ of the BSDE is called a set of adjoint processes.

The PMP necessary condition states that whenever X^α is an optimal state, it must hold that $H^{\mu_t}(t, X_t, Y_t, \alpha_t) = \mathcal{H}^{\mu_t}(t, X_t, Y_t)$, $t \in [0, T]$. Moreover, if the Hamiltonian H^{μ_t} is convex w.r.t. the variables (x, α) and the terminal g is convex w.r.t. the variable x , then the system given by Equations (13) and (8) characterizes the optimal states of the MFG problem.

3. Main Result

In this section, we generalize the approach proposed in [5], providing an application of the Mean-Field (MF) optimal control to Deep Learning (DL). In particular, we start by considering the learning process characterizing supervised learning as a population risk minimization problem, hence considering its probabilistic nature in the sense that the optimal control parameters, corresponding to the trainable weights in the associated Neural Network (NN) model, depending on the population distribution of input-target pairs constituting the randomness source.

3.1. Neural Network as a Dynamical System

In order to study DL as an optimal control problem, it is necessary to express the NN learning process as a dynamical system [6,21]. In the simplest form, the feed-forward propagation in a T layer, $T \geq 1$, network can be expressed by the following difference equation:

$$x_{t+1} = x_t + f(x_t, \theta_t), \quad t = 0, \dots, T-1, \quad (14)$$

where x_0 is the input, e.g., an image, several time-series, etc., while x_T is the final output, to be compared to some target y_T by means of a given loss function. By moving from a discrete time formulation to a continuous one, the forward dynamics we are interested in will be described by a differential equation that takes the role of (14). The learning aim is to tune the trainable parameters $\theta_0, \dots, \theta_{T-1}$ to have x_T as close as possible to y_T , according to a specified metric and knowing that the target y_T is joined to the input x_0 by means of a probability measure μ_0 .

Following the dynamical systems approach developed in [6], the supervised learning method aims to approximate some function \mathcal{F} , usually called the oracle, denoted by $\mathcal{F} : \mathcal{X} \rightarrow \mathcal{Y}$.

As stated before, the set $\mathcal{X} \subset \mathbb{R}^d$ contains the d -dimensional array of inputs, e.g., images, financial time-series, sound recorded data, text, etc., while \mathcal{Y} are the targets modelling the corresponding images, numerical forecast, or predicted texts.

In this setting, it is standard to define what is called a hypothesis space as:

$$\mathcal{H} = \{\mathcal{F}_\theta : \mathcal{X} \rightarrow \mathcal{Y} \mid \theta \in \Theta\}.$$

Training moves from a collection of K samples of input-target pairs $\{x^i, y^i = \mathcal{F}(x^i)\}_{i=1}^K$, the goal being to approximate \mathcal{F} exploiting these training data points.

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space supporting random variables $x_0 \in \mathbb{R}^d$ and $y_T \in \mathbb{R}^l$, jointly distributed according to $\mu_0 := \mathbb{P}_{(x_0, y_T)}$ with μ_0 modelling the distribution of the input-target pairs. The set of controls $\Theta \subseteq \mathbb{R}^m$ denotes the admissible training weights that are assumed to be essentially bounded, measurable $L^\infty([0, T], \Theta)$ functions. The network depth, i.e., the number of layers, is denoted by $T > 0$. We also introduce the functions:

- the feed-forward dynamics $f : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}^d$;
- the terminal loss function $\Phi : \mathbb{R}^d \times \mathbb{R}^l \rightarrow \mathbb{R}$;
- the regularization term $L : \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}$.

State dynamics are described by an Ordinary Differential Equation (ODE) of the form:

$$\dot{x}_t = f(x_t, \theta_t), \quad (15)$$

representing the continuous version of Equation (14), equipped with an initial condition x_0 , which is a random variable responsible for the randomness term characterizing Equation (15).

The population risk minimization problem in DL can be expressed by the following MF-optimal control problem (see [5] (p. 5)):

$$\inf_{\theta \in L^\infty([0, T], \Theta)} J(\theta) := \mathbb{E}_{\mu_0} \left[\Phi(x_T, y_T) + \int_0^T L(x_t, \theta_t) dt \right], \quad (16)$$

subject to the dynamics expressed by the stochastic ODE (15). Since the weights θ are shared by the distribution μ_0 of random variable (x_0, y_T) pairs, Equation (16) can be studied as an MF-optimal control problem.

On the other hand, the empirical risk minimization problem can be expressed by a sampled optimal control problem after computing i.i.d. samples $\{x_0^i, y_T^i\}_{i=1}^N$ modelled by $\mu_0 = \mathbb{P}_{(x_0, y_T)}$:

$$\inf_{\theta \in L^\infty([0, T], \Theta)} J_N(\theta) := \frac{1}{N} \sum_{i=1}^N \left[\Phi(x_T^i, y_T^i) + \int_0^T L(x_t^i, \theta_t) dt \right], \tag{17}$$

subject to the dynamics:

$$\dot{x}_t^i = f(x_t^i, \theta_t), \quad i = 1, \dots, N,$$

whose solutions, moving from random initial conditions through a deterministic path, correspond to random variables.

As in classical optimal control theory, the previous problem can be solved following two inter-connected approaches: a global theory, based on the Dynamic Programming Principle (DPP) leading to the HJB equation, or considering the Pontryagin Maximum Principle (PMP) approach, hence expressing the solution by a system of Forward Backward SDEs (FBSDEs) plus a local optimality condition.

3.2. HJB Equation

The idea behind the HJB formalism is to define a value function corresponding to the optimal loss of the control problem w.r.t. the general starting time and state. For the population risk minimization formulation expressed by Equation (16), the state argument of the value function corresponds to an infinite-dimensional object that models a joint distribution of the input-target as an element of a suitable Wasserstein space.

As regards random variables and their distribution, a suitable space must be defined for the rigorous treatment of the optimal control problem. In particular, we use the shorthand notation $L^2(\Omega, \mathbb{R}^{d+1})$ for $L^2((\Omega, \mathcal{F}, \mathbb{P}), \mathbb{R}^{d+1})$ to denote the set of \mathbb{R}^{d+1} -valued square integrable random variables w.r.t. a given probability measure \mathbb{P} . Then, we deal with a Hilbert space considering the norm:

$$\|X\|_{L^2} := \left(\mathbb{E}(\|X\|^2) \right)^{\frac{1}{2}}, \quad X \in L^2(\Omega, \mathbb{R}^{d+1}),$$

The set $\mathcal{P}_2(\mathbb{R}^{d+1})$ denotes the integrable probability measures defined on the Euclidean space \mathbb{R}^{d+1} . Let us recall that the random variable X is square integrable in $L^2(\Omega, \mathbb{R}^{d+1})$ if and only if its law $\mathbb{P}_X \in \mathcal{P}_2(\mathbb{R}^{d+1})$. The space $\mathcal{P}_2(\mathbb{R}^{d+1})$ can be endowed with a metric by considering the Wasserstein distance defined in Equation (2). For $p = 2$, the two-Wasserstein distance reads:

$$W_2(\mu, \nu) := \inf \left\{ \left(\int_{\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}} \|w - z\|^2 \pi(dw, dz) \right)^{\frac{1}{2}} \mid \pi \in \mathcal{P}_2(\mathbb{R}^{d+1} \times \mathbb{R}^{d+1}) \text{ with marginals } \mu \text{ and } \nu \right\},$$

according to the marginals introduced in Section 2.1, or equivalently:

$$W_2(\mu, \nu) := \inf \left\{ \|X - Y\|_{L^2} \mid X, Y \in L^2(\Omega, \mathbb{R}^{d+1}) \text{ with } \mathbb{P}_X = \mu, \mathbb{P}_Y = \nu \right\},$$

see, e.g., [5] (p. 6). Moreover, $\forall \mu \in \mathcal{P}_2(\mathbb{R}^{d+1})$, we define the associated norm:

$$\|\mu\|_{L^2} = \left(\int_{\mathbb{R}^{d+1}} \|w\|^2 \mu(dw) \right)^{\frac{1}{2}}.$$

Given a measurable function $\psi : \mathbb{R}^{d+1} \rightarrow \mathbb{R}^q$ that is square integrable w.r.t. the probability distribution μ , the following notation is introduced:

$$\langle \psi(\cdot), \mu \rangle := \int_{\mathbb{R}^{d+1}} \psi(w) \mu(dw). \tag{18}$$

Concerning the dynamical evolution of probability measures, let us fix $\zeta \in L^2(\Omega, \mathbb{R}^{d+l})$ and the control process $\theta \in L^\infty([0, T], \Theta)$. Then, the dynamics of the system can be written as:

$$W_s^{t,\zeta,\theta} = \zeta + \int_t^s \bar{f}(W_s^{t,\zeta,\theta}, \theta_t) ds, \quad s \in [t, T],$$

μ being the law associated with the variable ψ defined by $\mu = \mathbb{P}_\zeta \in \mathcal{P}_2(\mathbb{R}^{d+l})$, and we can rewrite the law of $W_s^{t,\zeta,\theta}$ as:

$$\mathbb{P}_s^{t,\mu,\theta} := \mathbb{P}_{W_s^{t,\zeta,\theta}}. \tag{19}$$

Indeed, the law involving the dynamics $W_s^{t,\zeta,\theta}$ depends only on the law of ζ and not on the random variable itself; see, e.g., [5] (p. 7).

It turns out that, to obtain the HJB Equation (5) corresponding to the above introduced formulation, it is necessary to define the concept of the derivative w.r.t. a probability measure. To begin with, it is useful to consider probability measures on \mathbb{R}^{d+l} as laws expressing probabilistic features of the \mathbb{R}^{d+l} -valued random variables defined over the probability space $(\Omega, \mathcal{F}, \mathbb{P})$. Then, we define the Banach space of random variables to define the derivatives. Moreover, if we define a function $u : \mathcal{P}_2(\mathbb{R}^{d+l}) \rightarrow \mathbb{R}$, it is possible to lift it into its extension U defined on $L^2([0, T], \mathbb{R}^{d+l})$, as follows:

$$U(X) = u(\mathbb{P}_X), \quad \forall X \in L^2(\Omega, \mathbb{R}^{d+l}),$$

then the definition of the derivative w.r.t. a probability measure can be expressed in terms of U in the usual Banach space setting. In particular, we have that u is $\mathcal{C}^1(\mathcal{P}_2(\mathbb{R}^{d+l}))$, if the lifted function U is Fréchet differentiable with continuous derivatives.

Since $L^2(\Omega, \mathbb{R}^{d+l})$ can be identified with its dual, if the Fréchet derivative $DU(X)$ exists, by Riesz's theorem, it can be identified with an element of $L^2(\Omega, \mathbb{R}^{d+l})$, i.e.,

$$DU(X)(Y) = \mathbb{E}[DU(X) \cdot Y], \quad \forall Y \in L^2(\Omega, \mathbb{R}^{d+l}).$$

It is worth underlining that $DU(X)$ does not depend on X , but only on the law described by X ; hence, the derivative of u at $\mu = \mathbb{P}_X$ is described by $\partial_\mu u(\mathbb{P}_X) : \mathbb{R}^{d+l} \rightarrow \mathbb{R}^{d+l}$, defined as:

$$DU(X) = \partial_\mu u(\mathbb{P}_X)(X).$$

By duality, we know that $\partial_\mu u(\mathbb{P}_X)$ is square integrable w.r.t. μ . To define a notion for the chain rule in $\mathcal{P}_2(\mathbb{R}^{d+l})$, a dynamical system is described by:

$$W_t = \zeta + \int_0^t \bar{f}(W_s) ds, \quad \zeta \in L^2(\Omega, \mathbb{R}^{d+l}),$$

where \bar{f} denotes the feed-forward dynamics. If a function $u \in \mathcal{C}^1(\mathcal{P}_2(\mathbb{R}^{d+l}))$, meaning that it is differentiable with a continuous derivative w.r.t. a probability measure, then, for all $t \in [0, T]$, we have:

$$u(\mathbb{P}_{W_t}) = u(\mathbb{P}_{W_0}) + \int_0^t \langle \partial_\mu u(\mathbb{P}_{W_s})(\cdot) \cdot \bar{f}(\cdot), \mathbb{P}_{W_s} \rangle ds, \tag{20}$$

where \cdot denotes the usual inner product between the vector in \mathbb{R}^{d+l} . Equivalently, exploiting the lifted function of u , we can state:

$$U(W_t) = U(W_0) + \int_0^t \mathbb{E}[DU(W_s) \cdot \bar{f}(W_s)] ds.$$

Moreover, the variable w denotes the concatenated $(d+l)$ -dimensional variable (x, y) , where $x \in \mathbb{R}^d$ and $y \in \mathbb{R}^l$. Correspondingly, $\bar{f}(w, \theta) := (f(x, \theta), 0)$ is the extended $(d+l)$ -

dimensional Feed-Forward Function (FFF), $\bar{L}(w, \theta) = L(x, \theta)$ is the extended $(d + l)$ -dimensional regularization loss and $\bar{\Phi}(w) := \Phi(x, y)$ represents the terminal loss function.

Since the state variable is identified with a probability distribution $\mu \in \mathcal{P}_2(\mathbb{R}^{d+l})$, the resulting objective functional can be defined as:

$$J(t, \mu, \theta) := \mathbb{E}_{(x_t, y_t) \sim \mu} \left[\Phi(x_T, y_T) + \int_t^T L(x_s, \theta_s) ds \right],$$

which can be written, with the concatenated variable w and the bracket notation introduced in (18), as:

$$J(t, \mu, \theta) := \langle \bar{\Phi}(w), \mathbb{P}_T^{t, \mu, \theta} \rangle + \int_t^T \langle \bar{L}(w, \theta_s), \mathbb{P}_s^{t, \mu, \theta} \rangle ds.$$

In this setting, some assumptions for the value function are needed to solve Equation (16). In particular:

1. f, L and Φ are bounded;
2. f, L and Φ are Lipschitz w.r.t. x , and the Lipschitz constant of f and L are independent of θ ;
3. $\mu_0 \in \mathcal{P}_2(\mathbb{R}^{d+l})$.

The value function $v^*(t, \mu)$ is defined as the real-valued function on $[0, T] \times \mathcal{P}_2(\mathbb{R}^{d+l})$, corresponding to the infimum of the functional J over the training parameters θ :

$$v^*(t, \mu) = \inf_{\theta \in L^\infty([0, T], \Theta)} J(t, \mu, \theta).$$

It is essential to observe how the value function satisfies a recursive relation based on the Dynamic Programming Principle (DPP). This implies that, for any optimal trajectory starting from any intermediate point, the remaining part of the trajectory still has to be optimal. The latter principle can be expressed by defining the value function as:

$$v^*(t, \mu) = \inf_{\theta \in L^\infty([0, T], \Theta)} \left[\int_t^{\hat{t}} \langle \bar{L}(\cdot, \theta_s), \mathbb{P}_s^{t, \mu, \theta} \rangle ds + v^*(\hat{t}, \mathbb{P}_{\hat{t}}^{t, \mu, \theta}) \right],$$

$\forall 0 \leq t \leq \hat{t} \leq T$ and $\mu \in \mathcal{P}_2(\mathbb{R}^{d+l})$.

Considering a small increment of time $\hat{t} = t + \delta t$ with $\delta > 0$, we can compute the Taylor expansion in the Wasserstein sense, hence obtaining:

$$0 = \inf_{\theta \in L^\infty([0, T], \Theta)} \left[v^*(t + \delta t, \mathbb{P}_{t+\delta t}^{t, \mu, \theta}) - v^*(t, \mu) + \int_t^{t+\delta t} \langle \bar{L}(w, \theta_s), \mathbb{P}_s^{t, \mu, \theta} \rangle ds \right].$$

By the chain rule in $\mathcal{P}_2(\mathbb{R}^{d+l})$, we have:

$$0 \approx \inf_{\theta \in L^\infty([0, T], \Theta)} \left[\partial_t v(t, \mu) \delta t + \int_t^{t+\delta t} \langle \partial_\mu v(t, \mu)(w) \cdot \bar{f}(w, \theta) + \bar{L}(w, \theta_s), \mu \rangle ds \right].$$

Since the infinitesimal δt does not affect the distribution μ and the controls θ (see [5] (p. 13)), integrating the second term, we have:

$$0 \approx \delta t \inf_{\theta \in L^\infty([0, T], \Theta)} \left[\partial_t v(t, \mu) + \langle \partial_\mu v(t, \mu)(w) \cdot \bar{f}(w, \theta) + \bar{L}(w, \theta), \mu \rangle \right].$$

Taking $\delta t \rightarrow 0$, we have:

$$\begin{cases} \frac{\partial v}{\partial t} + \inf_{\theta \in \Theta} \langle \partial_\mu v(t, \mu)(w) \cdot \bar{f}(w, \theta) + \bar{L}(w, \theta), \mu \rangle = 0, & \text{on } [0, T] \times \mathcal{P}_2(\mathbb{R}^{d+l}) \\ v(T, \mu) = \langle \bar{\Phi}(w), \mu \rangle, & \text{on } \mathcal{P}_2(\mathbb{R}^{d+l}). \end{cases} \quad (21)$$

Since the value function should solve the HJB equation, it is essential to find the precise link between the solution of this PDE and the value function obtained from the minimization of the functional J . To provide the result, we use a verification argument allowing the following consideration: if the solution of the HJB is smooth enough, then it corresponds to the value function v^* ; moreover, it allows computing the optimal control θ^* .

Theorem 1 (The verification argument). *Let v be a function in $C^{1,1}([0, T] \times \mathcal{P}_2(\mathbb{R}^{d+l}))$. If v is a solution of the HJB equation in (21) and there exists $\bar{\theta}$ that is mapping $(t, \mu) \rightarrow \Theta$ attaining the infimum in the HJB equation, then $v(t, \mu) = v^*(t, \mu)$, and $\bar{\theta}$ is an optimal feedback control policy, i.e., $\theta = \theta^*$ is a solution of the population risk minimization problem expressed by Equation (16), where $\theta_t^* := \bar{\theta}(t, \mathbb{P}_{w_t^*})$ with $\mathbb{P}_{w_0^*} = \mu_0$ and $\frac{dw_t^*}{dt} = \bar{f}(w_t^*, \theta_t^*)$.*

Proof of Theorem 1. Given any control process θ , applying Formula (20) between $s = t$ and $s = T$ with explicit time dependence gives:

$$v(T, \mathbb{P}_T^{t, \mu, \theta}) = v(t, \mu) + \int_t^T \frac{\partial v}{\partial t}(s, \mathbb{P}_s^{t, \mu, \theta}) + \langle \partial_\mu v(s, \mathbb{P}_s^{t, \mu, \theta})(\cdot) \cdot \bar{f}(\cdot, \theta_s), \mathbb{P}_s^{t, \mu, \theta} \rangle ds,$$

Equivalently:

$$\begin{aligned} v(t, \mu) &= v(T, \mathbb{P}_T^{t, \mu, \theta}) - \int_t^T \frac{\partial v}{\partial t}(s, \mathbb{P}_s^{t, \mu, \theta}) + \langle \partial_\mu v(s, \mathbb{P}_s^{t, \mu, \theta})(\cdot) \cdot \bar{f}(\cdot, \theta_s), \mathbb{P}_s^{t, \mu, \theta} \rangle ds \\ &\leq v(T, \mathbb{P}_T^{t, \mu, \theta}) + \int_t^T \langle \bar{L}(\cdot, \theta_s), \mathbb{P}_s^{t, \mu, \theta} \rangle ds \\ &= \langle \bar{\Phi}(\cdot), \mathbb{P}_t^{T, \mu, \theta} \rangle + \int_t^T \langle \bar{L}(\cdot, \theta_s), \mathbb{P}_s^{t, \mu, \theta} \rangle ds \\ &= J(t, \mu, \theta), \end{aligned}$$

where the first inequality comes from the infimum condition in (21).

Since the control is arbitrary, we have:

$$v(t, \mu) \leq v^*(t, \mu),$$

then it can be substituted with θ^* where $\theta_t^* = \bar{\theta}(t, \mathbb{P}_s^{t, \mu, \theta^*})$ is computed by the optimal feedback control. Repeating the above argument, the inequality becomes an equality since the infimum is attained for $\bar{\theta}$:

$$v(t, \mu) = J(t, \mu, \theta^*) \geq v^*(t, \mu).$$

Thus, $v(t, \mu) = v^*(t, \mu)$, and $\bar{\theta}$ defines an optimal control policy. For more details, see [5] (Proposition 3, pp. 13–14). \square

The importance of Theorem 1 consists of linking smooth solutions of the parabolic PDE to the solutions of the population risk minimization problem, becoming a natural candidate for the DL problem.

Moreover, the optimal control policy $\bar{\theta} : [0, T] \times \mathcal{P}_2(\mathbb{R}^{d+l}) \rightarrow \Theta$ is identified by computing the infimum in (21). Hence, it turns out that the HJB equation strongly characterizes the learning problem's solution for the feedback, or closed-loop networks: control weights are actively adjusted according to the outputs, and this is the essential feature of closed-loop control. Nevertheless, the solution comes from a PDE that is in general difficult to solve, even numerically. On the other hand, open-loop solutions can be obtained from the closed-loop control policy by sequentially setting $\theta_t^* = \bar{\theta}(t, \mathbb{P}_{w_t^*})$, w_t^* being the solution of the feed-forward ODE describing the dynamics of the state variable, with $\theta = \theta^*$ up to time t . Usually within DL settings, open-loops are used during training or to measure the inference of a trained model, since trained weights for each neuron will have a fixed value.

The great limit of such a formulation relies in assuming that the value function $v^*(t, \mu)$ is continuously differentiable. It is straightforward to study a more flexible characterization for v^* dealing with weak solutions, also denoted as viscosity solutions. Thus, it is worth considering a weaker formulation of the PDE to go beyond the concept of classical solutions, by introducing the notion of viscosity ones, hence allowing obtaining relevant results when dealing with weaker assumptions on the coefficients defining the (stochastic) differential problem we are interested in; see, e.g., [5] (Section 5 pp. 14–22) for more details.

The key idea relies on exploiting the lifting identification between measures and random variables and moving from the Wasserstein space $\mathcal{P}_2(\mathbb{R}^{d+1})$ to the Hilbert space $L^2(\Omega, \mathbb{R}^{d+1})$, using tools developed to study viscosity solutions.

We introduce a functional defined as the Hamiltonian for viscosity formulation $\mathcal{H}(\xi, P) : L^2(\Omega, \mathbb{R}^{d+1}) \times L^2(\Omega, \mathbb{R}^{d+1}) \rightarrow \mathbb{R}$ through:

$$\mathcal{H}(\xi, P) = \inf_{\theta \in \Theta} \mathbb{E}[P \cdot \bar{f}(\xi, \theta) + \bar{L}(\xi, \theta)]. \quad (22)$$

Then, the lifted Bellman equation can be written w.r.t. $V(t, \xi) = v(t, \mathbb{P}_\xi)$ as follows:

$$\begin{cases} \frac{\partial V}{\partial t} + \mathcal{H}(\xi, DV(t, \xi)) = 0 & \text{on } [0, T] \times L^2(\Omega, \mathbb{R}^{d+1}), \\ V(T, \xi) = \mathbb{E}[\bar{\Phi}(\xi)] & \text{on } L^2(\Omega, \mathbb{R}^{d+1}); \end{cases} \quad (23)$$

hence, the PDE we are analysing is now set within a larger space corresponding to $L^2(\Omega, \mathbb{R}^{d+1})$.

We say that a bounded, uniformly continuous function $u : [0, T] \times \mathcal{P}_2(\mathbb{R}^{d+1}) \rightarrow \mathbb{R}$ is a viscosity solution of HJB Equation (21) if its lifted function $U : [0, T] \times L^2(\Omega, \mathbb{R}^{d+1}) \rightarrow \mathbb{R}$ defined by:

$$U(t, \xi) = u(t, \mathbb{P}_\xi),$$

is a viscosity solution to the lifted Bellman Equation (23), namely:

1. $U(T, \xi) \leq \mathbb{E}[\bar{\Phi}(\xi)]$ and for any test function $\psi \in \mathcal{C}^{1,1}([0, T] \times L^2(\Omega, \mathbb{R}^{d+1}))$ such that $(U - \psi)$ has a local maximum at $(t_0, \xi_0) \in [0, T] \times L^2(\Omega, \mathbb{R}^{d+1})$, ψ solves:

$$\partial_t \psi(t_0, \xi_0) + \mathcal{H}(\xi_0, D\psi(t_0, \xi_0)) \geq 0.$$

2. $U(T, \xi) \geq \mathbb{E}[\bar{\Phi}(\xi)]$ and for any test function $\psi \in \mathcal{C}^{1,1}([0, T] \times L^2(\Omega, \mathbb{R}^{d+1}))$ such that the map $(U - \psi)$ has a local minimum at $(t_0, \xi_0) \in [0, T] \times L^2(\Omega, \mathbb{R}^{d+1})$, ψ solves:

$$\partial_t \psi(t_0, \xi_0) + \mathcal{H}(\xi_0, D\psi(t_0, \xi_0)) \leq 0.$$

Actually, the unique solution of this formulation corresponds to the value function v^* from the minimization problem; see, e.g., [5] (Theorem 1, p. 15). Therefore, the HJB equation provides both the necessary and sufficient condition for the optimality of the learning procedure.

Adopting the MF-optimal control viewpoint implies that the population risk minimization problem of DL can be studied as a variational problem, whose solution is characterized by a suitable HJB equation, in analogy with the classical calculus of variations. In other words, the HJB equation is a global characterization of the value function to be solved over the entire space $\mathcal{P}_2(\mathbb{R}^{d+1})$ of input-target distributions. From the numerical point of view, it is a hard task to get a solution for the entire space; this is why the learning problem is typically locally solved, around some (small set of) trajectories generated according to the initial condition $\mu_0 \in \mathcal{P}_2(\mathbb{R}^{d+1})$, then applying the obtained feedback to nearby input-target distributions.

3.3. Mean-Field Pontryagin Maximum Principle

We have seen how the HJB approach provides a characterization of the optimal solution for the population risk minimization problem that holds globally in $\mathcal{P}_2(\mathbb{R}^{d+1})$,

at the price of being difficult to handle in practice. Moving from this consideration, the MF-PMP aims to show a local condition for optimality, expressed in terms of $\mathbb{E}[H]$, i.e., the expectation of the Hamiltonian function.

Starting from the population risk minimization problem defined in Equation (16) and given a collection of K sample input-target pairs, the single i th input sample is considered. The prediction of the network can be approximated by a deterministic transformation of the terminal state $g(X_T^i)$ for some $g : \mathbb{R}^d \rightarrow \mathcal{Y}$ that models a function both of the initial input x^i and of the control parameters θ . Moreover, we define a loss function $\Phi : \mathcal{Y} \times \mathcal{Y} \rightarrow \mathbb{R}$, which is minimized when the arguments are equal. Therefore, the goal is to minimize:

$$\sum_{i=1}^K \Phi(g(X_T^i), y^i).$$

Since g is fixed, it can be absorbed into the definition of the loss function by defining the array $\Phi_i(\cdot) := \Phi(\cdot, y_i)$.

Then, the supervised learning problem can be expressed as:

$$\min_{\theta \in \mathcal{U}} \left[\sum_{i=1}^K \Phi_i(X_T^i) + \int_0^T L(\theta_t) dt \right], \quad (24)$$

where $L : \Theta \rightarrow \mathbb{R}$ acts as a regularizer term to model a running cost.

Input variables $x = (x^1, \dots, x^K)$ can be considered as the elements of a Euclidean space $\mathbb{R}^{d \times K}$, representing the initial conditions of the following ODE system:

$$\dot{X}_t^i = f_{\theta_t}(t, X_t^i), \quad X_0^i = x^i, \quad 0 \leq t \leq T, \quad i = 1, \dots, K, \quad (25)$$

where $\theta : [0, T] \rightarrow \Theta$ are the control parameters to be trained. The dynamics (25) are decoupled except for the control. A general space \mathcal{U} for controls θ is then defined as: $\mathcal{U} := \{\theta : [0, T] \rightarrow \Theta : \theta \text{ is Lebesgue measurable}\}$, and we are aiming to choose θ in \mathcal{U} to have $g(X_T^i)$ closer to y^i for $i = 1, \dots, K$.

To formulate the PMP as a set of necessary conditions for optimal solutions, it is useful to define the Hamiltonian $H : [0, T] \times \mathbb{R}^d \times \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}$ given by:

$$H(t, x, p, \theta) := p \cdot f(t, x, \theta) - L(\theta),$$

with p modelling an adjoint process as in Equation (6).

Let us underline that all input-target pairs (x_0, y_0) , connected by the distribution μ_0 , share a common control parameter, and this feature suggests the idea to develop a maximum condition that has to hold in the average sense. Indeed, the control is now enforced on the continuity equation that describes the evolution of probability densities.

The following assumptions are needed:

1. f is bounded and f, L are continuous w.r.t. θ ;
2. f, L and Φ are continuously differentiable w.r.t. x ;
3. the distribution μ_0 has bounded support in $\mathbb{R}^d \times \mathbb{R}^l$, which means there exists $M > 0$ such that $\mu(\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^l : \|x\| + \|y\| \leq 1\}) = 1$.

Theorem 2 (Mean-Field Pontryagin Maximum Principle). *Let assumptions 1–3 hold and $\theta^* \in L^\infty([0, T], \Theta)$ be the minimizer for $J(\theta)$ corresponding to the optimal control of the population risk minimization problem (16). Define the Hamiltonian $H : \mathbb{R}^d \times \mathbb{R}^d \times \Theta \rightarrow \mathbb{R}$ as*

$$H(x, p, \theta) = p \cdot f(x, \theta) - L(x, \theta). \quad (26)$$

Then, there exist absolutely continuous stochastic processes x^* and p^* solving the following forward backward SDEs:

$$\dot{x}_t^* = f(x_t^*, \theta_t^*), \quad x_t^* = x_0, \tag{27}$$

$$\dot{p}_t^* = -\nabla_x H(x_t^*, p_t^*, \theta_t^*), \quad p_T^* = -\nabla_x \Phi(x_T^*, y_0), \tag{28}$$

and the related optimality condition expressed in terms of the expectation of the Hamiltonian function:

$$\mathbb{E}_{\mu_0} H(x_t^*, p_t^*, \theta_t^*) \geq \mathbb{E}_{\mu_0} H(x_t^*, p_t^*, \theta), \quad \forall \theta \in \Theta, \text{ for a.e. } t \in [0, T]. \tag{29}$$

Proof of Theorem 2. For the sake of simplicity, let us introduce a new coordinate x^0 satisfying the dynamics $\dot{x}_t^0 = L(x_t^*, \theta_t^*)$ with $x_0^0 = 0$. Through this choice, the definition of the Hamiltonian in Equation (26) can be rewritten without running loss L by redefining:

$$x \rightarrow (x^0, x), \quad f \rightarrow (L, f), \quad \Phi(x_T, y_0) \rightarrow \Phi(x_T, y_0) + x_T^0.$$

Assumptions 1–3 are still preserved, but now we consider without loss of generality the case $L \equiv 0$.

Let some $\tau \in (0, T]$ be a Lebesgue point of $\hat{f}(t) := f(x_t^*, \theta_t^*)$; in this setting, these points are dense in $[0, T]$. Now, for $\varepsilon \in (0, \tau)$, define the family of perturbed controls:

$$\theta_t^{\tau, \varepsilon} = \begin{cases} \omega & t \in [\tau - \varepsilon, \tau] \\ \theta_t^* & \text{otherwise} \end{cases},$$

where $\omega \in \Theta$ is an admissible control; this kind of perturbation is called needle perturbation. Accordingly, define $x_t^{\tau, \varepsilon}$ by:

$$x_t^{\tau, \varepsilon} = x_0 + \int_0^t f(x_s^{\tau, \varepsilon}, \theta_s^{\tau, \varepsilon}) ds,$$

that is the solution of the forward propagation equation with the perturbed control $\theta^{\tau, \varepsilon}$. Clearly, $x_t^* = x_t^{\tau, \varepsilon}$ for every $t \leq \tau - \varepsilon$ and every x_0 since the perturbation is not present. At the limit point $t = \tau$, the following holds:

$$\frac{1}{\varepsilon}(x_\tau^{\tau, \varepsilon} - x_\tau^*) = \frac{1}{\varepsilon} \int_{\tau - \varepsilon}^\tau f(x_s^{\tau, \varepsilon}, \omega) - f(x_s^*, \theta_s^*) ds,$$

and since τ is a Lebesgue point of F :

$$v_\tau := \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon}(x_\tau^{\tau, \varepsilon} - x_\tau^*) = f(x_\tau^*, \omega) - f(x_\tau^*, \theta_\tau^*).$$

It is possible to characterize v_τ as the leading order perturbation on the state due to the needle perturbation introduced in the infinitesimal interval $[\tau - \varepsilon, \tau]$. In interval $(\tau, T]$, the dynamics is the same before applying the perturbation since the controls are the same.

Now, it is necessary to consider how the perturbation v_τ propagates. Thus, define for $t \geq \tau$:

$$v_t^\varepsilon := \frac{1}{\varepsilon}(x_t^{\tau, \varepsilon} - x_t^*),$$

and:

$$v_t := \lim_{\varepsilon \rightarrow 0} v_t^\varepsilon,$$

v_t is well-defined for almost every t , which are every Lebesgue point of the map $t \mapsto x^*(t)$, and it satisfies the following linearised equation:

$$\begin{cases} \dot{v}_t = \nabla_x f(x_t^*, \theta_t^*)^T v_t, & t \in (\tau, T] \\ v_\tau = f(x_\tau^*, \omega) - f(x_\tau^*, \theta_\tau^*). \end{cases} \quad (30)$$

In particular, $v(T)$ represents the perturbation of the final state introduced by this control. By the optimality assumption of θ^* , it follows that:

$$\mathbb{E}_{\mu_0} \Phi(x_T^{\tau, \varepsilon}, y_0) \geq \mathbb{E}_{\mu_0} \Phi(x_T^*, y_0).$$

Assumptions 1 and 2 (p. 13) imply $\nabla_x \Phi$ is bounded. By the dominated convergence theorem, we know that:

$$\begin{aligned} 0 &\leq \lim_{\varepsilon \rightarrow 0} \frac{1}{\varepsilon} \mathbb{E}_{\mu_0} [\Phi(x_T^{\tau, \varepsilon}, y_0) - \Phi(x_T^*, y_0)] \\ &= \mathbb{E}_{\mu_0} \left. \frac{d}{d\varepsilon} \Phi(x_T^{\varepsilon, \tau}, y_0) \right|_{\varepsilon=0^+} \\ &= \mathbb{E}_{\mu_0} \nabla_x \Phi(x_T^{\varepsilon, \tau}, y_0) \cdot v_T. \end{aligned} \quad (31)$$

Let us define p^* as the solution of the adjoint of Equation (30), hence:

$$\dot{p}_t^* = -\nabla_x f(x_t^*, \theta_t^*) p_t^*, \quad p_T^* = -\nabla_x \Phi(x_T^*, y_0).$$

By Equation (31), it follows that $\mathbb{E}_{\mu_0} p_T^* \cdot v_T \leq 0$, and moreover, for all $t \in [\tau, t]$:

$$\frac{d}{dt} (p_t^* \cdot v_t) = \dot{p}_t^* \cdot v_t + \dot{v}_t \cdot p_t^* = 0;$$

thus,

$$\mathbb{E}_{\mu_0} p_\tau^* \cdot v_t = \mathbb{E}_{\mu_0} p_T^* \cdot v_T \leq 0, \quad \forall t \in [\tau, T],$$

so that taking $t = \tau$:

$$\mathbb{E}_{\mu_0} p_\tau^* \cdot f(x_\tau^*, \theta_\tau^*) \geq \mathbb{E}_{\mu_0} p_T^* \cdot f(x_\tau^*, \omega).$$

Since ω is arbitrarily chosen, this completes the proof by recalling that $H(x, p, \theta) = p \cdot f(x, \theta)$. See [5] (Theorem 3, pp. 23–24) for more details. \square

MF-PMP refers only to the control of the open-loop type; Equation (27) is a feed-forward ODE, describing the state dynamics under optimal controls θ^* . Equation (28) defines the evolution of the co-state variable p_s^* , characterizing the evolution of an adjoint variational condition backwards in time. It is interesting to note how the optimality condition described in Equation (29) does not involve first order partial derivatives, being expressed in terms of expectations. In particular, it requires that optimal solutions must globally maximize the Hamiltonian function. This aspect allows considering also cases of non-differentiable dynamics w.r.t. the controls weights, as well as cases characterized by optimal weights lying on the boundary of the training set Θ . Moreover, the usual first order optimality condition can be derived from (29). Comparing this MF formulation to the classical PMP, we can see that the main difference lies in the fact that the maximization condition is expressed in terms of the expectation above a probability density μ_0 . The latter result is not surprising, since the mean-field-optimal control must depend on the probability distribution of input-target pairs.

Let us also note that the mean-field PMP expressed in Theorem 2 can be written more compactly as follows. For each control process $\theta \in L^\infty([0, T], \Theta)$, we denote by $x^\theta := \{x_t^\theta : 0 \leq t \leq T\}$ and $p^\theta := \{p_t^\theta : 0 \leq t \leq T\}$ the solutions of Hamilton's

Equations (27) and (28), and we enforce the control expressed by the random variables $(x_0, y_T) \sim \mu_0$, through:

$$\begin{aligned}x_t^\theta &= f(x_t^\theta, \theta_t), & x_0^\theta &= x_0, \\ \dot{p}_t^\theta &= -\nabla_x H(x_t^\theta, p_t^\theta, \theta_t), & p^\theta &= -\nabla_x \Phi(x_T^\theta, y_t).\end{aligned}$$

Then, θ^* satisfies the PMP if and only if:

$$\mathbb{E}_{\mu_0} H(x_t^{\theta^*}, p_t^{\theta^*}, \theta^*) \geq \mathbb{E}_{\mu_0} H(x_t^{\theta^*}, p_t^{\theta^*}, \theta), \quad \forall \theta \in \Theta. \quad (32)$$

Furthermore, observe that the mean-field PMP includes, as a special case, the necessary conditions for the optimality of the sampled optimal control problem (17). In order to point out this aspect, define the empirical measure:

$$\mu_0^N := \frac{1}{N} \sum_{i=1}^N \delta_{(x_0^i, y_T^i)},$$

and apply the mean-field PMP with μ_0^N in place of μ_0 to obtain:

$$\frac{1}{N} \sum_{i=1}^N H(x_t^{\theta^*, i}, p_t^{\theta^*, i}, \theta^*) \geq \frac{1}{N} \sum_{i=1}^N H(x_t^{\theta^*, i}, p_t^{\theta^*, i}, \theta), \quad \forall \theta \in \Theta, \quad (33)$$

where $x_t^{\theta^*, i}$ and $p_t^{\theta^*, i}$ are defined through the input-target pair (x_0^i, y_T^i) . Moreover, since μ_0^N is a random measure, (33) is a random equation whose solutions correspond to random variables.

Concerning possible numerical analysis of the DL algorithm based on the maximum principle, we refer to [4] (pp. 13–15), where a comparison can be found between usual gradient approaches to the discrete formulation of the Mean-Field PMP stated in Theorem 2, with a loss function based on Equation (24). The test and train losses of some variants of SGD algorithms are compared to the mean-field algorithm based on the discrete PMP. It is possible to observe that the latter algorithm is characterized by a better convergence rate, being then faster. This improvement is mainly due to the fact that it allows avoiding possibly getting stuck, caused by the flat regions, as clearly shown by the graphs reported in [4] (p. 14).

3.4. Connection between the HJB Equation and the PMP

In what follows, we provide connections between the global and local formulation of the HJB formalism via the PMP, exploiting the connection between Hamilton's canonical equations (ODEs) and the Hamilton–Jacobi equations (PDEs). The Hamiltonian dynamics of Equations (27) and (28) describe the trajectory of a random variable that is completely determined by random variables (x_0, y_T) . On the other hand, the optimality condition described by Equation (29) does not depend on the particular probability measure of the initial input-target pairs. Notice that the maximum principle can be expressed in terms of a Hamiltonian flow that depends on a probability measure in a suitable Wasserstein space and where Equation (29) is the corresponding lifting version. Analogously, in order to have both solutions in the same functional space, HJB has to be lifted in the $L^2(\Omega, \mathbb{R}^{d+1})$ space.

Starting from the lifted Bellman Equation (23), lying in $L^2(\Omega, \mathbb{R}^{d+1})$, it is possible to apply the method of the characteristics and define the following system of equations, after introducing $P_t = DV(t, \xi_t)$:

$$\begin{cases} \dot{\xi}_t = D_P \mathcal{H}(\xi_t, P_t) \\ \dot{P}_t = -D_\xi \mathcal{H}(\xi_t, P_t). \end{cases} \quad (34)$$

Suppose Equation (34) has a solution that satisfies an initial condition given by:

$$\mathbb{P}_{\xi_0} = \mu_0,$$

and a terminal one involving Bellman equation given by:

$$P_T = \nabla_w \bar{\Phi}(\xi_T).$$

We also assume that $\bar{\theta}(\xi, P)$ is the optimal control achieving the infimum in (21), as an interior point of Θ , then we can explicitly write the equation of the Hamiltonian w.r.t. the optimal control as:

$$H = \mathbb{E}[P \cdot \bar{f}(\xi, \bar{\theta}(\xi, P)) + \bar{L}(\xi, \bar{\theta}(\xi, P))].$$

Therefore, by the first order condition, we have:

$$\mathbb{E}[\nabla_{\theta} \bar{f}(\xi, \bar{\theta}(\xi, P))P + \nabla_{\theta} \bar{L}(\xi, \bar{\theta}(\xi, P))] = 0,$$

so that, taking into consideration Equation (34), we obtain the Hamilton-type equations:

$$\begin{cases} \dot{\xi}_t = \bar{f}(\xi_t, \bar{\theta}(\xi_t, P_t)) \\ \dot{P}_t = -\nabla_w \bar{f}(\xi_t, \bar{\theta}(\xi_t, P_t))P_t - \nabla_w \bar{L}(\xi_t, \bar{\theta}(\xi_t, P_t)). \end{cases} \quad (35)$$

Use $w = (x, y)$ as concatenated variable and $\theta^* = \bar{\theta}(\xi_t, P_t)$ to remark that the last l components of \bar{f} are zero and by considering only the first d components:

$$\begin{cases} \dot{x}_t = f(\xi_t, \theta^*) \\ \dot{p}_t = -\nabla_x f(x_t, \theta^*)p_t - \nabla_x L(x_t, \theta_t^*). \end{cases} \quad (36)$$

Summing up: Hamilton's equation of the system (36) can be viewed as the characteristic equations of the HJB equation in its lifted formulation described by Equation (23). Essentially, the PMP gives a necessary condition for any characteristic of the HJB equation: any characteristic originating from μ_0 , that is the initial law of the random variables, must satisfy a local necessary optimal condition constituted by the mean-field PMP. This justifies the claim that the PMP constitutes a local condition if compared to the HJB equation.

3.5. Small-Time Uniqueness

A natural question is to understand when the PMP solutions also provide sufficient conditions to have optimality. We start by considering that the uniqueness of the solution implies sufficiency, and we investigate which assumptions are needed to have a unique solution of the mean-field PMP equations. Equations (27) and (28) model highly non-linear two-point boundary value problems in terms of x^* and p^* , which are also coupled through their laws. In general, even without the coupling, this kind of PDE is known to not have a unique solution; see, e.g., Ch. 7 of [22]. In order to prove the uniqueness, strong assumptions are needed to prove the small-time case.

Theorem 3 (Small-time uniqueness). *Suppose that:*

- f is bounded;
- f, L and Φ are continuously differentiable w.r.t. both x and θ with bounded and Lipschitz partial derivatives;
- the distribution μ_0 has bounded support in $\mathbb{R}^d \times \mathbb{R}^l$, which means there exists $M > 0$ such that $\mu\{(x, y) \in \mathbb{R}^d \times \mathbb{R}^l : \|x\| + \|y\| \leq 1\} = 1$;
- $H(x, p, \theta)$ is strongly concave in θ and uniformly in $x, p \in \mathbb{R}^d$, i.e., $\nabla_{xx}^2 H(x, p, \theta) + \lambda_0 \preceq 0$ for some $\lambda_0 \geq 0$.

Then, for sufficiently small T , if θ^{*1} and θ^{*2} are solutions of the mean-field PMP derived in Theorem 2, then $\theta^{*1} = \theta^{*2}$.

Before proving the theorem, we report the following lemma, which provides an estimate of the difference between flow-maps driven by two different controls in terms of the small-time parameter T :

Lemma 1. Let $\theta^1, \theta^2 \in L^\infty([0, T], \Theta)$. Then, there exists a constant T_0 such that for all $T \in [0, T_0)$, it holds that:

$$\|x^{\theta^1} - x^{\theta^2}\|_{L^\infty} + \|p^{\theta^1} - p^{\theta^2}\|_{L^\infty} \leq C(T)\|\theta^1 - \theta^2\|_{L^\infty},$$

where $C(T) > 0$ satisfies $C(T) \rightarrow 0$ as $T \rightarrow 0$.

Proof of Lemma 1. We denote $\delta\theta := \theta^1 - \theta^2$, $\delta x := x^{\theta^1} - x^{\theta^2}$ and $\delta p = p^{\theta^1} - p^{\theta^2}$, respectively. Since $x_0^{\theta^1} = x_0^{\theta^2} = x_0$, integrating the respective ODEs while considering the first two assumptions of Theorem 6 leads to:

$$\|\delta x_t\| \leq \int_0^t \|f(x_s^{\theta^1}, \theta_s^1) - f(x_s^{\theta^2}, \theta_s^2)\| ds \leq K_L \int_0^t \|\delta x_s\| ds + K_L \int_0^t \|\delta p_s\| ds,$$

and so:

$$\|\delta x\|_{L^\infty} \leq K_L T \|\delta x\|_{L^\infty} + K_L T \|\delta\theta\|_{L^\infty}.$$

Now, if $T \leq T_0 := \frac{1}{K_L}$. We then have:

$$\|\delta x\|_{L^\infty} \leq \frac{K_L T}{1 - K_L T} \|\delta\theta\|_{L^\infty}. \quad (37)$$

Similarly:

$$\|\delta p_t\| \leq K_L \|\delta x_T\| + K_L \int_t^T \|\delta x_s\| ds + K_L \int_t^T \|\delta p_s\| ds$$

$$\|\delta p\| \leq (K_L + K_L T) \|\delta x\|_{L^\infty} + K_L T \|\delta p\|_{L^\infty},$$

and hence:

$$\|\delta p\|_{L^\infty} \leq \frac{K_L(1 + T)}{1 - K_L T} \|\delta x\|_{L^\infty},$$

which combined with Equation (37) proves the lemma. \square

Through the above estimate, it is now possible to prove Theorem 3.

Proof of Theorem 3. In what follows, we exploit [5] (p. 29). By uniform strong concavity, the function $\theta \rightarrow \mathbb{E}_{\mu_0} H(x_t^{\theta^1}, p_t^{\theta^1}, \theta)$ is strongly concave. Thus, consider a $\lambda_0 > 0$ such that:

$$\frac{\lambda_0}{2} \|\theta_t^1 - \theta_t^2\|^2 \leq \left[\mathbb{E}_{\mu_0} \nabla H(x_t^{\theta^1}, p_t^{\theta^1}, \theta_t^2) - \mathbb{E}_{\mu_0} \nabla H(x_t^{\theta^1}, p_t^{\theta^1}, \theta_t^1) \right] \cdot (\theta_t^1 - \theta_t^2).$$

A similar expression holds for $\theta \rightarrow \mathbb{E}_{\mu_0} H(x_t^{\theta^2}, p_t^{\theta^2}, \theta)$, and by combining the two inequalities, also exploiting the smoothness of functions involved, we have:

$$\begin{aligned} \lambda_0 \|\theta_t^1 - \theta_t^2\|^2 &\leq \left[\mathbb{E}_{\mu_0} \nabla H(x_t^{\theta^1}, p_t^{\theta^1}, \theta_t^2) - \mathbb{E}_{\mu_0} \nabla H(x_t^{\theta^1}, p_t^{\theta^1}, \theta_t^1) \right] \cdot (\theta_t^1 - \theta_t^2) \\ &\quad + \left[\mathbb{E}_{\mu_0} \nabla H(x_t^{\theta^2}, p_t^{\theta^2}, \theta_t^1) - \mathbb{E}_{\mu_0} \nabla H(x_t^{\theta^2}, p_t^{\theta^2}, \theta_t^2) \right] \cdot (\theta_t^1 - \theta_t^2) \\ &\leq \mathbb{E}_{\mu_0} \|\nabla H(x_t^{\theta^1}, p_t^{\theta^1}, \theta_t^1) - \nabla H(x_t^{\theta^2}, p_t^{\theta^2}, \theta_t^1)\| \|\theta_t^1 - \theta_t^2\| \\ &\quad + \mathbb{E}_{\mu_0} \|\nabla H(x_t^{\theta^1}, p_t^{\theta^1}, \theta_t^2) - \nabla H(x_t^{\theta^2}, p_t^{\theta^2}, \theta_t^2)\| \|\theta_t^1 - \theta_t^2\| \\ &\leq K_L \|\delta\theta\|_{L^\infty} (\|\delta x\|_{L^\infty} + \|\delta p\|_{L^\infty}). \end{aligned}$$

Combining the above inequality with Lemma 1, we have:

$$\|\delta\theta\|_{L^\infty}^2 \leq \frac{K_L}{\lambda_0} C(T) \|\delta\theta\|_{L^\infty}^2.$$

However, $C(T) = o(1)$, and by taking T sufficiently small, so that $K_L C(T) < \lambda_0$, this implies that $\|\delta\theta\|_{L^\infty} = 0$. For more details, see [5] (Lemma 2, p. 28). \square

Within the ML setting, the small-period T roughly corresponds to the regime where the reachable set of the forward dynamics is small. This can be interpreted as the case where the model has low capacity or low expressive power. In other words, Theorem 3 states that the optimal solution is unique if the network capacity is low, even when there is a huge number of parameters. It is essential to underline that the hypothesis of the strong concavity of the Hamiltonian H does not imply convexity for the loss function J , which can also be highly non-convex due to the non-linear transformation introduced by the activation functions σ .

4. Conclusions

In this paper, typical NN structures are considered as the evolution of a dynamical system to rigorously state the population risk minimization problem related to DL. Two parallel and connected perspectives are followed. The result expressed in Theorem 2 represents the generalization of the PMP in the calculus of variations and a local characterization of optimal trajectories derived from HJB Equation (21). Deriving the necessary condition for the optimality of the PMP provides several advantages: there is no reference to the derivative w.r.t. the probability measure in the Wasserstein sense; the parameter set Θ is highly general; maximization is point-wise in t , once x^* and p^* are known. Moreover, it allows the possibility to develop a learning algorithm without referring to the classical methods of DL such as the SGD. As an interest point with a general relevance within the optimal control theory, we remark that the controls, and then the weights θ , are only assumed to be measurable and essentially bounded in time. In particular, we allow for discontinuous terms. Even in the case where the number of parameters is infinite, as in the MFG setting, it is feasible to derive non-trivial stability estimates. This aspect results in contrast with generalized bounds based on classical statistical learning, where the increasing number of parameters adversely affects the generalization. On the other hand, it will be interesting to consider algorithms with weights having only discrete values, e.g., binary, values, requiring small memory amounts, with corresponding efficiency and speed improvements. A further direction of our ongoing research is based on the study of the PMP from a discrete-time perspective in order to relate the theoretical framework directly to the applications.

Author Contributions: Conceptualization, L.D.P. and M.G.; Investigation, L.D.P. and M.G.; Methodology, L.D.P. and M.G.; Writing—original draft, L.D.P. and M.G.; Writing—review and editing, L.D.P. and M.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. Cardaliguet, P. *Notes on Mean-Field Games*; Technical Report, from P.-L. Lions' lectures at Collège de France; Collège de France: Paris, France, 2012.
2. Touzi, N. *Stochastic Control and Application to Finance*; Chapter 1–4; Ecole Polytechnique: Paris, France, 2018.
3. Carmona, R.; Delarue, F.; Lachapelle, A. *Control of Mc-Kean-Vlasov Dynamics Versus Mean-Field Games*; Technical Report; Springer: Berlin/Heidelberg, Germany, 2012.
4. Li, Q.; Chen, L.; Tai, C.; E W. Maximum principle based algorithms for deep learning. *J. Mach. Learn. Res.* **2018**, *18*, 1–29.
5. Weinan, E.; Han, J.; Li, Q. A Mean-Field Optimal Control Formulation of Deep Learning. *arXiv* **2018**, arXiv:1807.01083.
6. Li, Q.; Lin, T.T.; Shen Z. Deep Learning via Dynamical Systems: An Approximation Perspective. *arXiv* **2019**, arXiv:1912.10382v1.

7. Athans, M.; Falb, P.L. *Optimal Control: An Introduction to the Theory and Its Applications*; Courier Corporation, Dover Publications, Inc.: Mineola, NY, USA, 2013.
8. Lacker, D. *Mean-Field Games and Interacting Particle Systems*. 2018. Available online: <http://www.columbia.edu/~dl3133/MFGSpring2018.pdf> (accessed on 17 November 2020).
9. Gangbo, W.; Kim, H.K.; Pacini, T. Differential forms on Wasserstein space and infinite-dimensional Hamiltonian systems. *arXiv* **2009**, arXiv:0807.1065.
10. Sagitov, S. *Weak Convergence of Probability Measures*; Chalmers University of Technology and Gothenburg University: 2015. Available online: <http://www.math.chalmers.se/~serik/WeakConv/C-space.pdf> (accessed on 17 November 2020).
11. Billingsley, P. *Weak Convergence in Metric Spaces*; Wiley Series in Probability and Statistics; Hoboken, NJ, USA, 1999.
12. Bergström, H. Weak Convergence of Measures. In *A Volume in Probability and Mathematical Statistics: A Series of Monographs and Textbooks*; Springer: New York, NY, USA, 1982.
13. Benazzoli, C.; Campi, L.; Di Persio, L. Mean field games with controlled jump–diffusion dynamics: Existence results and an illiquid interbank market model. *Stoch. Process. Their Appl.* **2020**, *130*, 6927–6964. [[CrossRef](#)]
14. Benazzoli, C.; Campi, L.; Di Persio, L. ϵ –Nash equilibrium in stochastic differential games with mean-field interaction and controlled jumps. *Stat. Probab. Lett.* **2019**. [[CrossRef](#)]
15. Carrillo, J.A.; Pimentel, E.A.; Voskanyan, V.K. On a mean-field optimal control problem. *Nonlinear Anal. Theory Methods Appl.* **2020**, *199*, 112039. [[CrossRef](#)]
16. Liberzon, D. *Calculus of Variations and Optimal Control Theory: A Concise Introduction*; Princeton University Press: Princeton, NJ, USA, 2012.
17. Ma, J.; Yong, J. *Forward-Backward Stochastic Differential Equations and Their Applications*; Springer: New York, NY, USA, 2007.
18. Hadikhanloo, S. Learning in Mean-Field Games. Ph.D. Thesis, Dauphine Université Paris, Paris, France, 2018.
19. Fleming, W.H.; Soner, H.M. *Controlled Markov Processes and Viscosity Solutions*; Stochastic Modelling and Applied Probability; Springer: New York, NY, USA, 2006.
20. Frankowska, H. Hamilton-Jacobi equations: Viscosity solutions and generalized gradients. *J. Math. Anal. Appl.* **1989**, *141*, 1. [[CrossRef](#)]
21. Weinan, E. A Proposal on Machine Learning via Dynamical Systems. *arXiv* **2019**, arXiv:1912.10382v1.
22. Kelley, W.G.; Peterson, A.C. *The Theory of Differential Equations: Classical and Qualitative*; Springer Science and Business Media: Berlin, Germany, 2010.