

## NORMALIZATION EFFECTS ON DEEP NEURAL NETWORKS

JIAHUI YU<sup>✉1</sup> AND KONSTANTINOS SPILIOPOULOS<sup>✉\*1</sup>

<sup>1</sup>Department of Mathematics and Statistics, Boston University, Boston, USA

(Communicated by Feng Bao)

**ABSTRACT.** We study the effect of normalization on the layers of deep neural networks of feed-forward type. A given layer  $i$  with  $N_i$  hidden units is allowed to be normalized by  $1/N_i^{\gamma_i}$  with  $\gamma_i \in [1/2, 1]$  and we study the effect of the choice of the  $\gamma_i$  on the statistical behavior of the neural network's output (such as variance) as well as on the test accuracy on the MNIST data set. We find that in terms of variance of the neural network's output and test accuracy the best choice is to choose the  $\gamma_i$ 's to be equal to one, which is the mean-field scaling. We also find that this is particularly true for the outer layer, in that the neural network's behavior is more sensitive in the scaling of the outer layer as opposed to the scaling of the inner layers. The mechanism for the mathematical analysis is an asymptotic expansion for the neural network's output. An important practical consequence of the analysis is that it provides a systematic and mathematically informed way to choose the learning rate hyperparameters. Such a choice guarantees that the neural network behaves in a statistically robust way as the  $N_i$  grow to infinity.

**1. Introduction.** The last few years have experienced an explosion in the study of neural networks. Neural networks are parametric models and their coefficients are estimated from data using gradient descent methods. Early classical results regarding the approximation power of neural networks [4, 17, 16] set the stage and then advances in technology led to great successes in text, speech and image recognition, see for example [24, 14, 6, 43, 3, 26, 46] to name a few. Later on, neural networks showed a lot of promise in other fields such as robotics, medicine, finance, and applied mathematics, see for example [27, 28, 15, 34, 10, 1, 37, 36, 38]. Their success in applications has made clearer the need for a better understanding of their mathematical properties.

The goal of this paper is to investigate the performance of multilayer neural networks as a function of normalization features. In particular, let us consider the following neural network with two hidden layers:

$$g_{\theta}^{N_1, N_2}(x) = \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C^i \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W^{2,j,i} \sigma(W^{1,j}x) \right), \quad (1)$$

---

2020 *Mathematics Subject Classification.* Primary: 60F05, 68T01, 60G99.

*Key words and phrases.* Machine learning, neural networks, normalization effect, scaling effects, asymptotic expansions, out-of-sample performance.

K.S. was partially supported by the National Science Foundation (DMS 2107856) and Simons Foundation Award 672441].

\*Corresponding author: Konstantinos Spiliopoulos.

where  $C^i, W^{2,j,i} \in \mathbb{R}$ ,  $x, W^{1,j} \in \mathbb{R}^d$ , and  $\gamma_1, \gamma_2 \in [1/2, 1]$  are fixed scaling parameters. For convenience, we write  $W^{1,j}x = \langle W^{1,j}, x \rangle_{l^2}$  as the standard  $l^2$  inner product for the vectors. The neural network model has parameters

$$\theta = (C^1, \dots, C^{N_2}, W^{2,1,1}, \dots, W^{2,N_1,N_2}, W^{1,1}, \dots, W^{1,N_1}),$$

which are to be estimated from data  $(X, Y) \sim \pi(dx, dy)$ . For fixed  $\theta, \gamma_1, \gamma_2$ , the neural network  $g_\theta^{N_1, N_2}$  is a map from  $\mathbb{R}^d$  to  $\mathbb{R}$ .

Our goal is to understand the effect of the choice of the values of the scaling parameters  $\gamma_1, \gamma_2 \in [1/2, 1]$  on the behavior of the neural network. The choice  $\gamma_1 = \gamma_2 = 1$  corresponds to the mean field scaling that has been studied in the literature in recent years, see for example [7, 30, 35, 39, 40, 41]. On the other side of the spectrum, i.e, when  $\gamma_1 = \gamma_2 = 1/2$ , then we have the so-called Xavier normalization [13], giving rise to the so-called neural tangent kernel, that has been analyzed in a number of works, see for example [20, 9, 18, 8, 42]. Even though, most of the discussion of this paper is focused on the two-layer neural network, in Section 3.2, see also Section 4, we discuss the three-layer neural network case demonstrating that our conclusions extend to general feed-forward multilayer neural networks.

In the case of shallow neural networks (SNN), i.e, when  $g_\theta^N(x) = \frac{1}{N^\gamma} \sum_{i=1}^N C^i \sigma(W^i x)$ , the question on the effect of  $\gamma \in [1/2, 1]$  on the performance of the neural network has been recently studied in [45]. In [45] we developed an asymptotic expansion for the neural network's statistical output  $g^N$  after training with stochastic gradient descent (SGD) pointwise with respect to the scaling parameter  $\gamma \in (1/2, 1)$  as the number of hidden units  $N$  grows to infinity. Based on this expansion [45] demonstrates mathematically that to leading order in  $N$  both bias and variance (both explicitly characterized) decrease as the number of hidden units increases and time grows. In addition, it is shown there that to leading order in  $N$ , the variance of the neural network's statistical output  $g^N$  is monotonically decreasing in  $\gamma$  and thus the lowest variance is attained at  $\gamma = 1$ . Numerical studies on the MNIST and CIFAR10 datasets in [45] showed that test and train accuracy monotonically improve as the neural network's normalization gets closer to the mean field normalization  $\gamma = 1$ . An additional useful conclusion of the mathematical analysis in [45] is that in order for the asymptotic results to be true (without trivial limits) one needs to choose the learning rate in SGD in a very specific way with respect to  $N$  and  $\gamma$ .

The content of this paper is the corresponding analysis in the case of deep neural networks (DNN). As we will see the analysis in the case of DNNs is considerably more complicated than in the case of SNN. However, the end conclusions are of similar nature with the additional interesting observation that the outer layer plays a more special role. In addition, the analysis of this paper offers a mathematically principled way to appropriately choose the learning rates. We base our analysis on a typical two-layer neural network, however as we shall see in Section 3 this is done without loss of generality.

In particular, we derive an asymptotic expansion of the neural network's output as  $N_2 \rightarrow \infty$  with  $N_1$  fixed. This expansion shows mathematically that to leading order in  $N_2$ , the variance of the neural network's statistical output is monotonically decreasing with respect to  $\gamma_2 \in [1/2, 1]$ . At the same time, the same expression (after appropriately choosing the learning rates) shows that the effect of  $\gamma_1$  is perhaps less prominent in the sense that it appears through terms that are averages and are also bounded (for bounded activated functions). The mathematical conclusion is that,

at least under our assumptions (as presented in Section 2) one would optimally choose the outer layer normalization to be  $\gamma_2 = 1$  and subsequently choosing  $\gamma_1 = 1$  would be optimal. This conclusion is also validated numerically. Indeed, in Section 3 we study the test accuracy of two and three layer neural networks for different parametrizations in terms of  $\gamma_1, \gamma_2 \in [1/2, 1]$  (and  $\gamma_3 \in [1/2, 1]$  in the three-layer neural network case) when trained with standard SGD on the MNIST dataset [25]. As we shall see there, the test accuracy is sensitive to the choice of the normalization of the outer layer  $\gamma_2$  with the optimal choice being  $\gamma_2 = 1$ , but having done that, the effect of the choice of the normalization of the inner layer, i.e., of  $\gamma_1$  is less profound. The end optimal choice is to choose  $\gamma_2 = \gamma_1 = 1$ , i.e., the mean-field normalization in all layers.

An additional important conclusion of this work is that it provides a systematic and mathematically informed way to choose the learning rates hyperparameters, see (6) for the model (1), Section 3.2 for the three-layer case and Section 4 for the general case. Without choosing the learning rates to be of the indicated order with respect to the  $N_i$ 's and  $\gamma_i$ 's the neural network as a statistical object will have trivial limits, i.e., it will either converge to zero or to infinity. If however, they are chosen in the indicated way then the neural network will behave nicely as a statistical quantity in the sense of not being trivial and having finite variance at least.

Our analysis is based upon the quadratic error loss function

$$L(\theta) = \frac{1}{2} \mathbb{E}_{X,Y} \left[ \left( Y - g_{\theta}^{N_1, N_2}(x) \right)^2 \right],$$

and the model parameters  $\theta$  are trained by the stochastic gradient descent algorithm, for  $k \in \mathbb{N}$

$$\begin{aligned} C_{k+1}^i &= C_k^i + \frac{\alpha_c^{N_1, N_2}}{N_2^{\gamma_2}} \left( y_k - g_k^{N_1, N_2}(x_k) \right) H_k^{2,i}(x_k), \\ W_{k+1}^{1,j} &= W_k^{1,j} + \frac{\alpha_{W,1}^{N_1, N_2}}{N_1^{\gamma_1}} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \\ &\quad \left( \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_k^i \sigma'(Z_k^{2,i}(x_k)) W_k^{2,j,i} \right) \sigma'(W_k^{1,j} x_k) x_k, \\ W_{k+1}^{2,j,i} &= W_k^{2,j,i} + \frac{\alpha_{W,2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{\gamma_2}} \left( y_k - g_k^{N_1, N_2}(x_k) \right) C_k^i \sigma'(Z_k^{2,i}(x_k)) H_k^{1,j}(x_k), \end{aligned} \tag{2}$$

where

$$H_k^{1,j}(x) = \sigma(W_k^{1,j} x), \quad Z_k^{2,i}(x) = \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} H_k^{1,j}(x), \quad H_k^{2,i}(x) = \sigma(Z_k^{2,i}(x)).$$

and  $\alpha_c^{N_1, N_2}, \alpha_{W,1}^{N_1, N_2}, \alpha_{W,2}^{N_1, N_2}$  are the learning rates.

For fixed  $N_1$ , we define the empirical measure

$$\tilde{\gamma}_k^{N_1, N_2} = \frac{1}{N_2} \sum_{i=1}^{N_2} \delta_{C_k^i, W_k^{2,1,i}, \dots, W_k^{2,N_1,i}, W_k^{1,1}, \dots, W_k^{1,N_1}}, \tag{3}$$

and the time-scaled empirical measure

$$\gamma_t^{N_1, N_2} = \tilde{\gamma}_{\lfloor N_2 t \rfloor}^{N_1, N_2} \in \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}). \tag{4}$$

Here  $\mathcal{M}(\mathbb{R}^{1+N_1(1+d)})$  is the space of probability measures in  $\mathbb{R}^{1+N_1(1+d)}$ . Define the function  $Z^{2,N_1}(x) = Z^{2,N_1}(x; w^{2,j}, w^{1,j}) = N_1^{-\gamma_1} \sum_{j=1}^{N_1} w^{2,j} \sigma(w^{1,j}x)$ . The neural network output can be rewritten as

$$g_{\theta_k}^{N_1, N_2}(x) = \left\langle c\sigma(Z^{2,N_1}(x)), N_2^{1-\gamma_2} \tilde{\gamma}_k^{N_1, N_2} \right\rangle,$$

and the time-scaled neural network output is

$$h_t^{N_1, N_2}(x) = g_{\theta_{\lfloor N_2 t \rfloor}}^{N_1, N_2}(x). \quad (5)$$

For a fixed data set  $(x^{(i)}, y^{(i)})_{i=1}^M$ , let  $g_k^{N_1, N_2}$  and  $h_t^{N_1, N_2}$  denote the  $M$ -dimensional vectors whose  $i$ -th entries are  $g_k^{N_1, N_2}(x^{(i)})$  and  $h_t^{N_1, N_2}(x^{(i)})$ , respectively. In order to emphasize the dependence on  $\hat{\gamma} = (\gamma_1, \gamma_2)$  and on  $\hat{N} = (N_1, N_2)$  we will instead write sometimes  $h_t^{\hat{N}, \hat{\gamma}}$ .

As it will be demonstrated below, it turns out that in order to understand the main effects of  $\gamma_1, \gamma_2 \in (1/2, 1)$  on the behavior of  $h_t^{\hat{N}, \hat{\gamma}}$  it is enough to look at its asymptotic behavior as  $N_2 \rightarrow \infty$  with the  $N_1$  being thought of as large but fixed.

In addition, the learning rates need to be chosen to be of the right order with respect to the number of hidden units  $N_i$  and network normalization  $\gamma_i$  in order for the neural network to behave in a statistically robust way. In particular, for reasons that will become clearer later on, we shall choose the learning rates to be

$$\alpha_C^{N_1, N_2} = \frac{\alpha_C}{N_2^{2-2\gamma_2}}, \quad \alpha_{W,1}^{N_1, N_2} = \frac{\alpha_{W,1}}{N_1^{1-2\gamma_1} N_2^{3-2\gamma_2}}, \quad \alpha_{W,2}^{N_1, N_2} = \frac{\alpha_{W,2}}{N_1^{1-2\gamma_1} N_2^{2-2\gamma_2}}, \quad (6)$$

where the coefficients  $\alpha_C, \alpha_{W,1}, \alpha_{W,2} \in (0, \infty)$  are chosen to be of order one with respect to  $N_1, N_2$ .

Loosely speaking our main mathematical result is that for each fixed  $\gamma_2 \in (1/2, 1)$  one has that as  $N_2 \rightarrow \infty$ , and when  $\gamma_2 \in \left(\frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2}\right)$  for fixed  $\nu \in \{1, 2, 3, \dots\}$  and fixed  $\gamma_1$  and  $N_1$ :

$$\begin{aligned} h_t^{\hat{N}, \hat{\gamma}} \approx h_t^{N_1, \gamma_1} &+ \sum_{j=1}^{\nu-1} N_2^{-j(1-\gamma_2)} Q_{j,t}^{N_1, \gamma_1} \\ &+ N_2^{-(\gamma_2-1/2)} e^{-A^{N_1, \gamma_1} t} \mathcal{G}^{N_1} + \text{lower order terms in } N_2, \end{aligned} \quad (7)$$

where  $h_t^{N_1, \gamma_1}$  is the limit of  $h_t^{\hat{N}, \hat{\gamma}}$  as  $N_2 \rightarrow \infty$ ,  $Q_{j,t}^{N_1, \gamma_1}$  are deterministic quantities,  $A^{N_1, \gamma_1}$  is a positive definite matrix and  $\mathcal{G}^{N_1}$  is a Gaussian vector of mean zero and known variance-covariance structure. The formal definitions of these terms are presented in Section 2. Noticeably, all of  $h_t^{N_1, \gamma_1}$ ,  $Q_{j,t}^{N_1, \gamma_1}$ ,  $A^{N_1, \gamma_1}$  and  $\mathcal{G}^{N_1}$  are not only independent of  $N_2 < \infty$  and  $\gamma_2 > 0$ , but the dependence on  $N_1$  is through explicit averages of the form  $\frac{1}{N_1} \sum_{i=1}^{N_1} (\dots)$ , and the dependence on  $\gamma_1$  is only through the terms  $\sigma(Z_k^{2,i}(x))$ ,  $\sigma'(Z_k^{2,i}(x))$  which for bounded  $\sigma \in C_b^1(\mathbb{R})$  will be bounded.

Even though we do not show this here, as in [42, 45], one gets that for all  $\gamma_1, \gamma_2 \in (1/2, 1)$  and for all  $N_1 < \infty$ , the limit of the network output recovers the global minimum as  $t \rightarrow \infty$ , i.e.  $h_t^{N_1, \gamma_1} \rightarrow \hat{Y}$ , where  $\hat{Y} = (y^{(1)}, \dots, y^{(M)})$ . For fixed  $j \in \mathbb{N}$ , one can also show exactly as in [45] that  $Q_{j,t}^{N_1, \gamma_1} \rightarrow 0$  exponentially fast as  $t \rightarrow \infty$ . The Gaussian vector  $\mathcal{G}^{N_1}$  is related to the variance of the network at initialization which then propagates forward, see (9).

These conclusions immediately suggest that the variance of  $h_t^{\hat{N}, \hat{\gamma}}$  to leading order in  $N_2$  is monotonically decreasing in  $\gamma_2 \in [1/2, 1]$ , with the smallest possible variance

when  $N_2$  is large, but fixed, when  $\gamma_2 = 1$ . In addition, the fact that the dependence of the leading order terms in the right hand side of (7) on  $N_1$  and on  $\gamma_1$  is through averages of the form  $\frac{1}{N_1} \sum_{i=1}^{N_1} (\dots)$  for  $N_1$  and through bounded terms for  $\gamma_1$  (given that the activation function  $\sigma \in C_b^1(\mathbb{R})$ ), demonstrates that  $h_t^{\hat{N}, \hat{\gamma}}$  is less sensitive on the value of  $\gamma_1$ . The latter observation is also confirmed numerically in Section 3.

To further validate and demonstrate these conclusions we perform in Section 3 extensive numerical studies fitting two and three layer feed-forward neural networks on the MNIST dataset [25]. In all of the examples, the pattern is the same and corroborates the theoretical conclusions. Namely, the test accuracy is sensitive in the choice of the normalization of the outer layer  $\gamma_2$  with the optimal choice being  $\gamma_2 = 1$ , but having done that, the choice of the normalization of the inner layer, i.e., of  $\gamma_1$  has less of an impact on the performance. The end optimal choice is to choose  $\gamma_2 = \gamma_1 = 1$ , i.e., the mean-field normalization in all layers.

At this point we want to emphasize that the goal of this paper is not to study the limit as  $N_2, N_1 \rightarrow \infty$ . We refer the interested reader to [20, 9, 41, 2, 33] for related results. Our goal here is to disentangle the effect of different scalings in different layers. With this goal in mind, it turns out that it is enough to fix  $N_1$ , look at  $N_2 \rightarrow \infty$  and then observe that at least to leading order in  $N_2$  the effect of  $N_1$  is only through averages that converge to well defined limits. In addition, in the process of doing so, we obtain that the effect of  $\gamma_2$  is to scale the variance in a very simple and intuitive way as demonstrated by (7). On the other hand, the effect of  $\gamma_1$  is through bounded terms when at least the activation function and its derivatives are bounded. Also, we note that in order to obtain expansions like (7) one needs not only to characterize the asymptotic behavior of  $h_t^{\hat{N}, \hat{\gamma}}$ , but also needs to understand the fluctuations (central limit theorem) corrections, corrections to those corrections, etc. Lastly, our numerical studies indicate, see Figures 3 and 4, that test accuracy is better when  $N_2 > N_1$ , which also motivates looking at  $N_2 \rightarrow \infty$ .

The rest of the paper is organized as follows. In Section 2 we lay down our main assumptions and present the main mathematical results of the paper. In Section 3 we discuss the theoretical results further and we present our numerical studies. In Section 4 we present for completeness and without proof the mathematically motivated choice of the learning rates for a deep feedforward neural network of arbitrary depth. Conclusions are in Section 5. The proof of the main results presented in Section 2 are presented in the appendix of this paper. In Appendix A we establish apriori bounds on the learning parameters as they evolve in time. In Appendix B we prove Theorem 2.4. In Appendix C we prove Theorem 2.7. In Appendix D we prove Theorem 2.9. Then in Appendix E we complete the proof of the asymptotic expansion for  $h_t^{N_1, N_2}$  for  $\gamma_2 \in (1/2, 1)$  through an inductive argument.

**2. Assumptions and main results.** In this section, we describe our main assumptions under which the results of this paper hold and we present our main results. We also establish necessary notation. We work on a filtered probability space  $(\Omega, \mathcal{F}, \mathbb{P})$  where all the random variables are defined. The probability space is equipped with a filtration  $\mathfrak{F}_t$  that is right continuous and  $\mathfrak{F}_0$  contains all  $\mathbb{P}$ -negligible sets.

**Assumption 2.1.**

1. The activation function  $\sigma \in C_b^\infty(\mathbb{R})$ , i.e.  $\sigma$  is infinitely differentiable and bounded.
2. There is a fixed dataset  $\mathcal{X} \times \mathcal{Y} = (x^{(i)}, y^{(i)})_{i=1}^M$ , and we set  $\pi(dx, dy) = \frac{1}{M} \sum_{i=1}^M \delta_{(x^{(i)}, y^{(i)})}(dx, dy)$ .

3. The initialized parameters  $\{C_0^i\}_i, \{W_0^{2,j,i}\}_{i,j}, \{W_0^{1,j}\}_j$  are i.i.d., generated from mean-zero random variables and take values in compact sets  $\mathcal{C}, \mathcal{W}^1$ , and  $\mathcal{W}^2$ .

We recall that we shall choose the learning rates to be

$$\alpha_C^{N_1, N_2} = \frac{\alpha_C}{N_2^{2-2\gamma_2}}, \quad \alpha_{W,1}^{N_1, N_2} = \frac{\alpha_{W,1}}{N_1^{1-2\gamma_1} N_2^{3-2\gamma_2}}, \quad \alpha_{W,2}^{N_1, N_2} = \frac{\alpha_{W,2}}{N_1^{1-2\gamma_1} N_2^{2-2\gamma_2}},$$

where the coefficients  $\alpha_C, \alpha_{W,1}, \alpha_{W,2} \in (0, \infty)$  are chosen to be of order one with respect to  $N_1, N_2$ . For notational convenience and without loss of generality we shall set them to be  $\alpha_C = \alpha_{W,1} = \alpha_{W,2} = 1$ . Note that the weights in different layers are trained with different rates. This choice of learning rates is necessary for convergence to a non-trivial limit as  $N_2 \rightarrow \infty$ . If the parameters in all the layers are trained with the same learning rate, it can be mathematically shown that the network will not train as  $N_1, N_2$  become large in the sense of having convergence to trivial limits.

Before presenting our main mathematical results let us first discuss what happens at time  $t = 0$ . Denote  $\mu_C(dc)$  and  $\mu_{W^2}(dw)$  the probability distributions of  $\{C_0^i\}_i$  and  $\{W_0^{2,j,i}\}_{i,j}$  respectively. By law of large numbers, as  $N_2 \rightarrow \infty$ , we have that  $\tilde{\gamma}_0^{N_1, N_2} \xrightarrow{p} \gamma_0^{N_1}(dw^1, dw^2, dc)$ , where  $\tilde{\gamma}_0^{N_1, N_2}$  is as defined in (3) for  $k = 0$ , and

$$\begin{aligned} \gamma_0^{N_1}(dw^1, dw^2, dc) &= \delta_{W_0^{1,1}}(dw^{1,1}) \times \cdots \times \delta_{W_0^{1,N_1}}(dw^{1,N_1}) \\ &\quad \times \mu_{W^2}(dw^{2,1}) \times \cdots \times \mu_{W^2}(dw^{2,N_1}) \times \mu_C(dc). \end{aligned} \quad (8)$$

By the central limit theorem, we have in distribution

$$N_2^{(\gamma_2 - \frac{1}{2})} h_0^{N_1, N_2}(x) = \left\langle c\sigma(Z^{2,N_1}(x)), \sqrt{N_2} \tilde{\gamma}_0^{N_1, N_2} \right\rangle \xrightarrow{d} \mathcal{G}^{N_1}(x), \text{ as } N_2 \rightarrow \infty \quad (9)$$

where  $\mathcal{G}^{N_1}$  is a Gaussian random variable and variance  $\lambda_{N_1}^2(x) = \left\langle |c\sigma(Z^{2,N_1}(x))|^2, \gamma_0^{N_1} \right\rangle$ . Note that if  $\gamma_2 < 1/2$  the limit (9) suggests that  $h_0^{N_1, N_2}(x)$  grows to infinity as  $N_2 \rightarrow \infty$ , suggesting that the restriction to  $\gamma_2 \in [1/2, 1]$  is well motivated. From now on, we will use the notation  $\mathcal{G}^{N_1}$  to refer to this specific Gaussian random variable.

Hence, when  $\gamma_2 = 1/2$ , one has that  $h_0^{N_1, N_2}(x) \xrightarrow{d} \mathcal{G}^{N_1}(x)$ , and when  $\gamma_2 > 1/2$ ,  $h_0^{N_1, N_2}(x) \xrightarrow{d} 0$ .

**Remark 2.2.** Notice now that due to the independence assumption from (2.1), the sequence of random variables  $\{Z^{2,N_1}(x)\}_{N_1}$ , which is the input to the assumed bounded activation function  $\sigma$ , will also converge to a Gaussian with mean zero and finite variance in the limit  $N_2 \rightarrow \infty$  if  $\gamma_1 = 1/2$  and to the trivial limit  $Z^{2,N_1}(x) \rightarrow 0$  if  $\gamma_1 \in (1/2, 1)$ .

Certain quantities will appear many times, so let's define them here.

$$\begin{aligned} B_{x,x'}^1(\theta) &= \sigma(Z^{2,N_1}(x')) \sigma(Z^{2,N_1}(x)), \\ B_{x,x'}^{2,j}(\theta) &= (c)^2 \sigma'(Z^{2,N_1}(x')) \sigma'(Z^{2,N_1}(x)) \sigma(w^{1,j}x') \sigma(w^{1,j}x), \\ B_x^{3,j}(\theta) &= cw^{2,j} \sigma'(w^{1,j}x) \sigma'(Z^{2,N_1}(x)), \end{aligned} \quad (10)$$

and set

$$\begin{aligned} A_{x,x'}^{N_1} &= \left\langle B_{x,x'}^1(\theta), \gamma_0^{N_1} \right\rangle \\ &\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \left[ \left\langle B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1} \right\rangle + xx' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right] \end{aligned} \quad (11)$$

In addition, for a given  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$  let us define

$$\begin{aligned} C_{x'}^{N_1,f}(\theta) &= \partial_c f(\theta) \sigma(Z^{2,N_1}(x')) + \frac{1}{N_1^{1-\gamma_1}} c \sigma'(Z^{2,N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta) \\ &\quad + \frac{1}{N_1^{1-\gamma_1}} \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x' \end{aligned} \quad (12)$$

Even though we do not explore this further here, we note that the dependence of  $A^{N_1}$  on  $N_1$  is through averages of the form  $\frac{1}{N_1} \sum_{j=1}^{N_1} (\dots)$  and thus by Assumption 2.1 and law of large numbers convergence as  $N_1 \rightarrow \infty$  is expected to hold. A fully rigorous justification of the latter claim is beyond the scope and purposes of this article and is left for future work.

**Remark 2.3.** In a snapshot the theorems that follow essentially establish that for large  $N_2$  the neural network output behaves as

- $\gamma_2 \in (\frac{1}{2}, \frac{3}{4}]: h_t^{N_1, N_2} \approx h_t^{N_1} + \frac{1}{N^{1-\gamma_2-\frac{1}{2}}} K_t^{N_1}$  where  $K_t$  satisfies either of equations (16) or (17) and has a Gaussian distribution,
- $\gamma_2 \in (\frac{3}{4}, \frac{5}{6}]: h_t^{N_1, N_2} \approx h_t^{N_1} + \frac{1}{N^{1-\gamma_2}} K_t^{N_1} + \frac{1}{N^{1-\gamma_2-\frac{1}{2}}} \Psi_t^{N_1}$ , where  $K_t^{N_1}$  satisfies equation (17) with  $K_0^{N_1}(x) = 0$ ,  $\Psi_t^{N_1}$  satisfies either equations (19) or (20) and has a Gaussian distribution,

where  $h_t^{N_1}$  is the limit of  $h_t^{N_1, N_2}$  as  $N_2 \rightarrow \infty$ . Under the appropriate assumptions,  $h_t^{N_1}$  recovers the global minimum as  $t \rightarrow \infty$ . We note that, as expected this is in parallel to what one observes in the one layer case of [45]. However, what is potentially interesting here is that the outer layer dominates the behavior.

Our first result is related to the convergence of the pair  $(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2})$  as defined by (4) and (5) as  $N_2 \rightarrow \infty$ . We study the convergence in the Skorokhod space  $D_E([0, T])$ , where  $E = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M$ , and  $N_1 \in \mathbb{N}$  is fixed. Recall that  $\mathcal{M}(\mathbb{R}^{1+N_1(1+d)})$  is the space of probability measures in  $\mathbb{R}^{1+N_1(1+d)}$ .

**Theorem 2.4.** *Let  $T < \infty$  be given. Under Assumption 2.1, for fixed  $\gamma_1, \gamma_2 \in (1/2, 1)$  and learning rates chosen via (6), we get that as  $N_2 \rightarrow \infty$ , the process  $(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2})$  converges in probability in the space  $D_E([0, T])$  to  $(\gamma_t^{N_1}, h_t^{N_1})$ , which for  $t \in [0, T]$ , satisfies the evolution equation*

$$h_t^{N_1}(x) = h_0^{N_1}(x) + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) A_{x,x'}^{N_1} \pi(dx', dy) ds, \quad (13)$$

where  $h_0^{N_1}(x) = 0$ . In addition, we have that for any  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$  and  $t \in [0, T]$ ,  $\langle f, \gamma_t^{N_1} \rangle = \langle f, \gamma_0^{N_1} \rangle$ .

For some of our results we would need to further assume the following.

**Assumption 2.5.**

1. The activation function  $\sigma$  is smooth, non-polynomial and slowly increasing<sup>1</sup>.
2. The fixed dataset  $(x^{(i)}, y^{(i)})_{i=1}^M$  from part (ii) of Assumption 2.1 has data points that are in distinct directions (per definition on page 192 of [19]).

In a similar manner now to [45] and to [42] we get that under Assumption 2.5 and for any  $N_1 \in \mathbb{N}$  the matrix  $A^{N_1} \in \mathbb{R}^{M \times M}$ , whose elements are  $A_{x,x'}^{N_1}$  with  $x, x' \in \mathcal{X}$ ,

<sup>1</sup>A function  $\sigma(x)$  is called slowly increasing if  $\lim_{x \rightarrow \infty} \frac{\sigma(x)}{x^a} = 0$  for every  $a > 0$ .

is positive definite. The latter immediately says that we have convergence to the global minimum (see Theorem 4.2 in [42] for a proof)

$$h_t^{N_1} \rightarrow \hat{Y} \quad \text{as } t \rightarrow \infty. \quad (14)$$

where  $h_t^{N_1} = (h_t^{N_1}(x^{(1)}), \dots, h_t^{N_1}(x^{(M)}))$  and  $\hat{Y} = (y^{(1)}, \dots, y^{(M)})$ .

We note that with these choices of learning rates, the aforementioned convergence is true for any  $N_1 \in \mathbb{N}$ .

Since for  $\gamma_2 \in (1/2, 1)$  the first order limit is deterministic it makes sense to investigate the second order convergence. In particular, consider

$$K_t^{N_1, N_2} = N_2^\varphi (h_t^{N_1, N_2} - h_t^{N_1}),$$

where  $\varphi$  depends on the scaling parameters  $\gamma_1, \gamma_2$  and will be chosen appropriately momentarily. We also denote  $\eta_t^{N_1, N_2} = N_2^\varphi (\gamma_t^{N_1, N_2} - \gamma_0^{N_1})$ . For  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$  let us also define  $l_t^{N_1, N_2}(f) = \langle f, \eta_t^{N_1, N_2} \rangle$ .

Then, we have the following results.

**Proposition 2.6.** *Let Assumption 2.1 hold and choose the learning rates via (6). Then, for fixed  $\gamma_1, \gamma_2 \in (1/2, 1)$  and fixed  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ , if  $\varphi \leq 1 - \gamma_2$ , the process  $\{l_t^{N_1, N_2}(f) = \langle f, \eta_t^{N_1, N_2} \rangle, t \in [0, T]\}_{N_2 \in \mathbb{N}}$  converges in probability in the space  $D_{\mathbb{R}}([0, T])$  as  $N_2 \rightarrow \infty$ , and*

*Case 1. If  $\varphi < 1 - \gamma_2$ ,  $\langle f, \eta_t^{N_1, N_2} \rangle \rightarrow 0$ .*

*Case 2. If  $\varphi = 1 - \gamma_2$ ,  $l_t^{N_1, N_2}(f) = \langle f, \eta_t^{N_1, N_2} \rangle \rightarrow l_t^{N_1}(f)$ , where  $l_t^{N_1}(f)$  is given by*

$$l_t^{N_1}(f) = \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \langle C_{x'}^{N_1, f}(\theta), \gamma_0^{N_1} \rangle \pi(dx', dy) ds \quad (15)$$

**Theorem 2.7.** *Let Assumption 2.1 hold and choose the learning rates via (6). Let  $\mathcal{G}^{N_1}(x)$  be the Gaussian random variable defined in (9). Then, as  $N_2 \rightarrow \infty$ , the sequence of processes  $\{K_t^{N_1, N_2}, t \in [0, T]\}_{N_2 \in \mathbb{N}}$  converges in distribution in the space  $D_{\mathbb{R}^M}([0, T])$  to  $K_t^{N_1}$ , such that, depending on the values of  $\gamma$  and  $\phi$ , we shall have*

*Case 1. When  $\gamma \in (\frac{1}{2}, \frac{3}{4})$  and  $\varphi \leq \gamma_2 - \frac{1}{2}$ , or when  $\gamma_2 \in [\frac{3}{4}, 1)$  and  $\varphi < 1 - \gamma_2 \leq \gamma_2 - \frac{1}{2}$ ,*

$$K_t^{N_1}(x) = K_0^{N_1}(x) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') A_{x, x'}^{N_1} \pi(dx', dy) ds \quad (16)$$

*where  $K_0^{N_1}(x) = 0$  if  $\varphi < \gamma_2 - \frac{1}{2}$ , and  $K_0^{N_1}(x) = \mathcal{G}^{N_1}(x)$  if  $\varphi = \gamma_2 - \frac{1}{2}$ .*

*Case 2. When  $\gamma_2 \in [\frac{3}{4}, 1)$  and  $\varphi = 1 - \gamma_2$ ,*

$$\begin{aligned} & K_t^{N_1}(x) \\ &= K_0^{N_1}(x) + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left[ l_t^{N_1} (B_{x, x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_t^{N_1} (B_{x, x'}^{2, j}(\theta)) \right] \\ & \quad \pi(dx', dy) ds \\ &+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' l_t^{N_1} (B_{x, x'}^{3, j}(\theta)) \langle B_{x, x'}^{3, j}(\theta), \gamma_0^{N_1} \rangle \pi(dx', dy) ds \end{aligned} \quad (17)$$

$$\begin{aligned}
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle l_t^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \pi(dx', dy) ds \\
& - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') A_{x,x'}^{N_1} \pi(dx', dy) ds
\end{aligned}$$

where  $K_0^{N_1}(x) = 0$  if  $\gamma_2 \in (\frac{3}{4}, 1)$ ,  $K_0^{N_1}(x) = \mathcal{G}^{N_1}(x)$  if  $\gamma_2 = \frac{3}{4}$ , and  $l_t^{N_1}(f)$  is given by equation (15) for any  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ .

Notice that when  $\gamma_2 > 3/4$ , Theorem 2.7 shows that the limit of  $K_t^{N_1, N_2}$  is deterministic. This motivates us to consider the next order correction. Namely, let us define the second order fluctuations  $\Psi_t^{N_1, N_2} = N_2^{\zeta-\varphi}(K_t^{N_1, N_2} - K_t^{N_1})$  for  $\gamma_2 \in (3/4, 1)$  and for some  $\zeta > \varphi$  to be determined.

**Proposition 2.8.** *Let Assumption 2.1 hold and choose the learning rates via (6). Fix  $\gamma_2 \in (3/4, 1)$ ,  $\varphi = 1 - \gamma_2$ , and  $f \in C_b^3(\mathbb{R}^{1+N_1(1+d)})$ . Letting  $\zeta \leq 2\varphi$ , the process  $\{L_t^{N_1, N_2}(f) = N_2^{\zeta-\varphi}[l_t^{N_1, N_2}(f) - l_t^{N_1}(f)], t \in [0, T]\}_{N_2 \in \mathbb{N}}$  converges in probability in the space  $D_{\mathbb{R}}([0, T])$  as  $N_2 \rightarrow \infty$ , and*

*Case 1. If  $\zeta < 2\varphi = 2 - 2\gamma_2$ ,  $L_t^{N_1, N_2}(f) \rightarrow 0$ .*

*Case 2. If  $\zeta = 2\varphi = 2 - 2\gamma_2$ ,  $L_t^{N_1, N_2}(f) \rightarrow L_t^{N_1}(f)$ , where  $L_t^{N_1}(f)$  is given by*

$$\begin{aligned}
& L_t^{N_1}(f) \\
& = \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) l_s^{N_1} \left( \partial_c f(\theta) \sigma(Z^{2, N_1}(x')) \right) \pi(dx', dy) ds \\
& + \frac{1}{N_1^{1-\gamma_1}} \\
& \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) l_s^{N_1} \left( c \sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta) \right) \pi(dx', dy) ds \\
& + \frac{1}{N_1^{1-\gamma_1}} \\
& \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) l_s^{N_1} \left( \left\langle c \sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x' \right) \pi(dx', dy) ds \\
& + \frac{1}{N_1^{1-\gamma_1}} \\
& \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle l_s^{N_1} \left( c \sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2 \right) \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
& - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left\langle C_{x'}^{N_1, f}(c, w), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds
\end{aligned} \tag{18}$$

**Theorem 2.9.** *Let Assumption 2.1 hold and choose the learning rates via (6). Let also  $\mathcal{G}^{N_1}(x)$  be the Gaussian random variable defined in (9). Then, for fixed  $\gamma_2 \in (3/4, 1)$  and  $\varphi = 1 - \gamma_2$ , the sequence of processes  $\{\Psi_t^{N_1, N_2}, t \in [0, T]\}_{N_2 \in \mathbb{N}}$  converges in distribution in the space  $D_{\mathbb{R}^M}([0, T])$  to  $\Psi_t^{N_1}$ , which satisfies the following evolution equations, depending on the values of  $\gamma_2$  and  $\zeta$ :*

*Case 1. When  $\gamma_2 \in (\frac{3}{4}, \frac{5}{6})$  and  $\zeta \leq \gamma_2 - \frac{1}{2}$ , or when  $\gamma_2 \in [\frac{5}{6}, 1)$  and  $\zeta < 2 - 2\gamma_2 \leq \gamma_2 - \frac{1}{2}$ ,*

$$\Psi_t^{N_1}(x) = \Psi_0^{N_1}(x) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1}(x') A_{x,x'}^{N_1} \pi(dx', dy) ds, \tag{19}$$

where  $\Psi_0^{N_1}(x) = 0$  if  $\zeta < \gamma_2 - \frac{1}{2}$ , and  $\Psi_0^{N_1}(x) = \mathcal{G}^{N_1}(x)$  if  $\zeta = \gamma_2 - \frac{1}{2}$ .

Case 2. When  $\gamma_2 \in [\frac{5}{6}, 1)$  and  $\zeta = 2 - 2\gamma_2$ ,

$$\begin{aligned}
\Psi_t^{N_1}(x) &= \Psi_0^{N_1}(x) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1}(x') A_{x,x'}^{N_1} \pi(dx', dy) ds \\
&\quad + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left[ L_s^{N_1}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1}((B_{x,x'}^{2,j}(\theta))) \right] \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) L_s^{N_1}(B_x^{3,j}(\theta)) \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle x x' B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle L_s^{N_1}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
&\quad - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left[ l_s^{N_1}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_s^{N_1}(B_{x,x'}^{2,j}(\theta)) \right] \pi(dx', dy) ds \\
&\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' l_s^{N_1}(B_x^{3,j}(\theta)) \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle l_s^{N_1}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds.
\end{aligned} \tag{20}$$

where  $\Psi_0^{N_1}(x) = 0$  if  $\gamma_2 \in (\frac{5}{6}, 1)$ ,  $\Psi_0^{N_1}(x) = \mathcal{G}^{N_1}(x)$  if  $\gamma_2 = \frac{5}{6}$ ,  $K_s^{N_1}$  satisfies equation (17), and  $L_s^{N_1}$  satisfies (18).

These results suggest that there is an expansion of  $\left\langle f, \gamma_t^{N_1, N_2} \right\rangle$  and  $h_t^{N_1, N_2}$  as  $N_2 \rightarrow \infty$  for all  $\gamma_2 \in \left[ \frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2} \right]$  with  $\nu \in \mathbb{N}$ . The aforementioned results obtain the leading order of such expansions when  $\nu = 1$  and  $\nu = 2$ . In Appendix E we obtain the leading order of such asymptotic expansions for all  $\nu \in \mathbb{N}$  and as a consequence for all  $\gamma_2 \in (1/2, 1)$  using an inductive argument.

In particular, when  $\gamma_2 \in \left[ \frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2} \right)$ , we obtain that for any fixed  $f \in C_b^\infty(\mathbb{R}^{1+N_1(1+d)})$ , as  $N_2 \rightarrow \infty$ ,

$$\left\langle f, \gamma_t^{N_1, N_2} \right\rangle \approx \sum_{n=0}^{\nu-1} \frac{1}{N_2^{n(1-\gamma_2)}} l_{n,t}^{N_1}(f) + \text{lower order terms in } N_2, \tag{21}$$

where we have identified  $l_{0,t}^{N_1}(f) = \left\langle f, \gamma_0^{N_1} \right\rangle$ ,  $l_{1,t}^{N_1}(f) = l_t^{N_1}(f)$ ,  $l_{2,t}^{N_1}(f) = L_t^{N_1}(f)$ . When  $\nu \geq 3$ , the inductive expressions for  $l_{n,t}^{N_1}(f)$  are given in (61).

As  $N_2 \rightarrow \infty$  and when  $\gamma_2 \in \left( \frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2} \right]$ , we have the asymptotic expansion

$$h_t^{N_1, N_2}(x) \approx \sum_{n=0}^{\nu-1} \frac{1}{N_2^{n(1-\gamma_2)}} Q_{j,t}^{N_1}(x) + \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} Q_{\nu,t}^{N_1}(x) + \text{lower order terms in } N_2, \tag{22}$$

where  $Q_{0,t}^{N_1} = h_t^{N_1}$ ,  $Q_{1,t}^{N_1} = K_t^{N_1}$ ,  $Q_{2,t}^{N_1} = \Psi_t^{N_1}$ . For  $n = 1, \dots, \nu - 1$ ,  $Q_{n,t}^{N_1}$  satisfy the deterministic evolution equations (63), (64) and (65). We do not show this here, but for fixed  $j \in \mathbb{N}$ , one can also show exactly as in [45] that, under Assumptions 2.1 and 2.5,  $Q_{j,t}^{N_1} \rightarrow 0$  exponentially fast as  $t \rightarrow \infty$ .

For the sake of presentation and due to the length of the formulas we present the associated formulas on the right hand side of these expansions (and their derivation) in Appendix E.

**Remark 2.10.** Examining the proofs in Appendices A-E makes it clear that the convergence theorems of this paper also hold if we assume that the sequence of data points  $(x^{(i)}, y^{(i)})$  is i.i.d. sampled from a probability distribution  $\pi(dx, dy)$  that is compactly supported. We make the somewhat more restrictive assumption in part (ii) of Assumption 2.1 for brevity and uniformity of presentation.

**3. Numerical studies.** The goal of this section is to compare the numerical performance of two and three-layer neural networks of the form (1) for different values of  $\gamma_i \in [1/2, 1]$ . In Section 2, we demonstrated the neural network's output statistical properties can be approximated via the limit to  $\infty$  of the hidden layers of the outer layer. This analysis showed that the variance of the neural network's output is minimized when the outer layer is in the mean-field scaling ( $\gamma_2 = 1$  in the case of (1)) while the scaling of the inner layer i.e. the value of  $\gamma_1$ , plays a less prominent role.

In this section we demonstrate a number of numerical studies to compare test accuracy for two and three layer neural networks for different values of the normalization parameters. Our numerical studies involve the well known MNIST [25] data sets. The MNIST dataset [25], which includes 70,000 images of handwritten integers from 0 to 9. For the two layer network case, the learning rates satisfy (6), as suggested by our theoretical analysis. The neural networks are trained to identify the handwritten numbers using the image pixels as an input. In the MNIST dataset, each image has 784 pixels, 60,000 images are used as train images and 10,000 images are test images.

We find numerically that test accuracy of the fitted neural networks increases monotonically in  $\gamma_2 \in [1/2, 1]$ , suggesting that the mean-field normalization  $1/N_2$  for the outer layer that corresponds to  $\gamma_2 = 1$ , has certain advantages over scalings  $1/N_2^{\gamma_2}$  for  $\gamma_2 \in [1/2, 1)$  when it comes to test accuracy. The numerical studies in both the two and the three layer neural networks demonstrate that as long as the outer layer is scaled in the mean-field scaling, the scalings of the inner layers plays a less prominent role. With that being said, the optimal choice, as seen by these numerical studies, is to scale all layers in the mean-field scaling.

Before presenting the numerical studies, we remark that even though the mathematical analysis of the paper is carried over for the mean square error loss (for mathematical convenience), we use the cross-entropy loss for the simulation studies as this loss function is more appropriate for the image dataset and classification problem at hand. As was demonstrated in [45] in the shallow case this is done without loss of generality.

**3.1. Numerical results for the two layer case.** In this subsection we fit the model (1) to the MNIST dataset and we compare the effect of different values of  $\gamma_1, \gamma_2$ .

In Figure 1, we fix in each sub-figure the value of  $\gamma_2$  and plot test accuracy curves with respect to values of  $\gamma_1$ . We find that for each  $\gamma_2$ , after an initial phase, the

behavior is monotonic with respect to  $\gamma_1$ . We also find that the best behavior is when  $\gamma_1 = \gamma_2 = 1$  with the neural network's behavior being more sensitive on the choice of the value for  $\gamma_2$ .

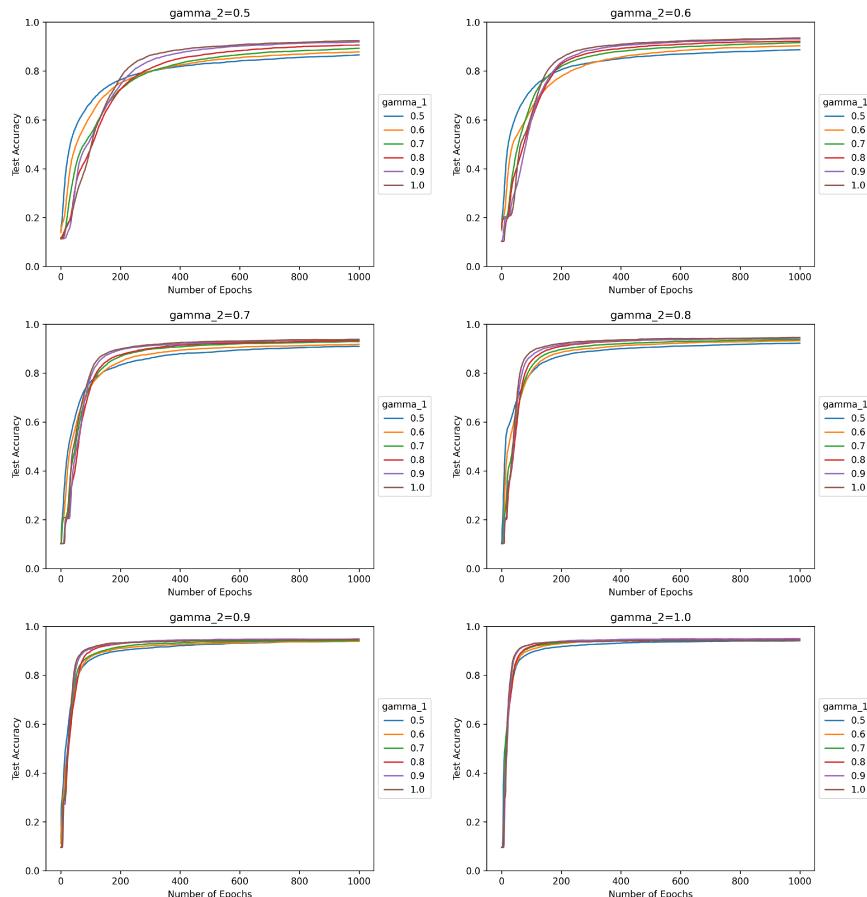


FIGURE 1. Performance of scaled neural networks on MNIST test dataset: cross entropy loss,  $N_1 = N_2 = 100$ , batch size = 20, Number of Epoch = 1000. Each subfigure plots various  $\gamma_1$  for a fixed  $\gamma_2$ .

In Figure 2, we fix in each sub-figure the value of  $\gamma_1$  and plot test accuracy curves with respect to values of  $\gamma_2$ . We find that for each  $\gamma_1$  the test accuracy is clearly monotonic with respect to  $\gamma_2$ . Independently of the value of  $\gamma_1$ , the best test accuracy is obtained when  $\gamma_2 = 1$ .

In Figures 3 and 4 we illustrate the effect of unequal choices for  $N_1$  and  $N_2$ . We find that the best test accuracy is always when  $N_2 > N_1$ , which also motivates taking first  $N_2 \rightarrow \infty$  and then  $N_1 \rightarrow \infty$ . Also overall, best test accuracy is also when  $\gamma_1 = \gamma_2 = 1$ .

**3.2. The three layer neural network case.** The purpose of this section is to demonstrate that the same qualitative conclusions that hold for the two-layer case

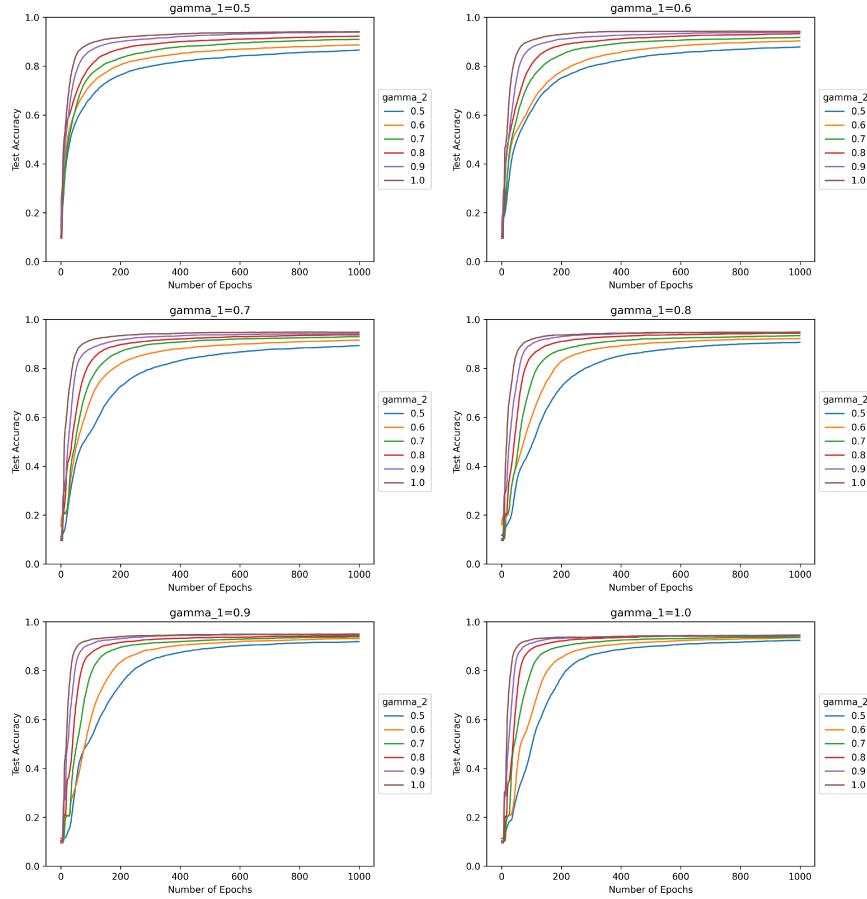


FIGURE 2. Performance of scaled neural networks on MNIST test dataset: cross entropy loss,  $N_1 = N_2 = 100$ , batch size = 20, Number of Epoch = 1000. Each subfigure plots various  $\gamma_2$  for a fixed  $\gamma_1$ .

also hold for neural networks with more layers. For this purpose, let us consider the following three-layer scaled neural networks:

$$g_{\theta}^{N_1, N_2, N_3}(x) = \frac{1}{N_3^{\gamma_3}} \sum_{i=1}^{N_3} C^i \sigma \left( \frac{1}{N_2^{\gamma_2}} \sum_{j=1}^{N_2} W^{3,i,j} \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{\nu=1}^{N_1} W^{2,j,\nu} \sigma(W^{1,\nu} x) \right) \right), \quad (23)$$

where  $C^i, W^{3,i,j}, W^{2,j,\nu} \in \mathbb{R}$ ,  $x, W^{1,\nu} \in \mathbb{R}^d$ , and  $\gamma_1, \gamma_2 \in [1/2, 1)$  are fixed scaling parameters. For convenience, we write  $W^{1,\nu} x = \langle W^{1,\nu}, x \rangle_{l^2}$  as the standard  $l^2$  inner product for the vectors. The neural network model has parameters

$$\theta = (C^1, \dots, C^{N_3}, W^{3,1,1}, \dots, W^{3,N_3,N_2}, W^{2,1,1}, \dots, W^{2,N_2,N_1}, W^{1,1}, \dots, W^{1,N_1}),$$

which are to be estimated from data  $(X, Y) \sim \pi(dx, dy)$ . We consider the loss function

$$L(\theta) = \frac{1}{2} \mathbb{E}_{X,Y} \left[ \left( Y - g_{\theta}^{N_1, N_2, N_3}(x) \right)^2 \right],$$

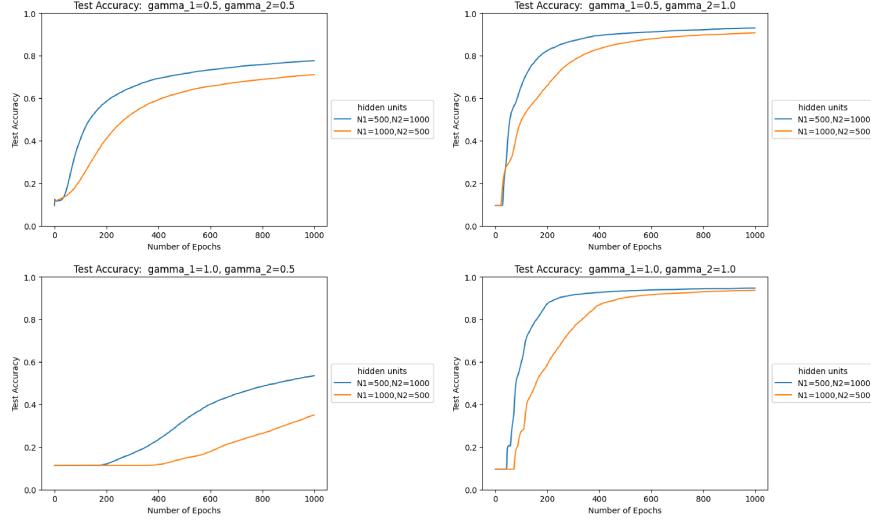


FIGURE 3. Performance of scaled neural networks on MNIST test dataset: cross entropy loss, batch size = 20, Number of Epoch = 1000. For each fixed sets of  $\gamma_1, \gamma_2$ , each subfigure compares the performances of models with different  $N_1, N_2$ .

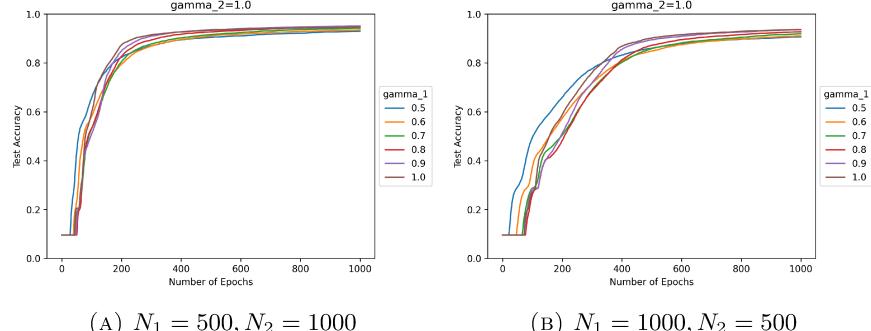


FIGURE 4. Performance of scaled neural networks on MNIST test dataset: cross entropy loss,  $\gamma_2 = 1.0$ , batch size = 20, Number of Epoch = 1000. Each subfigure plots for different sets of hidden units.

and the model parameters  $\theta$  are trained by the stochastic gradient descent algorithm, for  $k \in \mathbb{N}$ ,  $\nu = 1, \dots, N_1$ ,  $j = 1, \dots, N_2$  and  $i = 1, \dots, N_1$ ,

$$C_{k+1}^i = C_k^i + \frac{\alpha_c^{N_1, N_2, N_3}}{N_3^{\gamma_3}} (y_k - g_k^{N_1, N_2, N_3}(x_k)) H_k^{3,i}(x_k),$$

$$W_{k+1}^{1,\nu} = W_k^{1,\nu} + \frac{\alpha_{W,1}^{N_1, N_2, N_3}}{N_1^{\gamma_1}} (y_k - g_k^{N_1, N_2, N_3}(x_k)) \left( \frac{1}{N_3^{\gamma_3}} \sum_{i=1}^{N_3} C_k^i \sigma'(Z_k^{3,i}(x_k)) \left( \frac{1}{N_2^{\gamma_2}} \sum_{j=1}^{N_2} W_k^{3,i,j} \sigma'(Z^{2,j}(x_k)) W_k^{2,j,\nu} \right) \right)$$

$$\begin{aligned}
& \times \sigma'(W_k^{1,\nu} x_k) x_k, \\
W_{k+1}^{2,i,\nu} &= W_k^{2,i,\nu} + \frac{\alpha_{W,2}^{N_1, N_2, N_3}}{N_1^{\gamma_1} N_2^{\gamma_2}} \left( y_k - g_k^{N_1, N_2, N_3}(x_k) \right) \\
& \quad \frac{1}{N_3^{\gamma_3}} \sum_{i=1}^{N_3} C_k^i \sigma'(Z_k^{3,i}(x_k)) W_k^{3,i,j} \sigma'(Z_k^{2,j}(x_k)) H_k^{1,\nu}(x_k), \\
W_{k+1}^{3,i,j} &= W_k^{3,i,j} + \frac{\alpha_{W,3}^{N_1, N_2, N_3}}{N_2^{\gamma_2} N_3^{\gamma_3}} \left( y_k - g_k^{N_1, N_2, N_3}(x_k) \right) C_k^i \sigma'(Z_k^{3,i}(x_k)) H_k^{2,j}(x_k),
\end{aligned}$$

where

$$\begin{aligned}
H_k^{1,\nu}(x) &= \sigma(W_k^{1,\nu} x), \\
Z_k^{2,j}(x) &= \frac{1}{N_1^{\gamma_1}} \sum_{\nu=1}^{N_1} W_k^{2,j,\nu} H_k^{1,\nu}(x), \\
H_k^{2,j}(x) &= \sigma(Z_k^{2,j}(x)), \\
Z_k^{3,i}(x) &= \frac{1}{N_2^{\gamma_2}} \sum_{j=1}^{N_2} W_k^{3,i,j} H_k^{2,j}(x_k), \\
H_k^{3,i}(x) &= \sigma(Z_k^{3,i}(x)), \\
g_k^{N_1, N_2, N_3}(x) &= g_{\theta_k}^{N_1, N_2, N_3}(x) = \frac{1}{N_3^{\gamma_3}} \sum_{i=1}^{N_3} C_k^i H_k^{3,i}(x).
\end{aligned}$$

We investigate the numerical performance of the neural network (23) trained by the SGD algorithm (24) with various  $\gamma_1, \gamma_2, \gamma_3, N_1, N_2$  and  $N_3$ . Even though we do not show this here, following the mathematical analysis that led to the choice of the learning rates (6), we get that in the three layer case the learning rates should be given as follows

$$\begin{aligned}
\alpha_C^{N_1, N_2, N_3} &= \frac{1}{N_3^{2-2\gamma_3}}, & \alpha_{W,1}^{N_1, N_2, N_3} &= \frac{1}{N_1^{1-2\gamma_1} N_2^{2-2\gamma_2} N_3^{3-2\gamma_3}}, \\
\alpha_{W,2}^{N_1, N_2, N_3} &= \frac{1}{N_1^{1-2\gamma_1} N_2^{1-2\gamma_2} N_3^{3-2\gamma_3}}, & \alpha_{W,3}^{N_1, N_2, N_3} &= \frac{1}{N_2^{1-2\gamma_2} N_3^{2-2\gamma_3}}
\end{aligned} \tag{24}$$

Let us now investigate numerically the performance of neural networks scaled by  $1/N_1^{\gamma_1}$ ,  $1/N_2^{\gamma_2}$  and  $1/N_3^{\gamma_3}$  with  $\gamma_1, \gamma_2, \gamma_3 \in [1/2, 1]$ . The numerical studies are again on the MNIST data set.

In Figure 5 we fix in each sub-figure the value of  $\gamma_3$  and vary the values of  $\gamma_1, \gamma_2$ . We find that the best results in terms of test accuracy are when  $\gamma_i = 1$  for all  $i$ . Importantly, we also find that the neural network's test accuracy is more sensitive on the choice of the outer layer normalization, i.e., on  $\gamma_3$ .

In Figure 6 we fix in each sub-figure the value of  $\gamma_2$  and vary the values of  $\gamma_1, \gamma_3$ . We find that the best results in terms of test accuracy are when  $\gamma_i = 1$  for all  $i$ . Again, we find that the neural network is more sensitive on the choice for  $\gamma_3$ .

In Figure 7 we fix in each sub-figure the value of  $\gamma_1$  and vary the values of  $\gamma_2, \gamma_3$ . The conclusions are the same as before. Namely, the best results in terms of test accuracy are when  $\gamma_i = 1$  for all  $i$ . Again, we find that if  $\gamma_3 = 1$ , then the neural network behavior is less sensitive on the choice of  $\gamma_1, \gamma_2$ .

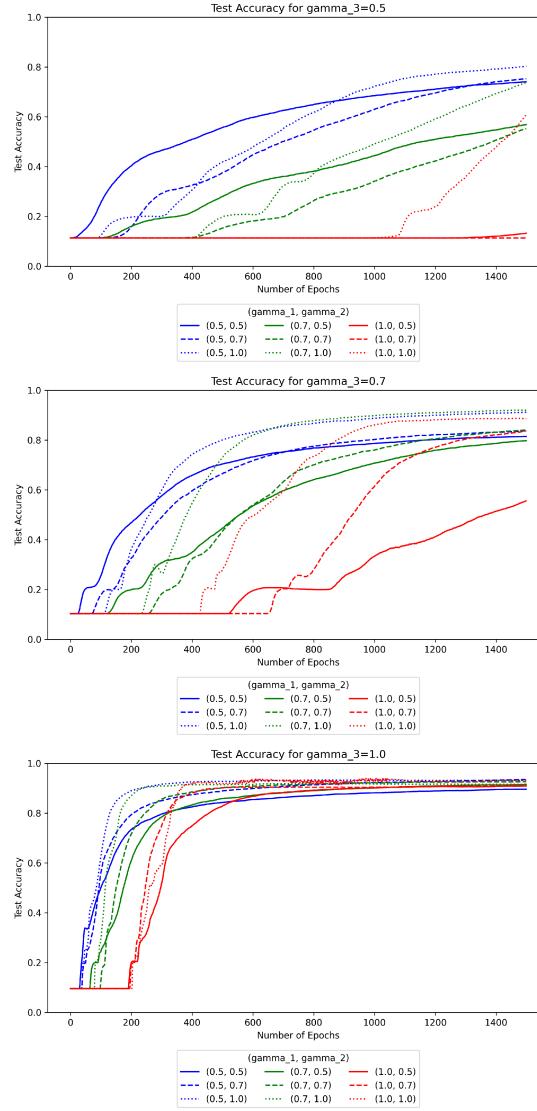


FIGURE 5. Performance of scaled neural networks on MNIST test dataset: cross entropy loss,  $N_1 = N_2 = N_3 = 100$ , batch size = 20, Number of Epoch = 1500. Each subfigure plots various  $\gamma_1, \gamma_2$  for a fixed  $\gamma_3$ .

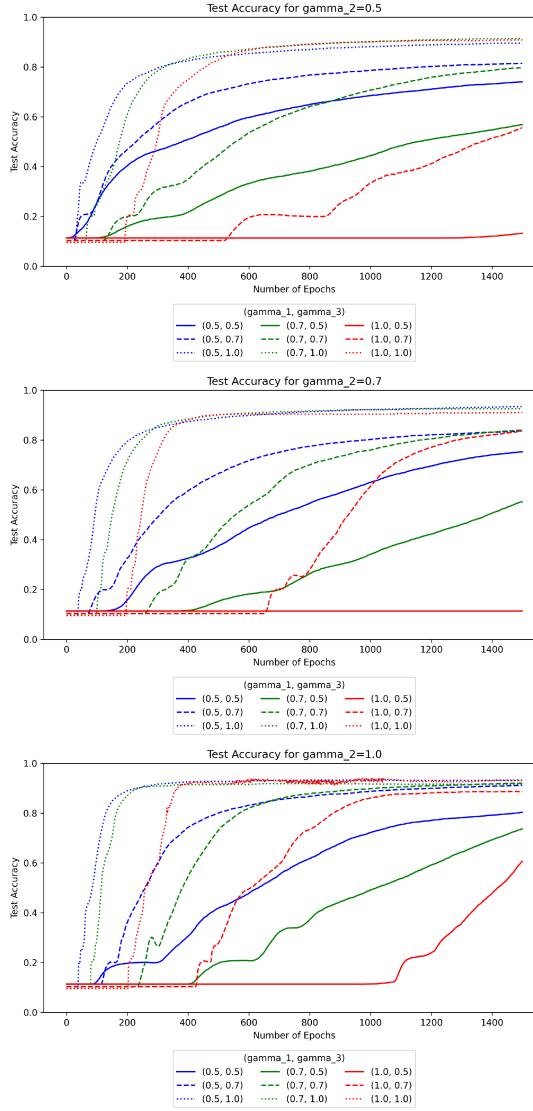


FIGURE 6. Performance of scaled neural networks on MNIST test dataset: cross entropy loss,  $N_1 = N_2 = N_3 = 100$ , batch size = 20, Number of Epoch = 1500. Each subfigure plots various  $\gamma_1, \gamma_3$  for a fixed  $\gamma_2$ .

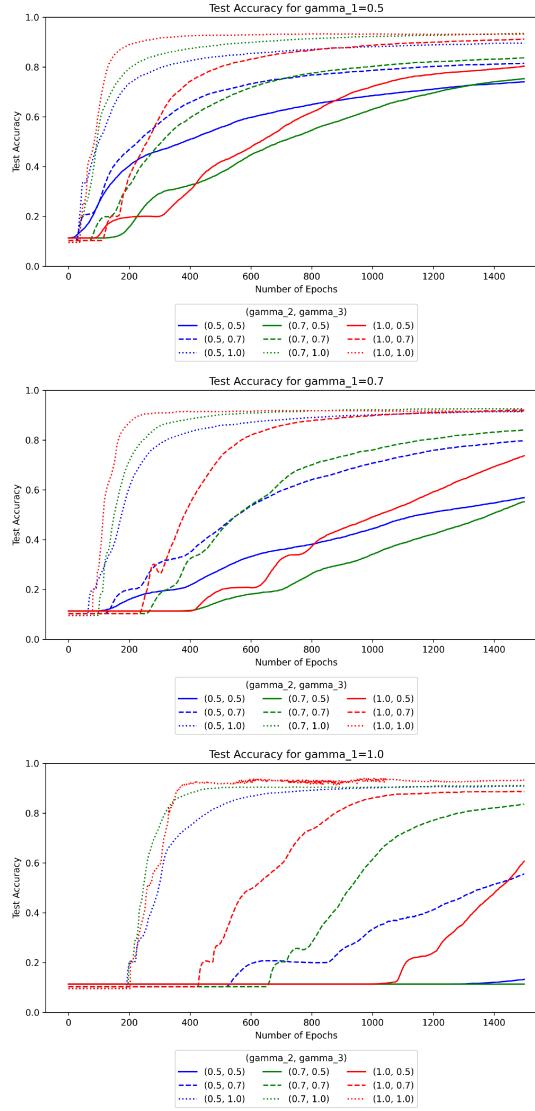


FIGURE 7. Performance of scaled neural networks on MNIST test dataset: cross entropy loss,  $N_1 = N_2 = N_3 = 100$ , batch size = 20, Number of Epoch = 1500. Each subfigure plots various  $\gamma_2, \gamma_3$  for a fixed  $\gamma_1$ .

#### 4. Learning rates definitions for deep neural networks of arbitrary depth.

Let us consider a typical deep feed-forward neural network that has depth  $m \in \mathbb{N}$  with  $\hat{\gamma} = (\gamma_1, \dots, \gamma_m) \in [1/2, 1]^{\otimes m}$  scalings that is defined inductively as follows

$$\begin{aligned}
g_{\theta}^{N_1, N_2, \dots, N_m}(x) &= \frac{1}{N_m^{\gamma_m}} \sum_{i_m=1}^{N_m} W^{N_m, i_m} \sigma_{i_m} \left( g_{\theta}^{N_1, \dots, N_{m-1}, i_m}(x) \right) \\
g_{\theta}^{N_1, \dots, N_{m-j}, i_{m-(j-1)}}(x) &= \frac{1}{N_{m-j}^{\gamma_{m-j}}} \sum_{i_{m-j}=1}^{N_{m-j}} W^{N_{m-j}, i_{m-j}, i_{m-(j-1)}} \sigma_{i_{m-j}} \\
&\quad \left( g_{\theta}^{N_1, \dots, N_{m-(j+1)}, i_{m-j}}(x) \right), j = 1, \dots, m-2 \\
&\quad \vdots \\
g_{\theta}^{N_1, i_2}(x) &= \frac{1}{N_1^{\gamma_1}} \sum_{i_1=1}^{N_1} W^{N_1, i_1, i_2} \sigma_{i_1} \left( g_{\theta}^{N_0, i_1}(x) \right) \\
g_{\theta}^{N_0, i_1}(x) &= \sigma_{i_0} \left( W^{N_0, i_1} x \right).
\end{aligned}$$

Even though  $N_0 = 1$  is redundant, we write it for notational consistency purposes.

The goal of this section is to provide the formulas for the choices of the learning rates as functions of  $N_i$  and  $\gamma_i$  for  $i = 1 \dots m$  so that in the end the neural network will be expected to converge to a well defined limit as  $N_i \rightarrow \infty$ .

We do not repeat the lengthy calculations here, but rather we only provide the formulas for the appropriate choice of the learning rate and leave the rest of the details to the interested reader. In the end, following the exact same procedure as in the two-layer and three-layer case, we obtain that the learning rates should be chosen according to the rules:

$$\begin{aligned}
a_{W^{N_m}} &= N_m^{2\gamma_m - 2} \\
a_{W^{N_{m-1}}} &= N_m^{2\gamma_m - 2} N_{m-1}^{2\gamma_{m-1} - 1} \\
a_{W^{N_{m-2}}} &= N_m^{2\gamma_m - 3} N_{m-1}^{2\gamma_{m-1} - 1} N_{m-2}^{2\gamma_{m-2} - 1} \\
a_{W^{N_{m-3}}} &= N_m^{2\gamma_m - 3} N_{m-1}^{2\gamma_{m-1} - 2} N_{m-2}^{2\gamma_{m-2} - 1} N_{m-3}^{2\gamma_{m-3} - 1} \\
a_{W^{N_{m-4}}} &= N_m^{2\gamma_m - 3} N_{m-1}^{2\gamma_{m-1} - 2} N_{m-2}^{2\gamma_{m-2} - 2} N_{m-3}^{2\gamma_{m-3} - 1} N_{m-4}^{2\gamma_{m-4} - 1} \\
&\quad \vdots \\
a_{W^{N_1}} &= N_m^{2\gamma_m - 3} N_{m-1}^{2\gamma_{m-1} - 2} N_{m-2}^{2\gamma_{m-2} - 2} N_{m-3}^{2\gamma_{m-3} - 2} \dots N_3^{2\gamma_3 - 2} N_2^{2\gamma_2 - 1} N_1^{2\gamma_1 - 1} \\
a_{W^{N_0}} &= N_m^{2\gamma_m - 3} N_{m-1}^{2\gamma_{m-1} - 2} N_{m-2}^{2\gamma_{m-2} - 2} N_{m-3}^{2\gamma_{m-3} - 2} \dots N_3^{2\gamma_3 - 2} N_2^{2\gamma_2 - 2} N_1^{2\gamma_1 - 1}.
\end{aligned}$$

Such a choice directly generalizes the formulas for the learning rates in the two and three layer case presented before and one can show that lead to formulas of the same type as those obtained in Section 2.

**5. Conclusions.** In this work, we have investigated the effect of layer normalization on the statistical behavior and test accuracy of deep neural networks. We have looked at all the scaling regimes between the square root normalization, i.e., the so-called Xavier normalization, see [13], all the way up to the mean-field normalization [7, 30, 35, 39, 40, 41]. Our two key findings are that (a): the mean field normalization leads to lower variance of the neural's network statistical output and better

test accuracy, and (b): given that the outer layer's normalization is the mean-field regime, the subsequent choice for the normalization of the inner layers does not affect test accuracy as much (mean field normalization remains the optimal choice, but there is less sensitivity in the inner layers). An important by-product of the mathematical analysis of this paper is a mathematically motivated way to define the learning rates. This is an important conclusion of our work since it gives a principled way to choose the related hyperparameters.

**Appendix A. A-priori bound for the parameters.** By specifying the learning rates  $\alpha_C^{N_1, N_2}, \alpha_{W,1}^{N_1, N_2}, \alpha_{W,2}^{N_1, N_2}$  as in (6), we can establish an important uniform bound for the parameters.

**Lemma A.1.** *For  $k = 0, 1, \dots, \lfloor TN_2 \rfloor$ ,  $i = 1, \dots, N_1$ , and  $j = 1, \dots, N_2$ , there exist a finite constant  $K > 0$  such that*

$$|C_k^i| + \|W_k^{1,j}\| + |W_k^{2,j,i}| < K.$$

Furthermore, as  $N_1, N_2$  grow

$$\begin{aligned} |C_{k+1}^i - C_k^i| &= O(N_2^{-1}), \\ \|W_{k+1}^{1,j} - W_k^{1,j}\| &= O(N_1^{-(1-\gamma_1)} N_2^{-1}), \\ |W_{k+1}^{2,j,i} - W_k^{2,j,i}| &= O(N_1^{-(1-\gamma_1)} N_2^{-1}). \end{aligned}$$

*Proof.* In this proof, we use  $K, K_1$  to represent unimportant constants that may change from line to line. We first establish a bound on  $C_k^i$ . For  $k = 0, 1, \dots, \lfloor TN_2 \rfloor$ , since  $\sigma(\cdot)$  is bounded,  $|H_k^{2,i}| < K$ , and by (2), we have

$$\begin{aligned} |C_{k+1}^i| &\leq |C_k^i| + \frac{\alpha_C^{N_1, N_2}}{N_2^{\gamma_2}} K \left[ K_1 + \frac{1}{N_2^{\gamma_2}} \sum_{m=1}^{N_2} |C_k^m| \right] \\ &\leq |C_k^i| + K \left[ \frac{\alpha_C^{N_1, N_2}}{N_2^{\gamma_2}} + \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2}} \sum_{m=1}^{N_2} |C_k^m| \right]. \end{aligned}$$

Since also

$$|C_k^i| = |C_0^i| + \sum_{j=1}^k (|C_j^i| - |C_{j-1}^i|) \leq |C_0^i| + K \frac{\alpha_C^{N_1, N_2}}{N_2^{\gamma_2-1}} + K \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2}} \sum_{j=1}^k \sum_{m=1}^{N_2} |C_{j-1}^m|,$$

we have

$$m_k^{N_2} \leq b^{N_2} + K \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-1}} \sum_{j=1}^k m_{j-1}^{N_2} = b_2^{N_2} + K \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-1}} \sum_{j=0}^{k-1} m_j^{N_2},$$

where

$$b^{N_2} = \frac{1}{N_2} \sum_{i=1}^{N_2} |C_0^i| + K \frac{\alpha_C^{N_1, N_2} N_2}{N_2^{\gamma_2}}, \quad m_k^{N_2} = \frac{1}{N_2} \sum_{i=1}^{N_2} |C_k^i|.$$

By the discrete Gronwall lemma and  $k \leq \lfloor TN_2 \rfloor$ ,

$$m_k^{N_2} \leq b^{N_2} \exp \left( K \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-2}} \right).$$

Thus, since  $C_0^i$  has compact support,

$$|C_k^i| \leq |C_0^i| + K \frac{\alpha_C^{N_1, N_2}}{N_2^{\gamma_2-1}} + K \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-2}} \left[ b^{N_2} \exp \left( K \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-2}} \right) \right]$$

is bounded if  $\alpha_C^{N_1, N_2} \leq 1/(N_2^{2-2\gamma_2})$ , for  $\gamma_2 \in [1/2, 1)$ .

Next, let's address the parameters  $W_k^{2,j,i}$ . By equation (2) and the boundedness of  $\sigma(\cdot), \sigma'(\cdot), H_k^{1,j}, C_k^i, m_k^{N_2}$ , we have

$$\begin{aligned} |W_{k+1}^{2,j,i}| &\leq |W_k^{2,j,i}| + \frac{\alpha_{W,2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{\gamma_2}} \left( |y_k| + \frac{K}{N_2^{\gamma_2}} \sum_{m=1}^{N_2} |C_k^m| \right) |C_k^i \sigma'(Z_k^{2,i}) H_k^{1,j}| \\ &\leq |W_k^{2,j,i}| + K \frac{\alpha_{W,2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{\gamma_2}} \left( |y_k| + \frac{1}{N_2^{\gamma_2}} \sum_{m=1}^{N_2} |C_k^m| \right) \\ &\leq |W_k^{2,j,i}| + K \frac{\alpha_{W,2}^{N_1, N_2} N_2^{1-\gamma_2}}{N_1^{\gamma_1} N_2^{\gamma_2}} \left( \frac{1}{N_2^{1-\gamma_2}} |y_k| + \frac{1}{N_2} \sum_{m=1}^{N_2} |C_k^m| \right) \\ &\leq |W_k^{2,j,i}| + K \frac{\alpha_{W,2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{2\gamma_2-1}}. \end{aligned}$$

Since  $k \leq \lfloor TN_2 \rfloor$  and  $W_0^{2,j,i}$  has compact support, we have

$$\begin{aligned} |W_k^{2,j,i}| &\leq |W_0^{2,j,i}| + \sum_{m=1}^k \left( |W_m^{2,j,i}| - |W_{m-1}^{2,j,i}| \right) \\ &\leq |W_0^{2,j,i}| + \sum_{m=1}^k K \frac{\alpha_{W,2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{2\gamma_2-1}} \\ &\leq K_1 + K \frac{\alpha_{W,2}^{N_1, N_2} N_2}{N_1^{\gamma_1} N_2^{2\gamma_2-1}}, \end{aligned}$$

which is bounded if  $\alpha_{W,2}^{N_1, N_2} \leq N_1^{\gamma_1} / N_2^{2-2\gamma_2}$ .

Lastly, for  $W_k^{1,j}$ , we have

$$\begin{aligned} \|W_{k+1}^{1,j}\| &\leq \|W_k^{1,j}\| + \frac{\alpha_{W,1}^{N_1, N_2}}{N_1^{\gamma_1}} \left| y_k - \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_k^i H_k^{2,i}(x_k) \right| \\ &\quad \left| \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_k^i \sigma'(Z_k^{2,i}) W_k^{2,j,i} \right| \|\sigma'(W_k^{1,j} x_k) x_k\| \\ &\leq \|W_k^{1,j}\| + K \frac{\alpha_{W,1}^{N_1, N_2} N_2^{2-2\gamma_2}}{N_1^{\gamma_1}} \left( \frac{1}{N_2^{1-\gamma_2}} |y_k| + \frac{1}{N_2} \sum_{i=1}^{N_2} |C_k^i| \right) \left( \frac{1}{N_2} \sum_{i=1}^{N_2} |C_k^i| \right) \\ &\leq \|W_k^{1,j}\| + K \frac{\alpha_{W,1}^{N_1, N_2} N_2^{2-2\gamma_2}}{N_1^{\gamma_1}}. \end{aligned}$$

Hence, for  $k \leq \lfloor TN_2 \rfloor$ ,

$$\begin{aligned}
\|W_k^{1,j}\| &\leq \|W_0^{1,j}\| + \sum_{m=1}^k \left( \|W_m^{1,j}\| - \|W_{m-1}^{1,j}\| \right) \\
&\leq \|W_0^{1,j}\| + \sum_{m=1}^k K \frac{\alpha_{W,1}^{N_1, N_2} N_2^{2-2\gamma_2}}{N_1^{\gamma_1}} \\
&\leq \|W_0^{1,j}\| + K \frac{\alpha_{W,1}^{N_1, N_2} N_2^{3-2\gamma_2}}{N_1^{\gamma_1}},
\end{aligned}$$

which is bounded since  $W_0^{1,j}$  has compact support and if  $\alpha_{W,1}^{N_1, N_2} \leq N_1^{\gamma_1}/N_2^{3-2\gamma_2}$ .

Collecting our results, for all  $k \leq \lfloor TN_2 \rfloor$  and  $i = 1, \dots, N_2$ , we have the desired uniform bound for the parameters.  $\square$

## Appendix B. Proof of Theorem 2.4.

**B.1. Evolution of the Pre-limit process.** We first analyze the evolution of the network output  $g_k^{N_1, N_2}(x)$ . Using Taylor expansion, we have

$$\begin{aligned}
g_{k+1}^{N_1, N_2}(x) - g_k^{N_1, N_2}(x) &= \\
&= \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_{k+1}^i \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_{k+1}^{2,j,i} \sigma(W_{k+1}^{1,j} x) \right) \\
&\quad - \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_k^i \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \\
&= \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} (C_{k+1}^i - C_k^i) \left[ \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_{k+1}^{2,j,i} \sigma(W_{k+1}^{1,j} x) \right) \right] \\
&\quad + \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_k^i \left[ \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_{k+1}^{2,j,i} \sigma(W_{k+1}^{1,j} x) \right) - \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \right] \\
&= \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} (C_{k+1}^i - C_k^i) \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \\
&\quad + \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_k^i \left[ \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \sigma(W_k^{1,j} x) (W_{k+1}^{2,j,i} - W_k^{2,j,i}) \right. \\
&\quad \left. + \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma'(W_k^{1,j} x) (W_{k+1}^{1,j} - W_k^{1,j}) x \right] \\
&\quad + R^{N_1, N_2}
\end{aligned}$$

where  $R^{N_1, N_2} = R_1^{N_1, N_2} + R_2^{N_1, N_2}$ , and

$$\begin{aligned}
R_1^{N_1, N_2} &= \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} (C_{k+1}^i - C_k^i) \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \check{W}_k^{2,j,i} \sigma(\check{W}_k^{1,j} x) \right) \\
&\quad \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \sigma(\check{W}_k^{1,j} x) (W_{k+1}^{2,j,i} - W_k^{2,j,i}) \\
&\quad + \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} (C_{k+1}^i - C_k^i) \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \check{W}_k^{2,j,i} \sigma(\check{W}_k^{1,j} x) \right)
\end{aligned}$$

$$\frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \check{W}_k^{2,j,i} \sigma'(\check{W}_k^{1,j} x) (W_{k+1}^{1,j} - W_k^{1,j}) x,$$

$$\begin{aligned} R_2^{N_1, N_2} &= \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_i^i \left\{ \frac{1}{2} \sigma'' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma(\tilde{W}_k^{1,j} x) \right) \left[ \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \sigma(\tilde{W}_k^{1,j} x) (W_{k+1}^{1,j} - W_k^{1,j}) \right]^2 \right. \\ &\quad + \frac{1}{2} \sigma'' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma(\tilde{W}_k^{1,j} x) \right) \left[ \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma'(\tilde{W}_k^{1,j} x) (W_{k+1}^{1,j} - W_k^{1,j}) x \right]^2 \\ &\quad + \frac{1}{2} \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma(\tilde{W}_k^{1,j} x) \right) \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma''(\tilde{W}_k^{1,j} x) [(W_{k+1}^{1,j} - W_k^{1,j}) x]^2 \\ &\quad + \sigma'' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma(\tilde{W}_k^{1,j} x) \right) \left[ \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma'(\tilde{W}_k^{1,j} x) (W_{k+1}^{1,j} - W_k^{1,j}) x \right] \\ &\quad \cdot \left[ \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \sigma(\tilde{W}_k^{1,j} x) (W_{k+1}^{2,j,i} - W_k^{2,j,i}) \right] \\ &\quad + \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \tilde{W}_k^{2,j,i} \sigma(\tilde{W}_k^{1,j} x) \right) \\ &\quad \left. \left[ \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} \sigma(\tilde{W}_k^{1,j} x) (W_{k+1}^{2,j,i} - W_k^{2,j,i}) (W_{k+1}^{1,j} - W_k^{1,j}) x \right] \right\}, \end{aligned}$$

for some  $(\check{W}_k^{2,j,i}, \check{W}_k^{1,j}), (\tilde{W}_k^{2,j,i}, \tilde{W}_k^{1,j})$  in the line segments connecting  $(W_{k+1}^{2,j,i}, W_{k+1}^{1,j})$  and  $(W_k^{2,j,i}, W_k^{1,j})$ . By Lemma A.1,  $R^{N_1, N_2} = O(N_2^{-(1+\gamma_2)})$ . Using equation (2) and definitions of the stochastic gradient descent algorithm and the empirical measure, we have

$$\begin{aligned} &g_{k+1}^{N_1, N_2}(x) - g_k^{N_1, N_2}(x) \\ &= \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-1}} (y_k - g_k^{N_1, N_2}(x_k)) \left\langle \sigma(Z^{2, N_1}(x_k)) \sigma(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\ &\quad + \frac{\alpha_{W, 2}^{N_1, N_2}}{N_1^{2\gamma_1} N_2^{2\gamma_2-1}} \sum_{j=1}^{N_1} (y_k - g_k^{N_1, N_2}(x_k)) \\ &\quad \left\langle (c)^2 \sigma'(Z^{2, N_1}(x_k)) \sigma'(Z^{2, N_1}(x)) \sigma(w^{1,j} x_k) \sigma(w^{1,j} x), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\ &\quad + \frac{\alpha_{W, 1}^{N_1, N_2}}{N_1^{2\gamma_1} N_2^{2\gamma_2-2}} \\ &\quad \sum_{j=1}^{N_1} (y_k - g_k^{N_1, N_2}(x_k)) x x_k \left\langle c w^{2,j} \sigma'(w^{1,j} x) \sigma'(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \times \\ &\quad \times \left\langle c w^{2,j} \sigma'(w^{1,j} x_k) \sigma'(Z^{2, N_1}(x_k)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle + O(N_2^{-(1+\gamma_2)}), \end{aligned}$$

where  $x_k$  is the data point in the  $k^{th}$  step of the stochastic gradient descent algorithm defined in (2), and  $Z^{2, N_1}(x) = \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} w^{2,j} \sigma(w^{1,j} x)$ . We can then write the

evolution of  $h_t^{N_1, N_2}(x)$  for  $t \in [0, T]$  as

$$\begin{aligned}
h_t^{N_1, N_2}(x) - h_0^{N_1, N_2}(x) &= \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left[ g_{k+1}^{N_1, N_2}(x) - g_k^{N_1, N_2}(x) \right] \\
&= \frac{\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-1}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \sigma(Z^{2, N_1}(x_k)) \sigma(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad + \frac{\alpha_{W, 2}^{N_1, N_2}}{N_1^{2\gamma_1} N_2^{2\gamma_2-1}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \sum_{j=1}^{N_1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \\
&\quad \left\langle (c)^2 \sigma'(Z^{2, N_1}(x_k)) \sigma'(Z^{2, N_1}(x)) \sigma(w^{1, j} x_k) \sigma(w^{1, j} x), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad + \frac{\alpha_{W, 1}^{N_1, N_2}}{N_1^{2\gamma_1} N_2^{2\gamma_2-2}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \sum_{j=1}^{N_1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \\
&\quad x x_k \left\langle c w^{2, j} \sigma'(w^{1, j} x) \sigma'(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad \cdot \left\langle c w^{2, j} \sigma'(w^{1, j} x_k) \sigma'(Z^{2, N_1}(x_k)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad + O(N_2^{-\gamma_2}),
\end{aligned}$$

and, using now the definitions of the learning rates from (6) we continue the last display as

$$\begin{aligned}
&= \frac{1}{N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \left\langle \sigma(Z^{2, N_1}(x')) \sigma(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \\
&\quad + \frac{1}{N_1 N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \sum_{j=1}^{N_1} \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \times \\
&\quad \times \left\langle (c)^2 \sigma'(Z^{2, N_1}(x')) \sigma'(Z^{2, N_1}(x)) \sigma(w^{1, j} x') \sigma(w^{1, j} x), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \\
&\quad + \frac{1}{N_1 N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \sum_{j=1}^{N_1} \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) x x' \left\langle c w^{2, j} \sigma'(w^{1, j} x) \sigma'(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad \cdot \left\langle c w^{2, j} \sigma'(w^{1, j} x') \sigma'(Z^{2, N_1}(x')), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \\
&\quad + M_t^{N_1, N_2} + O(N_2^{-\gamma_2}),
\end{aligned}$$

where  $M_t^{N_1, N_2} = M_{1, t}^{N_1, N_2} + M_{2, t}^{N_1, N_2} + M_{3, t}^{N_1, N_2}$  is a martingale term given by

$$\begin{aligned}
M_{1, t}^{N_1, N_2} &= \frac{1}{N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left\{ \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \sigma(Z^{2, N_1}(x_k)) \sigma(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \right. \\
&\quad \left. - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \right. \\
&\quad \left. \left\langle \sigma(Z^{2, N_1}(x')) \sigma(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \right\}, \tag{25}
\end{aligned}$$

$$\begin{aligned}
M_{2,t}^{N_1, N_2} = & \frac{1}{N_1 N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \sum_{j=1}^{N_1} \left\{ \left( y_k - g_k^{N_1, N_2}(x_k) \right) \right. \\
& \left\langle (c)^2 \sigma' (Z^{2, N_1}(x_k)) \sigma' (Z^{2, N_1}(x)) \sigma(w^{1,j} x_k) \sigma(w^{1,j} x), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
& - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \\
& \left. \left\langle (c)^2 \sigma' (Z^{2, N_1}(x')) \sigma' (Z^{2, N_1}(x)) \sigma(w^{1,j} x') \sigma(w^{1,j} x), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \right\}, \\
\end{aligned} \tag{26}$$

$$\begin{aligned}
M_{3,t}^{N_1, N_2} = & \frac{1}{N_1 N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \sum_{j=1}^{N_1} \left\{ \left( y_k - g_k^{N_1, N_2}(x_k) \right) \right. \\
& x x_k \left\langle c w^{2,j} \sigma'(w^{1,j} x) \sigma'(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
& \cdot \left. \left\langle c w^{2,j} \sigma'(w^{1,j} x_k) \sigma'(Z^{2, N_1}(x_k)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \right. \\
& - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) x x' \left\langle c w^{2,j} \sigma'(w^{1,j} x) \sigma'(Z^{2, N_1}(x)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
& \cdot \left. \left\langle c w^{2,j} \sigma'(w^{1,j} x') \sigma'(Z^{2, N_1}(x')), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \right\} \\
\end{aligned} \tag{27}$$

Recall that learning rates are as given in (6). As  $N_2 \rightarrow \infty$ ,  $h_t^{N_1, N_2}$  can further be re-written in terms of Riemann integrals and the scaled empirical measure  $\gamma_t^{N_1, N_2}$ ,

$$\begin{aligned}
& h_t^{N_1, N_2}(x) - h_0^{N_1, N_2}(x) \\
&= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \left\langle \sigma(Z^{2, N_1}(x')) \sigma(Z^{2, N_1}(x)), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \\
& \left\langle (c)^2 \sigma'(Z^{2, N_1}(x')) \sigma'(Z^{2, N_1}(x)) \sigma(w^{1,j} x') \sigma(w^{1,j} x), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) x x' \left\langle c w^{2,j} \sigma'(w^{1,j} x) \sigma'(Z^{2, N_1}(x)), \gamma_s^{N_1, N_2} \right\rangle \\
& \cdot \left\langle c w^{2,j} \sigma'(w^{1,j} x') \sigma'(Z^{2, N_1}(x')), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&+ M_t^{N_1, N_2} + O(N_2^{-\gamma_2}). \\
\end{aligned} \tag{28}$$

Finally, we analyze the evolution of the empirical measure  $\tilde{\gamma}_k^{N_1, N_2}$  in terms of its pairing with test functions  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ . Denote  $\theta_k^i = (C_k^i, W_k^{2,1,i}, \dots, W_k^{2,N_1,i}, W_k^{1,1}, \dots, W_k^{1,N_1})$ , first order Taylor expansion gives

$$\begin{aligned}
& \left\langle f, \tilde{\gamma}_{k+1}^{N_1, N_2} \right\rangle - \left\langle f, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&= \frac{1}{N_2} \sum_{i=1}^{N_2} [f(\theta_{k+1}^i) - f(\theta_k^i)]
\end{aligned}$$

$$\begin{aligned}
&= \frac{1}{N_2} \sum_{i=1}^{N_2} \partial_c f(\theta_k^i) (C_{k+1}^i - C_k^i) + \frac{1}{N_2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \partial_{w^{2,j}} f(\theta_k^i) (W_{k+1}^{2,j,i} - W_k^{2,j,i}) \\
&\quad + \frac{1}{N_2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \nabla_{w^{1,j}} f(\theta_k^i) (W_{k+1}^{1,j} - W_k^{1,j}) + O\left(\frac{1}{N_2^2}\right)
\end{aligned}$$

Using (2), we have

$$\begin{aligned}
&\left\langle f, \tilde{\gamma}_{k+1}^{N_1, N_2} \right\rangle - \left\langle f, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&= \frac{\alpha_C^{N_1, N_2}}{N_2^{\gamma_2}} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x_k)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad + \frac{\alpha_{W, 2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{\gamma_2}} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma(w^1 x_k) \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad + \frac{\alpha_{W, 1}^{N_1, N_2}}{N_1^{\gamma_1}} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma'(w^1 x_k) w^2, N_2^{1-\gamma_2} \tilde{\gamma}_k^{N_1, N_2} \right\rangle \right. \\
&\quad \left. \cdot \nabla_{w^1} f(\theta) x_k, \tilde{\gamma}_k^{N_1, N_2} \right\rangle + O\left(\frac{1}{N_2^2}\right).
\end{aligned}$$

In order to write the evolution in terms of the scaled measure  $\gamma_t^{N_1, N_2}$ , for  $t \in [0, 1]$ , we have

$$\begin{aligned}
&\left\langle f, \gamma_t^{N_1, N_2} \right\rangle - \left\langle f, \gamma_0^{N_1, N_2} \right\rangle = \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left\langle f, \nu_{k+1}^{N_1, N_2} \right\rangle - \left\langle f, \nu_k^{N_1, N_2} \right\rangle \\
&= \frac{\alpha_C^{N_1, N_2}}{N_2^{\gamma_2}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x_k)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad + \frac{\alpha_{W, 2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{\gamma_2}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma(w^1 x_k) \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad + \frac{\alpha_{W, 1}^{N_1, N_2}}{N_1^{\gamma_1}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \\
&\quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma'(w^1 x_k) w^2, N_2^{1-\gamma_2} \tilde{\gamma}_k^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x_k, \tilde{\gamma}_k^{N_1, N_2} \right\rangle + O\left(\frac{1}{N_2}\right) \\
&= \frac{\alpha_C^{N_1, N_2}}{N_2^{\gamma_2-1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x)) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x)), \gamma_s^{N_1, N_2} \right\rangle \pi(dx, dy) ds \\
&\quad + \frac{\alpha_{W, 2}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{\gamma_2-1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x)) \\
&\quad \left\langle c\sigma'(Z^{2, N_1}(x)) \sigma(w^1 x) \cdot \partial_{w^2} f(\theta), \gamma_s^{N_1, N_2} \right\rangle \pi(dx, dy) ds \\
&\quad + \frac{\alpha_{W, 1}^{N_1, N_2}}{N_1^{\gamma_1} N_2^{\gamma_2-2}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x)) \\
&\quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x)) \sigma'(w^1 x) w^2, \gamma_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x, \gamma_s^{N_1, N_2} \right\rangle \pi(dx, dy) ds \\
&\quad + M_{f, t}^{N_1, N_2} + O\left(\frac{1}{N_2}\right),
\end{aligned} \tag{29}$$

where  $M_{f,t}^{N_1,N_2} = M_{f,1,t}^{N_1,N_2} + M_{f,2,t}^{N_1,N_2} + M_{f,3,t}^{N_1,N_2}$  is a martingale term, and

$$\begin{aligned}
& M_{f,1,t}^{N_1,N_2} \\
&= \frac{\alpha_C^{N_1,N_2}}{N_2^{\gamma_2}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left\{ \left( y_k - g_k^{N_1,N_2}(x_k) \right) \left\langle \partial_c f(\theta) \sigma(Z^{2,N_1}(x_k)), \tilde{\gamma}_k^{N_1,N_2} \right\rangle \right. \\
&\quad \left. - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1,N_2}(x) \right) \left\langle \partial_c f(\theta) \sigma(Z^{2,N_1}(x)), \tilde{\gamma}_k^{N_1,N_2} \right\rangle \pi(dx, dy) \right\}, \\
& M_{f,2,t}^{N_1,N_2} \\
&= \frac{\alpha_{W,2}^{N_1,N_2}}{N_1^{\gamma_1} N_2^{\gamma_2}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left\{ \left( y_k - g_k^{N_1,N_2}(x_k) \right) \left\langle c\sigma'(Z^{2,N_1}(x_k))\sigma(w^1 x_k) \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1,N_2} \right\rangle \right. \\
&\quad \left. - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1,N_2}(x) \right) \left\langle c\sigma'(Z^{2,N_1}(x))\sigma(w^1 x) \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1,N_2} \right\rangle \pi(dx, dy) \right\}, \\
& M_{f,3,t}^{N_1,N_2} = \\
&= \frac{\alpha_{W,1}^{N_1,N_2}}{N_1^{\gamma_1} N_2^{\gamma_2-1}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left\{ \left( y_k - g_k^{N_1,N_2}(x_k) \right) \right. \\
&\quad \left\langle c\sigma'(Z^{2,N_1}(x_k))\sigma'(w^1 x_k) w^2, \tilde{\gamma}_k^{N_1,N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x_k, \tilde{\gamma}_k^{N_1,N_2} \right\rangle \\
&\quad - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1,N_2}(x) \right) \\
&\quad \left. \left\langle c\sigma'(Z^{2,N_1}(x))\sigma'(w^1 x) w^2, \tilde{\gamma}_k^{N_1,N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x, \tilde{\gamma}_k^{N_1,N_2} \right\rangle \pi(dx, dy).
\end{aligned}$$

Using learning rates as specified in (6), we have

$$\begin{aligned}
& \left\langle f, \gamma_t^{N_1,N_2} \right\rangle - \left\langle f, \gamma_0^{N_1,N_2} \right\rangle \\
&= \frac{1}{N_2^{1-\gamma_2}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1,N_2}(x) \right) \left\langle \partial_c f(\theta) \sigma(Z^{2,N_1}(x)), \gamma_s^{N_1,N_2} \right\rangle \pi(dx, dy) ds \\
&\quad + \frac{1}{N_1^{1-\gamma_1} N_2^{1-\gamma_2}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1,N_2}(x) \right) \\
&\quad \left\langle c\sigma'(Z^{2,N_1}(x))\sigma(w^1 x) \cdot \partial_{w^2} f(\theta), \gamma_s^{N_1,N_2} \right\rangle \pi(dx, dy) ds \\
&\quad + \frac{1}{N_1^{1-\gamma_1} N_2^{1-\gamma_2}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1,N_2}(x) \right) \\
&\quad \left\langle c\sigma'(Z^{2,N_1}(x))\sigma'(w^1 x) w^2, \gamma_s^{N_1,N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x, \gamma_s^{N_1,N_2} \right\rangle \pi(dx, dy) ds \\
&\quad + M_{f,t}^{N_1,N_2} + O\left(\frac{1}{N_2}\right)
\end{aligned} \tag{30}$$

In the following lemma, we prove a uniform bound for  $\mathbb{E} \left( |g_k^N(x)|^4 \right)$ .

**Lemma B.1.** *For any  $k \leq N_2 T$  and any  $x \in \mathcal{X}$ ,*

$$\sup_{N_1, N_2 \in \mathbb{N}, k \leq \lfloor N_2 T \rfloor} \mathbb{E} \left( |g_k^{N_1,N_2}(x)|^4 \right) < C,$$

for some finite constant  $C < \infty$ .

*Proof.* By equation (25), we have the following bound

$$\begin{aligned}
|g_{k+1}^{N_1, N_2}(x)| &\leq |g_k^{N_1, N_2}(x)| + \frac{C\alpha_C^{N_1, N_2}}{N_2^{2\gamma_2-1}} |y_k - g_k^{N_1, N_2}(x_k)| \\
&\quad + \frac{C\alpha_{W,2}^{N_1, N_2}}{N_1^{2\gamma_1} N_2^{2\gamma_2-1}} \sum_{j=1}^{N_1} |y_k - g_k^{N_1, N_2}(x_k)| \\
&\quad + \frac{C\alpha_{W,1}^{N_1, N_2}}{N_1^{2\gamma_1} N_2^{2\gamma_2-2}} \sum_{j=1}^{N_1} |y_k - g_k^{N_1, N_2}(x_k)| + \frac{C}{N_2^{\gamma_2+1}} \\
&\leq |g_k^{N_1, N_2}(x)| + \frac{C}{N_2} |g_k^{N_1, N_2}(x_k)| + \frac{C}{N_2} \left(1 + \frac{1}{N_2^{\gamma_2}}\right) \\
&\leq |g_k^{N_1, N_2}(x)| + \frac{C}{N_2} |g_k^{N_1, N_2}(x_k)| + \frac{C}{N_2},
\end{aligned} \tag{31}$$

where the last inequality holds because  $N_2 > 0$  is large. Squaring both sides of the (31) gives

$$\begin{aligned}
|g_{k+1}^{N_1, N_2}(x)|^2 &\leq |g_k^{N_1, N_2}(x)|^2 + 2|g_k^{N_1, N_2}(x)| \left( \frac{C}{N_2} |g_k^{N_1, N_2}(x_k)| + \frac{C}{N_2} \right) \\
&\quad + \left( \frac{C}{N_2} \right)^2 \left( |g_k^{N_1, N_2}(x_k)| + 1 \right)^2 \\
&\leq |g_k^{N_1, N_2}(x)|^2 + \frac{C}{N_2} |g_k^{N_1, N_2}(x)|^2 + \frac{C}{N_2} |g_k^{N_1, N_2}(x_k)|^2 + \frac{C}{N_2},
\end{aligned}$$

where the last inequality follows from the Young's inequality ( $ab \leq \frac{a^2}{2\epsilon} + \frac{\epsilon b^2}{2}$ , for  $\epsilon = \frac{1}{N_2}$ ). Similarly, squaring both sides one more time gives

$$|g_{k+1}^{N_1, N_2}(x)|^4 \leq |g_k^{N_1, N_2}(x)|^4 + \frac{C}{N_2} |g_k^{N_1, N_2}(x)|^4 + \frac{C}{N_2} |g_k^{N_1, N_2}(x_k)|^4 + \frac{C}{N_2}.$$

Therefore, for  $k \leq N_2 T$ ,

$$\begin{aligned}
|g_k^{N_1, N_2}(x)|^4 &= |g_0^{N_1, N_2}(x)|^4 + \sum_{j=1}^k \left( |g_j^{N_1, N_2}(x)|^4 - |g_{j-1}^{N_1, N_2}(x)|^4 \right) \\
&\leq |g_0^{N_1, N_2}(x)|^4 + \sum_{j=1}^k \left( \frac{C}{N_2} |g_{j-1}^{N_1, N_2}(x)|^4 + \frac{C}{N_2} |g_{j-1}^{N_1, N_2}(x_{j-1})|^4 + \frac{C}{N_2} \right) \\
&\leq |g_0^{N_1, N_2}(x)|^4 + C + \frac{C}{N_2} \sum_{j=1}^k |g_{j-1}^{N_1, N_2}(x)|^4 + \frac{C}{N_2} \sum_{j=1}^k |g_{j-1}^{N_1, N_2}(x_{j-1})|^4.
\end{aligned}$$

We then take expectation on both sides and get

$$\begin{aligned}
\mathbb{E} \left( \left| g_k^{N_1, N_2}(x) \right|^4 \right) &\leq \mathbb{E} \left( \left| g_0^{N_1, N_2}(x) \right|^4 \right) + C + \frac{C}{N_2} \sum_{j=1}^k \mathbb{E} \left( \left| g_{j-1}^{N_1, N_2}(x) \right|^4 \right) \\
&\quad + \frac{C}{N_2} \sum_{j=1}^k \mathbb{E} \left( \left| g_{j-1}^{N_1, N_2}(x_{j-1}) \right|^4 \right) \\
&\leq \mathbb{E} \left( \left| g_0^{N_1, N_2}(x) \right|^4 \right) + C + \frac{C}{N_2} \sum_{j=1}^k \mathbb{E} \left( \left| g_{j-1}^{N_1, N_2}(x) \right|^4 \right) \\
&\quad + \frac{C}{N_2} \sum_{j=1}^k \sum_{x' \in \mathcal{X}} \mathbb{E} \left( \left| g_{j-1}^{N_1, N_2}(x') \right|^4 \right), \tag{32}
\end{aligned}$$

where the last term in the last inequality holds because  $x_j$  are sampled from a fixed data set  $\mathcal{X}$  of size  $M$ .

Therefore, summing both side of (32) with respect to  $x$  gives

$$\begin{aligned}
&\sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| g_k^{N_1, N_2}(x) \right|^4 \right) \\
&\leq \sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| g_0^{N_1, N_1}(x) \right|^4 \right) + CM + \frac{C}{N_2} \sum_{j=1}^k \sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| g_{j-1}^{N_1, N_2}(x) \right|^4 \right) \\
&\quad + \frac{CM}{N_2} \sum_{j=1}^k \sum_{x' \in \mathcal{X}} \mathbb{E} \left( \left| g_{j-1}^{N_1, N_2}(x') \right|^4 \right) \\
&\leq \sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| g_0^{N_1, N_2}(x) \right|^4 \right) + C + \frac{C}{N_2} \sum_{j=1}^k \sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| g_{j-1}^{N_1, N_2}(x) \right|^4 \right). \tag{33}
\end{aligned}$$

Since  $(C_0^i, W_0^{1,j}, W_0^{2,j,i})$  are i.i.d. mean zero random variables, we have

$$\begin{aligned}
\mathbb{E} \left( \left| g_0^N(x) \right|^4 \right) &= \mathbb{E} \left[ \left| \frac{1}{N_2^{\gamma_2}} \sum_{i=1}^{N_2} C_0^i \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_0^{2,j,i} \sigma(W_0^{1,j} x) \right) \right|^4 \right] \\
&\leq \frac{C}{N_2^{4\gamma_2}} \sum_{i=1}^{N_2} \mathbb{E} \left( \left| C_0^i \right|^4 \right) \leq C.
\end{aligned}$$

Then, by applying the discrete Grönwall lemma to equation (33), for any  $0 \leq k \leq \lfloor N_2 T \rfloor$  and  $N_2 \in \mathbb{N}$

$$\sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| g_k^{N_1, N_2}(x) \right|^4 \right) \leq C.$$

The result in the lemma follows.  $\square$

Next, using conditional independence of the terms in the series for  $M_t^{N_1, N_2}$  and  $M_{f,t}^{N_1, N_2}$  as well as the bounds from Lemmas A.1 and B.1, we can establish the following  $L^2$  bounds for the martingale terms  $M_t^{N_1, N_2}$  and  $M_{f,t}^{N_1, N_2}$ , which implies that they converge to zero as  $N_2 \rightarrow \infty$ . The proof is similar to that for Lemma 3.1 in [41] and thus it is omitted.

**Lemma B.2.** *For large  $N_1, N_2 \in \mathbb{N}$  and some finite constant  $C > 0$ , we have*

$$\mathbb{E} \left[ \left( M_t^{N_1, N_2} \right)^2 \right] \leq \frac{C}{N_2}, \quad \mathbb{E} \left[ \left( M_{f,t}^{N_1, N_2} \right)^2 \right] \leq \frac{C}{N_2^{3-2\gamma_2}}.$$

**B.2. Relative compactness.** In this section, we prove the relative compactness of  $\{\gamma^{N_1, N_2}, h^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  in  $D_E([0, T])$ , where  $E = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M$ , and  $N_1 \in \mathbb{N}$  is fixed. Using Lemmas A.1, B.1, and Markov's inequality, we get the following lemma which shows compact containment for  $\{(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2}), t \in [0, T]\}_{N_2 \in \mathbb{N}}$ . The proof is analogous to that for Lemma 3.3 in [42] and thus omitted.

**Lemma B.3.** *For each  $\eta > 0$ , there is a compact subset  $\mathcal{K}$  of  $E$  such that*

$$\sup_{N_1 \in \mathbb{N}, t \in [0, T]} \mathbb{P} \left[ \left( \gamma_t^{N_1, N_2}, h_t^{N_1, N_2} \right) \notin \mathcal{K} \right] < \eta.$$

We now show the regularity of the process  $\gamma^{N_1, N_2}$  in  $D_{\mathcal{M}(\mathbb{R}^{1+N_1(1+d)})}([0, T])$ . For  $z_1, z_2 \in \mathbb{R}$ , define the function  $q(z_1, z_2) = \min\{|z_1 - z_2|, 1\}$ . Let  $\mathcal{F}_t^{N_1, N_2}$  be the  $\sigma$ -algebra generated by  $(C_0^i, W_0^{2,j,i}, W_0^{1,j})_{i,j}$  and  $(x_j, y_j)_{j=0}^{\lfloor N_2 t \rfloor - 1}$ .

**Lemma B.4.** *For any  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$  and  $\delta \in (0, 1)$ , there is a constant  $C < \infty$  such that for  $0 \leq u \leq \delta$ ,  $0 \leq v \leq \delta \wedge t$ , and  $t \in [0, T]$ ,*

$$\begin{aligned} & \mathbb{E} \left[ q \left( \langle f, \gamma_{t+u}^{N_1, N_2} \rangle, \langle f, \gamma_t^{N_1, N_2} \rangle \right) q \left( \langle f, \gamma_t^{N_1, N_2} \rangle, \langle f, \gamma_{t-v}^{N_1, N_2} \rangle \right) \mid \mathcal{F}_t^{N_1, N_2} \right] \\ & \leq \frac{C\delta}{N_2^{1-\gamma_2}} + \frac{C}{N_2^{2-\gamma_2}}. \end{aligned}$$

*Proof.* For  $0 \leq s < t \leq T$ , using a Taylor expansion, we have

$$\begin{aligned} & \left| \langle f, \gamma_t^{N_1, N_2} \rangle - \langle f, \gamma_s^{N_1, N_2} \rangle \right| = \left| \langle f, \tilde{\gamma}_{\lfloor N_2 t \rfloor}^{N_1, N_2} \rangle - \langle f, \tilde{\gamma}_{\lfloor N_2 s \rfloor}^{N_1, N_2} \rangle \right| \\ & \leq \frac{1}{N_2} \sum_{i=1}^{N_2} \left| f(\theta_{\lfloor N_2 t \rfloor}^i) - f(\theta_{\lfloor N_2 s \rfloor}^i) \right| \\ & \leq \frac{1}{N_2} \sum_{i=1}^{N_2} \left| \partial_c f(\bar{\theta}_{\lfloor N_2 t \rfloor}^i) \right| \left| C_{\lfloor N_2 t \rfloor}^i - C_{\lfloor N_2 s \rfloor}^i \right| \\ & + \frac{1}{N_2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \left| \partial_{w^{2,j}} f(\bar{\theta}_{\lfloor N_2 t \rfloor}^i) \right| \left| W_{\lfloor N_2 t \rfloor}^{2,j,i} - W_{\lfloor N_2 s \rfloor}^{2,j,i} \right| \\ & + \frac{1}{N_2} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \left\| \nabla_{w^{1,j}} f(\bar{\theta}_{\lfloor N_2 t \rfloor}^i) \right\| \left\| W_{\lfloor N_2 t \rfloor}^{1,j} - W_{\lfloor N_2 s \rfloor}^{1,j} \right\| \end{aligned} \tag{34}$$

for some  $\bar{\theta}_{\lfloor N_2 t \rfloor}^i$  in the line segments between  $\theta_{\lfloor N_2 s \rfloor}^i$  and  $\theta_{\lfloor N_2 t \rfloor}^i$ . With  $0 < t - s \leq \delta < 1$ , by Lemmas A.1, B.1, we have

$$\begin{aligned} & \mathbb{E} \left( \left| C_{\lfloor N_2 t \rfloor}^i - C_{\lfloor N_2 s \rfloor}^i \right| \mid \mathcal{F}_s^{N_1, N_2} \right) \leq \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor - 1} \mathbb{E} \left( \left| C_{k+1}^i - C_k^i \right| \mid \mathcal{F}_s^{N_1, N_2} \right) \\ & \leq \frac{1}{N_2^{2-\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor - 1} C \\ & \leq \frac{C\delta}{N_2^{1-\gamma_2}} + \frac{C}{N_2^{2-\gamma_2}}. \end{aligned}$$

Similar analysis shows

$$\begin{aligned}\mathbb{E} \left( \left| W_{\lfloor N_2 t \rfloor}^{2,j,i} - W_{\lfloor N_2 s \rfloor}^{2,j,i} \right| \middle| \mathcal{F}_s^{N_1, N_2} \right) &\leq \frac{C\delta}{N_1^{1-\gamma_1} N_2^{1-\gamma_2}} + \frac{C}{N_1^{1-\gamma_1} N_2^{2-\gamma_2}}, \\ \mathbb{E} \left( \left\| W_{\lfloor N_2 t \rfloor}^{1,j} - W_{\lfloor N_2 s \rfloor}^{1,j} \right\| \middle| \mathcal{F}_s^{N_1, N_2} \right) &\leq \frac{C\delta}{N_1^{1-\gamma_1} N_2^{1-\gamma_2}} + \frac{C}{N_1^{1-\gamma_1} N_2^{2-\gamma_2}}.\end{aligned}$$

By Lemma A.1,  $\bar{\theta}_{\lfloor N_2 t \rfloor}^i$  is bounded in expectation for  $0 < s < t \leq T$ . Taking conditional expectation on both sides of (34) and using bounds we derived above yields

$$\mathbb{E} \left[ \left\langle f, \gamma_t^{N_1, N_2} \right\rangle - \left\langle f, \gamma_s^{N_1, N_2} \right\rangle \middle| \mathcal{F}_s^{N_1, N_2} \right] \leq \frac{C\delta}{N_2^{1-\gamma_2}} + \frac{C}{N_2^{2-\gamma_2}},$$

for  $0 < s < t \leq T$  with  $0 < t - s \leq \delta < 1$ , and some unimportant positive constant  $C < \infty$ . Therefore, the statement of the lemma follows.  $\square$

We next establish the regularity of the process  $h^{N_1, N_2}$  in  $D_{\mathbb{R}^M}([0, T])$  in the following lemma. For the purpose of this lemma, we denote  $q(z_1, z_2) = \min\{\|z_1 - z_2\|_{l^1}, 1\}$  for  $z_1, z_2 \in \mathbb{R}^M$ .

**Lemma B.5.** *For any  $\delta \in (0, 1)$ , there is a constant  $C < \infty$  such that for  $0 \leq u \leq \delta$ ,  $0 \leq v \leq \delta \wedge t$ , and  $t \in [0, T]$ ,*

$$\mathbb{E} \left[ q \left( h_{t+u}^{N_1, N_2}, h_t^{N_1, N_2} \right) q \left( h_t^{N_1, N_2}, h_{t-v}^{N_1, N_2} \right) \middle| \mathcal{F}_t^{N_1, N_2} \right] \leq C\delta + \frac{C}{N_2}.$$

*Proof.* For  $0 < s < t \leq T$ , by the Taylor expansion of the network output  $g_k^{N_1, N_2}(x)$ , we have

$$\begin{aligned}\left| h_t^{N_1, N_2}(x) - h_s^{N_1, N_2}(x) \right| &\leq \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor-1} \left| g_{k+1}^{N_1, N_2}(x) - g_k^{N_1, N_2}(x) \right| \\ &\leq \frac{1}{N_2^{\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor-1} \sum_{i=1}^{N_2} \left| C_{k+1}^i - C_k^i \right| \left| \sigma \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \right| \\ &\quad + \frac{1}{N_1^{\gamma_1} N_2^{\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor-1} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \left| C_k^i \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \right. \\ &\quad \left. \sigma(W_k^{1,j} x) \left( W_{k+1}^{2,j,i} - W_k^{2,j,i} \right) \right| \\ &\quad + \frac{1}{N_1^{\gamma_1} N_2^{\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor-1} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \left| C_k^i \sigma' \left( \frac{1}{N_1^{\gamma_1}} \sum_{j=1}^{N_1} W_k^{2,j,i} \sigma(W_k^{1,j} x) \right) \right. \\ &\quad \left. W_k^{2,j,i} \sigma'(W_k^{1,j} x) \left( W_{k+1}^{1,j} - W_k^{1,j} \right) x \right| \\ &\leq \frac{C}{N_2^{\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor-1} \sum_{i=1}^{N_2} \left| C_{k+1}^i - C_k^i \right| \\ &\quad + \frac{C}{N_1^{\gamma_1} N_2^{\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor-1} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \left( \left| W_{k+1}^{2,j,i} - W_k^{2,j,i} \right| + \|W_{k+1}^i - W_k^i\| \right).\end{aligned}\tag{35}$$

By taking conditional expectation on both sides of (35) and using the bounds we derived in the proof of Lemma B.4,

$$\begin{aligned} \mathbb{E} \left[ \left| h_t^{N_1, N_2}(x) - h_s^{N_1, N_2}(x) \right| \middle| \mathcal{F}_s^{N_1, N_2} \right] &\leq \frac{C}{N_2^{\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor - 1} \sum_{i=1}^{N_2} \mathbb{E} \left[ \left| C_{k+1}^i - C_k^i \right| \middle| \mathcal{F}_s^{N_1, N_2} \right] \\ &+ \frac{C}{N_1^{\gamma_1} N_2^{\gamma_2}} \sum_{k=\lfloor N_2 s \rfloor}^{\lfloor N_2 t \rfloor - 1} \sum_{i=1}^{N_2} \sum_{j=1}^{N_1} \mathbb{E} \left[ \left| W_{k+1}^{2, j, i} - W_k^{2, j, i} \right| + \| W_{k+1}^i - W_k^i \| \middle| \mathcal{F}_s^{N_1, N_2} \right] \\ &\leq C\delta + \frac{C}{N_2}. \end{aligned}$$

Since  $x \in \mathcal{X}$  is arbitrary, the bound above implies that

$$\mathbb{E} \left[ \left\| h_t^{N_1, N_2} - h_s^{N_1, N_2} \right\|_{l^1} \middle| \mathcal{F}_s^{N_1, N_2} \right] \leq C\delta + \frac{C}{N_2}.$$

The statement of the lemma then follows.  $\square$

Combining Lemmas B.3 to B.5, we have the following lemma for the relative compactness of the processes  $\{\gamma^{N_1, N_2}, h^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  for fixed  $N_1$ . The proof is similar to that of Lemma 3.6 in [39], which is omitted here.

**Lemma B.6.** *The sequence of processes  $\{\gamma^{N_1, N_2}, h^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_E([0, T])$ , where  $E = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M$ .*

**B.3. Identification of the limit.** In this section, we show that for fixed  $N_1$  and as  $N_2 \rightarrow \infty$ , the process  $(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2})$  converges in distribution in the space  $D_E([0, T])$  to  $(\gamma_t^{N_1}, h_t^{N_1})$ , which satisfies the evolution equation

$$h_t^{N_1}(x) = h_0^{N_1}(x) + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) A_{x, x'}^{N_1} \pi(dx', dy) ds, \quad (36)$$

where  $\gamma_0^{N_1}$  is given by (8), and if  $\gamma_2 = 1/2$ ,  $h_0^{N_1}(x) = \mathcal{G}^{N_1}(x)$ , where  $\mathcal{G}^{N_1}$  is Gaussian, and if  $\gamma_2 > 1/2$ ,  $h_0^{N_1}(x) = 0$ .

Let  $\pi^{N_1, N_2} \in \mathcal{M}(D_E([0, T]))$  be the probability measure corresponding to  $(\gamma^{N_1, N_2}, h^{N_1, N_2})$ . Relative compactness implies that there is a subsequence  $\pi^{N_1, N_{2k}}$  that converges weakly. We must show that any limit point  $\pi^{N_1}$  of a convergent subsequence  $\pi^{N_1, N_{2k}}$  is a Dirac measure concentrated on  $(\gamma^{N_1}, h^{N_1}) \in D_E([0, T])$ , which satisfies equation (36) and  $\langle f, \gamma_t^{N_1} \rangle = \langle f, \gamma_0^{N_1} \rangle$  for any test function  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ .

We define the map  $F(\gamma^{N_1}, h^{N_1}) : D_E([0, T]) \rightarrow \mathbb{R}_+$  for each  $t \in [0, T]$ ,  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ ,  $g_1, \dots, g_p \in C_b(\mathbb{R}^{1+N_1(1+d)})$ ,  $q_{1,1}, \dots, q_{1,p}, q_{2,1}, \dots, q_{N_2,p} \in C_b(\mathbb{R}^M)$ ,  $m_1, \dots, m_p \in C_b(\mathbb{R}^M)$  and  $0 \leq s_1 < \dots < s_p \leq t$ :

$$\begin{aligned} F(\gamma, h) &= \left| \left( \langle f, \gamma_t^{N_1} \rangle - \langle f, \gamma_0^{N_1} \rangle \right) \times \langle g_1, \gamma_{s_1}^{N_1} \rangle \times \dots \times \langle g_p, \gamma_{s_p}^{N_1} \rangle \right| \\ &+ \sum_{x \in \mathcal{X}} \left| \left( h_t^{N_1}(x) - h_0^{N_1}(x) \right) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \langle \sigma(Z^{2, N_1}(x')) \sigma(Z^{2, N_1}(x)), \gamma_s^{N_1} \rangle \pi(dx', dy) ds \right| \end{aligned} \quad (37)$$

$$\begin{aligned}
& -\frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
& \quad \langle (c)^2 \sigma'(Z^{2,N_1}(x')) \sigma'(Z^{2,N_1}(x)) \sigma(w^{1,j}x') \sigma(w^{1,j}x), \gamma_s^{N_1} \rangle \pi(dx', dy) ds \\
& -\frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' \langle c w^{2,j} \sigma'(w^{1,j}x) \sigma'(Z^{2,N_1}(x)), \gamma_s^{N_1} \rangle \\
& \quad \cdot \langle c w^{2,j} \sigma'(w^{1,j}x') \sigma'(Z^{2,N_1}(x')), \gamma_s^{N_1} \rangle \pi(dx', dy) ds \times m_1(h_{s_1}^{N_1}) \times \cdots \times m_p(h_{s_p}^{N_1}) \Big|.
\end{aligned}$$

By equations (30), (28), Lemma B.2 and the Cauchy-Schwarz inequality, we have

$$\begin{aligned}
& \mathbb{E}_{\pi^{N_1, N_2}} [F(\gamma^{N_1}, h^{N_1})] \\
& = \mathbb{E} [F(\gamma^{N_1, N_2}, h^{N_1, N_2})] \\
& = \mathbb{E} \left[ \left| O\left(N_2^{-(1-\gamma_2)}\right) + M_{f,t}^{N_1, N_2} + O(N_2^{-1}) \times \prod_{i=1}^p \langle g_i, \gamma_{s_i}^{N_1, N_2} \rangle \right| \right] \\
& \quad + \sum_{x \in \mathcal{X}} \mathbb{E} \left[ \left| \left( M_t^{N_1, N_2} + O(N_2^{-\gamma_2}) \right) \times \prod_{i=1}^p m_i(h_{s_i}^{N_1, N_2}) \right| \right] \\
& \leq C \left( \mathbb{E} \left[ \left| M_{f,t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \left| M_t^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} \right) + O(N_2^{-(1-\gamma_2)}) \\
& \leq C \left( \frac{1}{N_2^{1-\gamma_2}} \right).
\end{aligned}$$

Therefore,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F(\gamma^{N_1}, h^{N_1})] = 0$ . Since  $F(\cdot)$  is continuous and  $F(\gamma^{N_1, N_2}, h^{N_1, N_2})$  is uniformly bounded, we have  $\mathbb{E}_{\pi^{N_1}} [F(\gamma^{N_1}, h^{N_1})] = 0$ . Hence,  $(\gamma^{N_1}, h^{N_1})$  satisfies the evolution equation (36) and  $\langle f, \gamma_t^{N_1} \rangle = \langle f, \gamma_0^{N_1} \rangle$  for any test function  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ .

Since equation (36) is a finite-dimensional, linear equation, it has a unique solution. By Prokhorov's theorem,  $\pi^{N_1, N_2}$  converges weakly to  $\pi^{N_1}$ , which is the distribution of  $(\gamma^{N_1}, h^{N_1})$ , the unique solution of (36). Hence, for fixed  $N_1$ ,  $(\gamma^{N_1, N_2}, h^{N_1, N_2})$  converges in distribution to  $(\gamma^{N_1}, h^{N_1})$  as  $N_2 \rightarrow \infty$ .

**Appendix C. Proof of Theorem 2.7.** In this section, we look at the convergence of the first order fluctuation process of the network output for fixed  $N_1$  and study its limiting behavior as  $N_2 \rightarrow \infty$ . In particular, consider

$$K_t^{N_1, N_2} = N_2^\varphi (h_t^{N_1, N_2} - h_t^{N_1}),$$

where  $\varphi$  is dependent on the scaling parameters  $\gamma_1, \gamma_2$ . We also denote  $\eta_t^{N_1, N_2} = N_2^\varphi (\gamma_t^{N_1, N_2} - \gamma_0^{N_1})$ .

For  $t \in [0, T]$  and  $x \in \mathcal{X}$ , by equations (28) and (36), the evolution of  $K_t^{N_1, N_2}(x)$  can be written as

$$\begin{aligned}
& K_t^{N_1, N_2}(x) \\
& = N_2^\varphi \left[ \left( h_t^{N_1, N_2} - h_0^{N_1, N_2} \right) + h_0^{N_1, N_2} - h_t^{N_1} \right] \\
& = N_2^\varphi \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \langle B_{x,x'}^1(\theta), \gamma_s^{N_1, N_2} \rangle \pi(dx', dy) ds \right\}
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left\langle B_{x, x'}^{2, j}(\theta), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' \\
& \quad \left\langle B_x^{3, j}(\theta), \gamma_s^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \Big\} \\
& - N_2^\varphi \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) \left\langle B_{x, x'}^1(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right. \\
& \quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) \left\langle B_{x, x'}^{2, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
& \quad \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right\} \\
& + K_0^{N_1, N_2} + N_2^\varphi M_t^{N_1, N_2} + O(N_2^{-\gamma_2 + \varphi})
\end{aligned}$$

By rearranging terms, we obtain

$$\begin{aligned}
K_t^{N_1, N_2}(x) &= \\
&= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) \left\langle B_{x, x'}^1(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x, x'}^1(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) \left\langle B_{x, x'}^{2, j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \right. \\
&\quad \left. - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x, x'}^{2, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right\} \\
&\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' \left\langle B_x^{3, j}(\theta), \eta_s^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right. \\
&\quad \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \right. \\
&\quad \left. - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right\} \\
&\quad + \Gamma_t^{N_1, N_2}(x) + K_0^{N_1, N_2} + N_2^\varphi M_t^{N_1, N_2} + O(N_2^{-\gamma_2 + \varphi})
\end{aligned} \tag{38}$$

where  $\Gamma_t^{N_1, N_2}(x) = \Gamma_{1, t}^{N_1, N_2}(x) + \Gamma_{2, t}^{N_1, N_2}(x) + \Gamma_{3, t}^{N_1, N_2}(x)$ , and

$$\begin{aligned}
\Gamma_{1, t}^{N_1, N_2}(x) &= -\frac{1}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x, x'}^1(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds, \\
\Gamma_{2, t}^{N_1, N_2}(x) &= -\frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x, x'}^{2, j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
\Gamma_{3, t}^{N_1, N_2}(x) &= -\frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \right.
\end{aligned}$$

$$\begin{aligned}
& \left\langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' \left\langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \Big\} .
\end{aligned}$$

Recall that when  $\gamma_2 > 1/2$ ,  $h_0^{N_1}(x) = 0$ . Therefore,

$$K_0^{N_1, N_2}(x) = N_2^\varphi h_0^{N_1, N_2}(x) = N_2^{-(\gamma_2 - \frac{1}{2} - \varphi)} \left\langle c\sigma(Z^{2, N_1}(x)), \sqrt{N_2} \tilde{\gamma}_0^{N_1, N_2} \right\rangle,$$

which, by the central limit theorem, converges to the Gaussian random variable  $\mathcal{G}^{N_1}(x)$  if  $\varphi = \gamma_2 - (1/2)$  and to 0 if  $\varphi < \gamma_2 - (1/2)$ .

We also need to consider the evolution of  $l_t^{N_1, N_2}(f) = \left\langle f, \eta_t^{N_1, N_2} \right\rangle$  for a fixed function  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ . By (29), for  $N_2$  large enough, we have

$$\begin{aligned}
& \left\langle f, \eta_t^{N_1, N_2} \right\rangle - \left\langle f, \eta_0^{N_1, N_2} \right\rangle = N_2^\varphi \left( \left\langle f, \gamma_t^{N_1, N_2} \right\rangle - \left\langle f, \gamma_0^{N_1, N_2} \right\rangle \right) \\
& = \frac{1}{N_2^{2-\gamma_2-\varphi}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x_k)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
& + \frac{1}{N_1^{1-\gamma_1} N_2^{2-\gamma_2-\varphi}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \\
& \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma(w^1 x_k) \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \tag{39} \\
& + \frac{1}{N_1^{1-\gamma_1} N_2^{2-\gamma_2-\varphi}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \\
& \left\langle \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma'(w^1 x_k) w^2, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x_k, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
& + O\left(\frac{1}{N_2^{1-\varphi}}\right).
\end{aligned}$$

The evolution equations (38) and (39) suggest that we consider the convergence of  $K_t^{N_1, N_2}$  and  $l_t^{N_1, N_2}(f)$  for  $\varphi \leq \min\{1 - \gamma_2, \gamma_2 - (1/2)\}$ . If  $\gamma_2 < \frac{3}{4}$ , we can take  $\varphi = \gamma_2 - \frac{1}{2} < 1 - \gamma_2$  in order to obtain a limiting Gaussian process for  $K_t^{N_1, N_2}$ . If  $\gamma_2 \geq \frac{3}{4}$ , the limiting process for  $K_t^{N_1, N_2}$  is Gaussian only if  $\gamma_2 = \frac{3}{4}$  and  $\varphi = 1 - \gamma_2 = \gamma_2 - \frac{1}{2}$ .

**C.1. Convergence of  $l_t^{N_1, N_2}(f) = \left\langle f, \eta_t^{N_1, N_2} \right\rangle$ .** In this section, we establish the convergence of the process  $l_t^{N_1, N_2}(f)$  as  $N_2 \rightarrow \infty$  in  $D_{\mathbb{R}}([0, T])$  for a fixed function  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ .

Following the same idea as in Section B, we first show that relative compactness holds. The following lemma implies compact containment of the process  $\{l_t^{N_1, N_2}(f)\}$ .

**Lemma C.1.** *For any fixed  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ , when  $\varphi \leq 1 - \gamma_2$ , there exist a constant  $C < \infty$ , such that*

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{E} \left[ \left| \left\langle f, \eta_t^{N_1, N_2} \right\rangle \right|^4 \right] < C.$$

Furthermore, for any  $\epsilon > 0$ , there exist a compact subset  $U \subset \mathbb{R}$  such that

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{P} \left( \left\langle f, \eta_t^{N_1, N_2} \right\rangle \notin U \right) < \epsilon.$$

*Proof.* By equation (39), we have

$$\begin{aligned} \left| \left\langle f, \eta_t^{N_1, N_2} \right\rangle \right| &\leq \left| \left\langle f, \eta_0^{N_1, N_2} \right\rangle \right| + \frac{C}{N_2^{2-\gamma_2-\varphi}} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left| y_k - g_k^{N_1, N_2}(x_k) \right| + \frac{C}{N_2^{1-\varphi}} \\ &\leq \left| \left\langle f, \eta_0^{N_1, N_2} \right\rangle \right| + \frac{C}{N_2^{1-\gamma_2-\varphi}} \left| g_k^{N_1, N_2}(x_k) \right| + \frac{C}{N_2^{1-\gamma_2-\varphi}}. \end{aligned}$$

Raising to the forth power on both sides, by Hölder's inequality, we have

$$\left| \left\langle f, \eta_t^{N_1, N_2} \right\rangle \right|^4 \leq 9 \left( \left| \left\langle f, \eta_0^{N_1, N_2} \right\rangle \right|^4 + \frac{C}{N_2^{4(1-\gamma_2-\varphi)}} \left| g_k^{N_1, N_2}(x_k) \right|^4 + \frac{C}{N_2^{4(1-\gamma_2-\varphi)}} \right). \quad (40)$$

Since  $\left\langle f, \eta_0^{N_1, N_2} \right\rangle = N_2^\varphi \left\langle f, \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle$ , and by independence,

$$\begin{aligned} \mathbb{E} \left[ \left| \left\langle f, \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle \right|^4 \right] &= \mathbb{E} \left[ \left| \frac{1}{N_2} \sum_{i=1}^{N_2} f(\theta_0^i) - \left\langle f, \gamma_0^{N_1} \right\rangle \right|^4 \right] \\ &= \frac{1}{N_2^4} \sum_{i=1}^{N_2} \mathbb{E} \left[ \left| f(\theta_0^i) - \left\langle f, \gamma_0^{N_1} \right\rangle \right|^4 \right] < \frac{C}{N_2^3}, \end{aligned}$$

we have  $\mathbb{E} \left[ \left| \left\langle f, \eta_0^{N_1, N_2} \right\rangle \right|^4 \right] \leq C(N_2^{4\varphi-3})$ . Taking expectation on both sides of equation (40), by Lemma B.1 and  $4\varphi - 3 < 0$ , we have

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{E} \left[ \left| \left\langle f, \eta_t^{N_1, N_2} \right\rangle \right|^4 \right] < C,$$

for some  $C < \infty$ . By Markov's inequality, the compact containment condition of  $\left\langle f, \eta_t^{N_1, N_2} \right\rangle$  follows.  $\square$

Next, we establish the regularity of  $\left\langle f, \eta_t^{N_1, N_2} \right\rangle$ . For the following lemma, we define the function  $q(z_1, z_2) = \min\{|z_1 - z_2|, 1\}$ , where  $z_1, z_2 \in \mathbb{R}$ .

**Lemma C.2.** *For  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ ,  $\delta \in (0, 1)$ , there exist a constant  $C < \infty$  such that for any  $0 \leq u \leq \delta$ ,  $0 \leq v \leq \delta \wedge t$ , and  $t \in [0, T]$ ,*

$$\begin{aligned} &\mathbb{E} \left[ q \left( \left\langle f, \eta_{t+u}^{N_1, N_2} \right\rangle, \left\langle f, \eta_t^{N_1, N_2} \right\rangle \right) q \left( \left\langle f, \eta_t^{N_1, N_2} \right\rangle, \left\langle f, \eta_{t-v}^{N_1, N_2} \right\rangle \right) \mid \mathcal{F}_t^{N_1, N_2} \right] \\ &\leq \frac{C\delta}{N_2^{1-\gamma-\varphi}} + \frac{C}{N_2^{2-\gamma-\varphi}}, \end{aligned}$$

where  $\varphi \leq 1 - \gamma_2$ .

*Proof.* Recall that  $\langle f, \gamma_t^{N_1} \rangle = \langle f, \gamma_0^{N_1} \rangle$  for any  $t \in [0, T]$  and  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ .

For any  $0 \leq s < t \leq T$ , by the regularity result for  $\gamma_t^{N_1, N_2}$  proved in Lemma B.4, we have

$$\begin{aligned} & \mathbb{E} \left[ \left| \langle f, \eta_t^{N_1, N_2} \rangle - \langle f, \eta_s^{N_1, N_2} \rangle \right| \middle| \mathcal{F}_s^{N_1, N_2} \right] \\ &= N_2^\varphi \mathbb{E} \left[ \left| \langle f, \gamma_t^{N_1, N_2} \rangle - \langle f, \gamma_s^{N_1, N_2} \rangle \right| \middle| \mathcal{F}_s^{N_1, N_2} \right] \\ &\leq \frac{C\delta}{N_2^{1-\gamma_2-\varphi}} + \frac{C}{N_2^{2-\gamma_2-\varphi}}, \end{aligned}$$

for  $0 < s < t \leq T$  with  $0 < t - s \leq \delta < 1$ . If  $\varphi \leq 1 - \gamma_2$ , both terms in the last inequality above are bounded as  $N_2$  grows. The statement of the lemma follows.  $\square$

Using Lemmas C.1 and C.2, we are now ready to present the proof of the convergence of  $l_t^{N_1, N_2}(f)$ . We first show the case when  $\varphi < 1 - \gamma_2$ . For fixed  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ , when  $\varphi \leq 1 - \gamma_2$ , the family of processes  $\{\langle f, \eta_t^{N_1, N_2} \rangle, t \in [0, T]\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{\mathbb{R}}([0, T])$  due to Lemmas C.1, C.2, and Theorem 8.6 of Chapter 3 of [11]. For simplicity, we denote  $l_t^{N_1, N_2} = \langle f, \eta_t^{N_1, N_2} \rangle$ . Let  $\pi^{N_1, N_2} \in \mathcal{M}(D_{\mathbb{R}}([0, T]))$  be the probability measure corresponding to  $l_t^{N_1, N_2}$ . Relative compactness implies that there is a subsequence  $\pi^{N_1, N_{2k}}$  that converges weakly to a limit point  $\pi^{N_1}$ . We show that  $\pi^{N_1}$  is a Dirac measure concentrated on zero when  $\varphi < 1 - \gamma_2$ .

For  $t \in [0, T]$ ,  $g_1, \dots, g_p \in C_b(\mathbb{R})$ , and  $0 \leq s_1 < \dots < s_p \leq t$ , define a map  $F(l) : D_{\mathbb{R}}([0, T]) \rightarrow \mathbb{R}_+$  as

$$F(l) = |(l_t - 0) \times g_1(l_{s_1}) \times \dots \times g_p(l_{s_p})|.$$

By equation (39) and the fact that  $\langle f, \eta_0^{N_1, N_2} \rangle = N_2^{\varphi - \frac{1}{2}} \langle f, \sqrt{N_2} (\gamma_0^{N_1, N_2} - \gamma_0^{N_1}) \rangle = O_p(N_2^{\varphi - \frac{1}{2}})$ , we have

$$\begin{aligned} & \mathbb{E}_{\pi^{N_1, N_2}} [F(l)] \\ &= \mathbb{E} [F(l^{N_1, N_2})] \\ &= \mathbb{E} \left[ \left| \left( \langle f, \eta_t^{N_1, N_2} \rangle - \langle f, \eta_0^{N_1, N_2} \rangle + \langle f, \eta_0^{N_1, N_2} \rangle \right) \times \prod_{i=1}^p g_i(l_{s_i}^{N_1, N_2}) \right| \right] \\ &\leq \mathbb{E} \left[ \left| \left( \langle f, \eta_t^{N_1, N_2} \rangle - \langle f, \eta_0^{N_1, N_2} \rangle \right) \times \prod_{i=1}^p g_i(l_{s_i}^{N_1, N_2}) \right| \right] \\ &\quad + \mathbb{E} \left[ \left| \langle f, \eta_0^{N_1, N_2} \rangle \times \prod_{i=1}^p g_i(l_{s_i}^{N_1, N_2}) \right| \right] \\ &\leq C \left( \frac{1}{N_2^{1-\gamma_2-\varphi}} + \frac{1}{N_2^{1-\varphi}} + \frac{1}{N_2^{\frac{1}{2}-\varphi}} \right). \end{aligned}$$

Since  $F(\cdot)$  is continuous and  $F(l^{N_1, N_2})$  is uniformly bounded, we have

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F(l)] = \mathbb{E}_{\pi^{N_1}} [F(l)] = 0,$$

where  $\pi^{N_1}$  is the Dirac measure concentrated on 0. We have shown that the limit point  $\pi$  of any convergence subsequence, which exists due to relative compactness, is the Dirac measure concentrated on 0. Therefore, by Prokhorov's theorem,  $\pi^{N_1, N_2}$

weakly converges to 0. As  $N_2 \rightarrow \infty$ ,  $l^{N_1, N_2}(f) \xrightarrow{d} 0$  and thus the limit is in probability. This concludes the proof for case 1:  $\varphi < 1 - \gamma_2$ .

The proof for case 2:  $\varphi = 1 - \gamma_2$  is more subtle and is given in different steps below. We see that the evolution of  $l_t^{N_1, N_2}(f)$  becomes

$$\begin{aligned}
& \left\langle f, \eta_t^{N_1, N_2} \right\rangle - \left\langle f, \eta_0^{N_1, N_2} \right\rangle \\
&= \frac{1}{N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \\
&+ \frac{1}{N_1^{1-\gamma_1} N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \\
&\quad \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \\
&+ \frac{1}{N_1^{1-\gamma_1} N_2} \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \\
&\quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \\
&+ M_{\eta, 1, t}^{N_1, N_2} + M_{\eta, 2, t}^{N_1, N_2} + M_{\eta, 3, t}^{N_1, N_2} + O\left(\frac{1}{N_2^{\gamma_2}}\right).
\end{aligned}$$

where

$$\begin{aligned}
M_{\eta, 1, t}^{N_1, N_2} &= \frac{1}{N_2} \left\{ \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x_k)), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \right. \\
&\quad \left. - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \right\},
\end{aligned}$$

$$\begin{aligned}
M_{\eta, 2, t}^{N_1, N_2} &= \\
&= \frac{1}{N_1^{1-\gamma_1} N_2} \left\{ \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma(w^1 x_k) \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \right. \\
&\quad \left. - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \right\},
\end{aligned}$$

$$\begin{aligned}
M_{\eta, 3, t}^{N_1, N_2} &= \\
&= \frac{1}{N_1^{1-\gamma_1} N_2} \left\{ \sum_{k=0}^{\lfloor N_2 t \rfloor - 1} \left( y_k - g_k^{N_1, N_2}(x_k) \right) \right. \\
&\quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x_k)) \sigma'(w^1 x_k) w^2, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x_k, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \\
&\quad - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \\
&\quad \left. \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \tilde{\gamma}_k^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \right\}.
\end{aligned}$$

As  $N_2$  grows, we can rewrite this equation in terms of Riemann integrals and scaled measure  $\gamma_t^{N_1, N_2}$ ,

$$\begin{aligned}
& \left\langle f, \eta_t^{N_1, N_2} \right\rangle - \left\langle f, \eta_0^{N_1, N_2} \right\rangle \\
&= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \\
&\quad \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \\
&\quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad + M_{\eta, 1, t}^{N_1, N_2} + M_{\eta, 2, t}^{N_1, N_2} + M_{\eta, 3, t}^{N_1, N_2} + O\left(\frac{1}{N_2^{\gamma_2}}\right). \tag{41}
\end{aligned}$$

Fir any fixed  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ , similar analysis as in Lemma 3.1 in [39], we have the following bound for terms  $M_{\eta, i, t}^{N_1, N_2}, i = 1, 2, 3$ .

**Lemma C.3.** *For any  $N \in \mathbb{N}$ , there is a constant  $C < \infty$  such that*

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left( \left| M_{\eta, 1, t}^{N_1, N_2} \right|^2 + \left| M_{\eta, 2, t}^{N_1, N_2} \right|^2 + \left| M_{\eta, 3, t}^{N_1, N_2} \right|^2 \right) \right] \leq \frac{C}{N_2}.$$

From equation (41), we see that the evolution of  $l_t^{N_1, N_2}(f)$  involves the evolution of  $\gamma_t^{N_1, N_2}$  and  $h_t^{N_1, N_2}$ . In the next lemma, we prove the convergence of the processes  $(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2}, l_t^{N_1, N_2}(f))$  in distribution in the space  $D_{E'}([0, T])$ , where  $E' = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M \times \mathbb{R}$ . The convergence of  $l_t^{N_1, N_2}(f)$  case 2:  $\varphi = 1 - \gamma$  then follows from Lemma C.4.

**Lemma C.4.** *For any fixed  $f \in C_b^2(\mathbb{R}^{1+N_1(1+d)})$ , if  $\varphi = 1 - \gamma_2$ , the processes  $(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2}, l_t^{N_1, N_2}(f))$  converges in distribution in  $D_{E'}([0, T])$  to  $(\gamma_0^{N_1}, h_t^{N_1}, l_t^{N_1}(f))$ , where  $h_t^{N_1}$  satisfies equation (36) and  $l_t^{N_1}(f)$  is given by*

$$\begin{aligned}
l_t^{N_1}(f) &= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
&\quad \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
&\quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \tag{42}
\end{aligned}$$

*Proof.* By Lemmas C.1 and C.2,  $\{l^{N_1, N_2}(f)\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{\mathbb{R}}([0, T])$ . By Lemma B.6,  $\{\gamma^{N_1, N_2}, h^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_E([0, T])$ , where  $E = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M$ . Since relative compactness is equivalent to tightness, we have that the probability measures of the family of processes  $\{l^{N_1, N_2}(f)\}_{N_2 \in \mathbb{N}}$

and the probability measures of the family of processes  $\{\gamma^{N_1, N_2}, h^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  are tight. Therefore,  $\{\gamma^{N_1, N_2}, h^{N_1, N_2}, l^{N_1, N_2}(f)\}_{N_2 \in \mathbb{N}}$  is tight, hence it is also relatively compact.

Denote  $\pi^{N_1, N_2} \in \mathcal{M}(D_{E'}([0, T]))$  the probability measure corresponding to  $(\gamma^{N_1, N_2}, h^{N_1, N_2}, l^{N_1, N_2}(f))$ . Relative compactness implies that there is a subsequence  $\pi^{N_1, N_{2k}}$  that converges weakly. We now show that any limit point  $\pi$  of a convergent subsequence  $\pi^{N_1, N_{2k}}$  is a Dirac measure concentrated on  $(\gamma^{N_1}, h^{N_1}, l^{N_1}(f)) \in D_{E'}([0, T])$ , where  $(\gamma^{N_1}, h^{N_1}, l^{N_1}(f))$  satisfies equations (36) and (42). Define a map  $F_1(\gamma^{N_1}, h^{N_1}, l^{N_1}(f)) : D_{E'}([0, T]) \rightarrow \mathbb{R}_+$  for each  $t \in [0, T]$ ,  $m_1, \dots, m_p \in C_b(\mathbb{R})$ , and  $0 \leq s_1 < \dots < s_p \leq t$ .

$$\begin{aligned}
& F_1(\gamma, h, l(f)) \\
&= F(\gamma^{N_1}, h^{N_1}) \\
&+ \left| \left( l_t^{N_1}(f) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_s^{N_1} \rangle \pi(dx', dy) ds \right. \right. \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
&\quad \langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_s^{N_1} \rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
&\quad \langle \langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_s^{N_1} \rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_s^{N_1} \rangle \pi(dx', dy) ds \\
&\quad \left. \left. \times m_1(l_{s_1}^{N_1}(f)) \times \dots \times m_p(l_{s_p}^{N_1}(f)) \right| \right|, \tag{43}
\end{aligned}$$

where  $F(\gamma^{N_1}, h^{N_1})$  is as given in equation (B.3). Using equation (41), Lemma C.3, the analysis of  $F(\gamma^{N_1}, h^{N_1})$  in Section B.3 and the fact that  $\langle f, \eta_0^{N_1, N_2} \rangle = O_p(N_2^{\varphi - \frac{1}{2}})$ , we obtain

$$\begin{aligned}
& \mathbb{E}_{\pi^{N_1, N_2}} [F_1(\gamma, h, l(f))] = \mathbb{E} [F(\gamma^{N_1, N_2}, h^{N_1, N_2})] \\
&+ \mathbb{E} \left[ \left| \left( \langle f, \eta_0^{N_1, N_2} \rangle + M_{\eta, 1, t}^{N_1, N_2} + M_{\eta, 2, t}^{N_1, N_2} + M_{\eta, 3, t}^{N_1, N_2} + O(N_2^{-\gamma_2}) \right) \right. \right. \\
&\quad \left. \left. \times \prod_{i=1}^p m_i(l_{s_i}^{N_1, N_2}(f)) \right| \right] \\
&\leq C \left( \frac{1}{N_2^{1-\gamma_2}} \right) + C \left( \mathbb{E} \left[ \left| M_{\eta, 1, t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \left| M_{\eta, 2, t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \left| M_{\eta, 3, t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} \right) \\
&\quad + C \left( \frac{1}{N_2^{\frac{1}{2}-\varphi}} \right) \\
&\leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\frac{1}{2}-\varphi}} \right).
\end{aligned}$$

Therefore,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_1(\gamma^{N_1}, h^{N_1}, l^{N_1}(f))] = 0$ . Since  $F(\cdot)$  is continuous and  $F(\gamma^{N_1, N_2}, h^{N_1, N_2})$  is uniformly bounded, together with analysis in Section C.1, we have that  $F_1(\cdot)$  is continuous and thus  $F_1(\gamma^{N_1, N_2}, h^{N_1, N_2}, l^{N_1, N_2}(f))$  is uniformly

bounded. Hence,

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_1(\gamma^{N_1}, h^{N_1}, l^{N_1}(f))] = 0.$$

We have shown that any limit point  $\pi^{N_1}$  of a convergent subsequence must be a Dirac measure concentrated  $(\gamma^{N_1}, h^{N_1}, l^{N_1}(f)) \in D_{E'}([0, T])$ , where  $(\mu^{N_1}, h^{N_1}, l^{N_1}(f))$  satisfies equations (36), (42) and  $\gamma_t^{N_1} = \gamma_0^{N_1}$  weakly. By Prokhorov's theorem, the processes  $(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2}, l_t^{N_1, N_2}(f))$  converges in distribution to  $(\gamma_0^{N_1}, h_t^{N_1}, l_t^{N_1}(f))$ .  $\square$

**C.2. Relative compactness of  $K_t^{N_1, N_2}$ .** We begin this section by proving the following lemma for the term  $N_2^\varphi M_t^{N_1, N_2}$ .

**Lemma C.5.** *For any  $N_2 \in \mathbb{N}$  and  $x \in \mathcal{X}$ , there is a constant  $C < \infty$  such that*

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left| N_2^\varphi M_t^{N_1, N_2}(x) \right|^2 \right] \leq \frac{C}{N_2^{1-2\varphi}}.$$

*Proof.* Recall that  $M_t^{N_1, N_2} = M_{1,t}^{N_1, N_2} + M_{2,t}^{N_1, N_2} + M_{3,t}^{N_1, N_2}$ , which are defined in (25) to (27). Let  $\mathfrak{F}_t$  be the  $\sigma$ -algebra generated by  $\gamma_s^{N_1, N_2}$ ,  $M_{1,s}^{N_1, N_2}$ ,  $M_{2,s}^{N_1, N_2}$  and  $M_{1,s}^{N_1, N_2}$  for  $s \leq t$ . Since for any  $t > r$ , we have

$$\begin{aligned} & \mathbb{E} \left[ N_2^\varphi \left( M_{1,t}^{N_1, N_2}(x) - M_{1,r}^{N_1, N_2}(x) \right) \mid \mathfrak{F}_r \right] \\ &= \frac{1}{N_2^{1-\varphi}} \sum_{k=\lfloor N_2 r \rfloor}^{\lfloor N_2 t \rfloor - 1} \mathbb{E} \left[ \left( y_k - g_k^{N_1, N_2}(x_k) \right) \left\langle \sigma \left( Z^{2, N_1}(x_k) \right) \sigma \left( Z^{2, N_1}(x) \right), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \right. \\ & \quad \left. - \int_{\mathcal{X} \times \mathcal{Y}} \left( y - g_k^{N_1, N_2}(x') \right) \left\langle \sigma \left( Z^{2, N_1}(x') \right) \sigma \left( Z^{2, N_1}(x) \right), \tilde{\gamma}_k^{N_1, N_2} \right\rangle \pi(dx', dy) \mid \mathcal{F}_r^{N_1, N_2} \right] \\ &= \frac{1}{N_2^{1-\varphi}} \cdot 0 = 0. \end{aligned}$$

Therefore, we have

$$\begin{aligned} & \mathbb{E} \left[ N_2^\varphi M_{1,t}^{N_1, N_2}(x) \mid \mathfrak{F}_r \right] \\ &= \mathbb{E} \left[ N_2^\varphi \left( M_{1,t}^{N_1, N_2}(x) - M_{1,r}^{N_1, N_2}(x) \right) \mid \mathfrak{F}_r \right] + \mathbb{E} \left[ N_2^\varphi M_{1,r}^{N_1, N_2}(x) \mid \mathfrak{F}_r \right] \\ &= 0 + N_2^\varphi M_{1,r}^{N_1, N_2}(x), \end{aligned}$$

proving the martingale property for the process  $N_2^\varphi M_{1,t}^{N_1, N_2}(x)$  and  $x \in \mathcal{X}$ . Hence, by Lemma B.2 and Doob's martingale inequality, we have

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left| N_2^\varphi M_{1,t}^{N_1, N_2}(x) \right|^2 \right] \leq C N_2^{2\varphi} \mathbb{E} \left[ \left| M_{1,T}^{N_1, N_2}(x) \right|^2 \right] \leq \frac{C}{N_2^{1-2\varphi}},$$

where the constant  $C < \infty$ . Note that since  $\gamma < 1$  and  $\varphi \leq \gamma - \frac{1}{2}$ , we have  $1 - 2\varphi > 0$ .

Similar analysis gives

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left| N_2^\varphi M_{2,t}^{N_1, N_2}(x) \right|^2 \right] \leq \frac{C}{N_2^{1-2\varphi}}, \quad \mathbb{E} \left[ \sup_{t \in [0, T]} \left| N_2^\varphi M_{3,t}^{N_1, N_2}(x) \right|^2 \right] \leq \frac{C}{N_2^{1-2\varphi}}.$$

Hence,

$$\mathbb{E} \left[ \sup_{t \in [0, T]} \left| N_2^\varphi M_t^{N_1, N_2}(x) \right|^2 \right] \leq C \sum_{i=1}^3 \mathbb{E} \left[ \sup_{t \in [0, T]} \left| N_2^\varphi M_{i,t}^{N_1, N_2}(x) \right|^2 \right] \leq \frac{C}{N_2^{1-2\varphi}}.$$

□

The next three lemmas prove relative compactness of the family  $\{K_t^{N_1, N_2}, t \in [0, T]\}_{N_2 \in \mathbb{N}}$  in  $D_{\mathbb{R}^M}([0, T])$ .

**Lemma C.6.** *There exist a constant  $C < \infty$ , such that for each  $x \in \mathcal{X}$ ,*

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{E} \left[ \left| K_t^{N_1, N_2}(x) \right|^2 \right] < C.$$

In particular, for any  $\epsilon > 0$ , there exist a compact subset  $U \subset \mathbb{R}^M$  such that

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{P} \left( K_t^{N_1, N_2} \notin U \right) < \epsilon.$$

*Proof.* By (38) and Cauchy-Schwarz inequality, we have

$$\begin{aligned} & \left| K_t^{N_1, N_2}(x) \right|^2 \\ & \leq C \left\{ (I)^2 + (II)^2 + \left| \Gamma_t^{N_1, N_2} \right|^2 + \left| N_2^{1-\gamma_2+\varphi} \left\langle c\sigma(Z^{2, N_1}(x)), \gamma_0^{N_1, N_2} \right\rangle \right|^2 + \left| N_2^\varphi M_t^{N_1, N_2} \right|^2 \right\} + \\ & \quad + O \left( N_2^{-2(\gamma_2-\varphi)} \right), \end{aligned}$$

where

$$\begin{aligned} (I) &= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| y - h_s^{N_1}(x') \right| \left( \left| \left\langle B_{x,x'}^1(\theta), \eta_s^{N_1, N_2} \right\rangle \right| + \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_{x,x'}^{2,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right| \right) \\ & \quad \pi(dx', dy) ds + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| y - h_s^{N_1}(x') \right| \\ & \quad \cdot \left( \left| \left\langle x x' B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right| \left| \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \right. \\ & \quad \left. + \left| \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \left| \left\langle x x' B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right| \right) \pi(dx', dy) ds \\ (II) &= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1, N_2}(x') \right| \left| \left\langle B_{x,x'}^1(\theta), \gamma_0^{N_1} \right\rangle \right| + \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \pi(dx', dy) ds \\ & \quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1, N_2}(x') \right| \left| \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \left| \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \pi(dx', dy) ds. \end{aligned}$$

By Assumption 2.1, definition of  $\gamma_0^{N_1}$ , and Lemma A.1, there exist some constant  $C < \infty$ , such that

$$\begin{aligned} & \sup_{x, x' \in \mathcal{X}} \left\{ \left| \left\langle B_{x,x'}^1(\theta), \gamma_0^{N_1} \right\rangle \right| + \frac{1}{N_1} \sum_{j=1}^{N_1} \left( \left| \left\langle B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \right. \\ & \quad \left. + \left| \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \left| \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right| \right) \right\} < C. \end{aligned} \tag{44}$$

Then, by the Cauchy-Schwarz inequality and equation (36), we have

$$\left| h_t^{N_1}(x) \right|^2 \leq C \left[ \left( \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |y| \pi(dx', dy) ds \right)^2 + \left( \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| h_s^{N_1}(x') \right| \pi(dx', dy) ds \right)^2 \right],$$

$$\leq C_1 t^2 + C_2 t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |h_s^{N_1}(x')|^2 \pi(dx', dy) ds,$$

which implies that,

$$\sup_{t \in [0, T]} \int_{\mathcal{X} \times \mathcal{Y}} |h_t^{N_1}(x)|^2 \pi(dx, dy) \leq C_1 T^2 + C_2 T \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |h_s^{N_1}(x')|^2 \pi(dx', dy) ds.$$

Therefore, by Grönwall's inequality,

$$\sup_{0 \leq t \leq T} \int_{\mathcal{X} \times \mathcal{Y}} |h_t^{N_1}(x)|^2 \pi(dx, dy) \leq \sup_{0 \leq t \leq T} C_1 T^2 \exp(C_2 T t) < C(T), \quad (45)$$

for some constant  $C(T) < \infty$  depending on  $T$ . By Cauchy-Schwarz inequality and (44), we also have

$$(II)^2 \leq C_3 t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |K_s^{N_1, N_2}(x')|^2 \pi(dx', dy) ds. \quad (46)$$

Since  $\sigma \in C_b^\infty(\mathbb{R})$ , by Lemma C.1, there exist some constant  $C < \infty$  such that

$$\begin{aligned} \mathbb{E} \left[ |\langle B_{x, x'}^1(\theta), \eta_s^{N_1, N_2} \rangle|^2 \right] &< C, \quad \mathbb{E} \left[ \left| \langle B_{x, x'}^{2, j}(\theta), \eta_s^{N_1, N_2} \rangle \right|^2 \right] < C, \\ \mathbb{E} \left[ \left| \langle x x' B_{x'}^{3, j}(\theta), \eta_s^{N_1, N_2} \rangle \right|^2 \right] &< C \end{aligned} \quad (47)$$

for  $t \in [0, T]$ ,  $j = 1, \dots, N_1$ , and  $N_2 \in \mathbb{N}$ . By the Cauchy-Schwarz inequality, equations (45), (47), and Assumption 2.1, we have

$$\begin{aligned} &\mathbb{E}[(I)^2] \\ &\leq C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left( \left| \langle B_{x, x'}^1(\theta), \eta_s^{N_1, N_2} \rangle \right|^2 + \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \langle B_{x, x'}^{2, j}(\theta), \eta_s^{N_1, N_2} \rangle \right|^2 \right) \pi(dx', dy) ds \\ &\quad + C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{N_1} \sum_{j=1}^{N_1} \mathbb{E} \left( \left| \langle x x' B_{x'}^{3, j}(\theta), \eta_s^{N_1, N_2} \rangle \right|^2 + \left| \langle x x' B_{x'}^{3, j}(\theta), \eta_s^{N_1, N_2} \rangle \right|^2 \right) \\ &\quad \pi(dx', dy) ds \\ &\leq C_4 t^2 \end{aligned}$$

Since  $|\Gamma_t^{N_1, N_2}|^2 \leq C \left( |\Gamma_{1, t}^{N_1, N_2}|^2 + |\Gamma_{2, t}^{N_1, N_2}|^2 + |\Gamma_{3, t}^{N_1, N_2}|^2 \right)$ , by Assumption 2.1 and Lemma A.1, we have

$$\begin{aligned} |\Gamma_{1, t}^{N_1, N_2}|^2 &\leq C \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |K_s^{N_1, N_2}(x')|^2 \pi(dx', dy) ds \int_0^t \\ &\quad \int_{\mathcal{X} \times \mathcal{Y}} \left| \langle B_{x, x'}^1(\theta), \gamma_s^{N_1, N_2} - \gamma_0^{N_1} \rangle \right|^2 \pi(dx', dy) ds \\ &\leq C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |K_s^{N_1, N_2}(x')|^2 \pi(dx', dy) ds, \end{aligned}$$

$$\begin{aligned} |\Gamma_{2, t}^{N_1, N_2}|^2 &\leq \\ &\leq C \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |K_s^{N_1, N_2}(x')|^2 \pi(dx', dy) ds \int_0^t \end{aligned}$$

$$\begin{aligned}
& \int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_{x,x'}^{2,j}(\theta), \gamma_s^{N_1,N_2} - \gamma_0^{N_1} \right\rangle \right|^2 \pi(dx', dy) ds \\
& \leq Ct \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1,N_2}(x') \right|^2 \pi(dx', dy) ds, \\
& \left| \Gamma_{3,t}^{N_1,N_2}(x) \right|^2 \\
& \leq C \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1,N_2}(x') \right|^2 \pi(dx', dy) ds \\
& \quad \cdot \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle xx' B_x^{3,j}(\theta), \gamma_s^{N_1,N_2} - \gamma_0^{N_1} \right\rangle \right|^2 \left| \left\langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1,N_2} \right\rangle \right|^2 \pi(dx', dy) ds \\
& \quad + C \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1,N_2}(x') \right|^2 \pi(dx', dy) ds \\
& \quad \cdot \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right|^2 \left| \left\langle xx' B_{x'}^{3,j}(\theta), \gamma_s^{N_1,N_2} - \gamma_0^{N_1} \right\rangle \right|^2 \pi(dx', dy) ds \\
& \quad + C \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| y - h_s^{N_1}(x') \right|^2 \pi(dx', dy) ds \\
& \quad \cdot \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle xx' B_x^{3,j}(\theta), \gamma_s^{N_1,N_2} - \gamma_0^{N_1} \right\rangle \right|^2 \left| \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1,N_2} \right\rangle \right|^2 \pi(dx', dy) ds \\
& \leq Ct \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1,N_2}(x') \right|^2 \pi(dx', dy) ds \\
& \quad + Ct \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1,N_2} \right\rangle \right|^2 \pi(dx', dy) ds.
\end{aligned}$$

Hence,

$$\begin{aligned}
\left| \Gamma_t^{N_1,N_2} \right|^2 & \leq C_5 t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1,N_2}(x') \right|^2 \pi(dx', dy) ds \\
& \quad + C_5 t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1,N_2} \right\rangle \right|^2 \pi(dx', dy) ds
\end{aligned} \tag{48}$$

By (46) to (48), and the definition of  $\pi(dx, dy)$ , we see that

$$\begin{aligned}
\mathbb{E} \left( \left| K_t^{N_1,N_2}(x) \right|^2 \right) & \leq C \left\{ (C_4 + C_5)t^2 + \frac{C_3 + C_5}{M} t \int_0^t \sum_{x' \in \mathcal{X}} \mathbb{E} \left( \left| K_t^{N_1,N_2}(x') \right|^2 \right) ds \right. \\
& \quad \left. + \mathbb{E} \left( \left| N_2^{1-\gamma_2+\varphi} \left\langle c\sigma(Z^{2,N_1}(x)), \gamma_0^{N_1,N_2} \right\rangle \right|^2 \right) \right. \\
& \quad \left. + \mathbb{E} \left( \left| N_2^\varphi M_t^{N_1,N_2} \right|^2 \right) + O \left( N_2^{-2(\gamma_2-\varphi)} \right) \right\}.
\end{aligned}$$

Summing both side of the above inequality over all  $x \in \mathcal{X}$ , where  $\mathcal{X}$  is a fixed data set of size  $M$  gives

$$\begin{aligned}
\sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| K_t^{N_1, N_2}(x) \right|^2 \right) &\leq CMT^2 + CT \int_0^t \sum_{x' \in \mathcal{X}} \mathbb{E} \left( \left| K_s^{N_1, N_2}(x') \right|^2 \right) ds \\
&+ \sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| N_2^{1-\gamma_2+\varphi} \left\langle c\sigma(Z^{2, N_1}(x)), \gamma_0^{N_1, N_2} \right\rangle \right|^2 \right) \\
&+ \sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| N_2^\varphi M_t^{N_1, N_2} \right|^2 \right) + O \left( N_2^{-2(\gamma_2-\varphi)} \right). \tag{49}
\end{aligned}$$

Since for  $\varphi \leq \gamma_2 - \frac{1}{2}$ ,  $2(\gamma_2 - \varphi) \geq 1$ , we have

$$\mathbb{E} \left( \left| N_2^{1-\gamma_2+\varphi} \left\langle c\sigma(Z^{2, N_1}(x)), \gamma_0^{N_1, N_2} \right\rangle \right|^2 \right) \leq \frac{C}{N_2^{2(\gamma_2-\varphi)}} \sum_{i=1}^{N_2} \mathbb{E} \left( |C_0^i|^2 \right) \leq C.$$

Therefore, by applying Grönwall's inequality to equation (49) and using Lemma C.5,

$$\sum_{x \in \mathcal{X}} \mathbb{E} \left( \left| K_t^{N_1, N_2}(x) \right|^2 \right) \leq C(M)T^2 \exp \left[ \tilde{C}Tt \right],$$

where  $C(M), \tilde{C}$  are some finite constants. Hence, for any  $x \in \mathcal{X}$ , there exist  $C < \infty$  such that

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{E} \left[ \left| K_t^{N_1, N_2}(x) \right|^2 \right] < C(M)T^2 \exp \left[ \tilde{C}T^2 \right] \leq C.$$

By Markov's inequality, the compact containment condition for  $K_t^{N_1, N_2}$  follows, concluding the proof of the lemma.  $\square$

We next establish the regularity of the process  $K_t^{N_1, N_2}$  in  $D_{\mathbb{R}^M}([0, T])$ . For the purpose of this lemma, we denote  $q(z_1, z_2) = \min\{\|z_1 - z_2\|_{l^1}, 1\}$  for  $z_1, z_2 \in \mathbb{R}^M$ .

**Lemma C.7.** *For any  $\delta \in (0, 1)$ , there is a constant  $C < \infty$  such that for  $0 \leq u \leq \delta$ ,  $0 \leq v \leq \delta \wedge t$ , and  $t \in [0, T]$ ,*

$$\mathbb{E} \left[ q \left( K_{t+u}^{N_1, N_2}, K_t^{N_1, N_2} \right) q \left( K_t^{N_1, N_2}, K_{t-v}^{N_1, N_2} \right) | \mathcal{F}_t^{N_1, N_2} \right] \leq C\delta + \frac{C}{N_2^{1-\varphi}}.$$

*Proof.* For  $0 \leq s < t \leq T$ , the leading terms in equation (38) gives

$$\begin{aligned}
&\left| K_t^{N_1, N_2}(x) - K_s^{N_1, N_2}(x) \right| \\
&\leq \int_s^t \int_{\mathcal{X} \times \mathcal{Y}} \left| y - h_\tau^{N_1}(x') \right| \left( \left| \left\langle B_{x, x'}^1(\theta), \eta_\tau^{N_1, N_2} \right\rangle \right| + \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_{x, x'}^{2, j}(\theta), \eta_\tau^{N_1, N_2} \right\rangle \right| \right) \\
&\quad \pi(dx', dy) d\tau + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_s^t \int_{\mathcal{X} \times \mathcal{Y}} \left| y - h_\tau^{N_1}(x') \right| \\
&\quad \cdot \left( \left| \left\langle x x' B_x^{3, j}(\theta), \eta_\tau^{N_1, N_2} \right\rangle \right| \left| \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \right| + \left| \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \right| \left| \left\langle x x' B_{x'}^{3, j}(\theta), \eta_\tau^{N_1, N_2} \right\rangle \right| \right) \\
&\quad \pi(dx', dy) d\tau \\
&+ \int_s^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_\tau^{N_1, N_2}(x') \right| \left| \left\langle B_{x, x'}^1(\theta), \gamma_0^{N_1} \right\rangle \right| + \frac{1}{N_1} \sum_{j=1}^{N_1} \left| \left\langle B_{x, x'}^{2, j}(\theta), \gamma_0^{N_1} \right\rangle \right| \pi(dx', dy) d\tau \\
&+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_s^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_\tau^{N_1, N_2}(x') \right| \left| \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \right| \left| \left\langle x x' B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \right| \pi(dx', dy) d\tau
\end{aligned}$$

$$+ \left| \Gamma_t^{N_1, N_2}(x) - \Gamma_s^{N_1, N_2}(x) \right| + N_2^\varphi \left| M_t^{N_1, N_2}(x) - M_s^{N_1, N_2}(x) \right|.$$

Taking expectation on both sides of the above inequality, by Assumption 2.1, Lemma A.1, and analysis in Lemmas C.5 and C.6, we have for  $0 \leq t - s \leq \delta < 1$

$$\begin{aligned} & \mathbb{E} \left[ \left| K_t^{N_1, N_2}(x) - K_s^{N_1, N_2}(x) \right| \middle| \mathcal{F}_s^{N_1, N_2} \right] \\ & \leq C(t-s) + C_1 \int_s^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| K_\tau^{N_1, N_2}(x') \right| \middle| \mathcal{F}_s^{N_1, N_2} \right] \pi(dx', dy) d\tau \\ & \quad + C \mathbb{E} \left[ \left| N_2^\varphi \left( M_t^{N_1, N_2}(x) - M_s^{N_1, N_2}(x) \right) \right|^2 \middle| \mathcal{F}_s^{N_1, N_2} \right]^{\frac{1}{2}} \\ & \leq C\delta + \frac{C}{N_2^{1-\varphi}}. \end{aligned}$$

Note that

$$\mathbb{E} \left[ \left| N_2^\varphi \left( M_t^{N_1, N_2}(x) - M_s^{N_1, N_2}(x) \right) \right|^2 \middle| \mathcal{F}_s^{N_1, N_2} \right] \leq \frac{C\delta}{N_2^{1-2\varphi}} + \frac{C}{N_2^{2-2\varphi}},$$

following an analysis similar to Lemma 3.1 of [39]. Since  $x \in \mathcal{X}$  is arbitrary, the statement of the lemma is then implied.  $\square$

By combining Lemmas C.6 and C.7, we have that the sequence of processes  $\{K_t^{N_1, N_2}, t \in [0, T]\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{\mathbb{R}^M}([0, T])$ , which follows from Theorem 8.6 of Chapter 3 of [11].

**C.3. Convergence of  $K_t^{N_1, N_2}$ .** Denote  $l_{1,t}^{N_1, N_2} = l_t^{N_1, N_2}(B_{x,x'}^1(\theta))$ ,  $l_{2,t}^{N_1, N_2}$  and  $l_{3,t}^{N_1, N_2}$  as  $N_1$ -dimensional vectors with  $j$ -th entry being  $l_t^{N_1, N_2}(B_{x,x'}^{2,j}(\theta))$  and  $l_t^{N_1, N_2}(B_x^{3,j}(\theta))$ , respectively. We also let  $l_{1,t}^{N_1}, l_{2,t}^{N_1}, l_{3,t}^{N_1}$  be the corresponding limits for  $l_{1,t}^{N_1, N_2}, l_{2,t}^{N_1, N_2}, l_{3,t}^{N_1, N_2}$  as  $N_2 \rightarrow \infty$ . Recall that from Section C.1, for  $\gamma_2 \in (1/2, 1)$ , if  $\varphi < 1 - \gamma_2$ ,  $l_{1,t}^{N_1} = 0, l_{2,t}^{N_1} = l_{3,t}^{N_1} = 0$ , and if  $\varphi = 1 - \gamma_2$ ,  $l_{1,t}^{N_1}, l_{2,t}^{N_1}, l_{3,t}^{N_1}$  are given by (42) for appropriate definitions of the function  $f$ .

In this section, we show that the processes  $(\gamma_t^{N_1, N_2}, h_t^{N_1, N_2}, l_{1,t}^{N_1, N_2}, l_{2,t}^{N_1, N_2}, l_{3,t}^{N_1, N_2}, K_t^{N_1, N_2})$  converges in distribution in  $D_{E_1}([0, T])$  to  $(\gamma_0^{N_1}, h_t^{N_1}, l_{1,t}^{N_1}, l_{2,t}^{N_1}, l_{3,t}^{N_1}, K_t^{N_1})$ , where  $E_1 = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^{N_1} \times \mathbb{R}^{N_1} \times \mathbb{R}^M$ , and  $K_t^{N_1}$  satisfies either of the following evolution equations:

*Case 1.* When  $\gamma_2 \in (\frac{1}{2}, \frac{3}{4})$  and  $\varphi \leq \gamma_2 - \frac{1}{2}$ , or when  $\gamma_2 \in [\frac{3}{4}, 1)$  and  $\varphi < 1 - \gamma_2 \leq \gamma_2 - \frac{1}{2}$  then  $K_t^{N_1}(x)$  is given by (16).

*Case 2.* When  $\gamma_2 \in [\frac{3}{4}, 1)$  and  $\varphi = 1 - \gamma_2$  then  $K_t^{N_1}(x)$  satisfies (17).

By Lemmas B.6, C.1, C.2, and Section C.2,  $\{\gamma^{N_1, N_2}, h^{N_1, N_2}, l_1^{N_1, N_2}, l_2^{N_1, N_2}, l_3^{N_1, N_2}, K^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{E_1}([0, T])$ . Denote  $\pi^{N_1, N_2} \in \mathcal{M}(D_{E_1}([0, T]))$  the probability measure corresponding to  $(\gamma^{N_1, N_2}, h^{N_1, N_2}, l_1^{N_1, N_2}, l_2^{N_1, N_2}, l_3^{N_1, N_2}, K^{N_1, N_2})$ . We now show that any limit point  $\pi^{N_1}$  of a convergence subsequence  $\pi^{N_1, N_{2k}}$  is a Dirac measure concentrated on  $(\gamma^{N_1}, h^{N_1}, l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1})$ , where  $(\gamma^{N_1}, h^{N_1})$  satisfies equation (36) and  $(l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1})$  satisfies Lemma C.4, equations (16), or (17) for different values of  $\gamma_2$  and  $\varphi$ .

*Case 1.* When  $\gamma_2 \in (\frac{1}{2}, \frac{3}{4})$  and  $\varphi_2 \leq \gamma_2 - \frac{1}{2}$ , or when  $\gamma \in [\frac{3}{4}, 1)$  and  $\varphi < 1 - \gamma_2 \leq \gamma_2 - \frac{1}{2}$ , for any  $t \in [0, T]$ ,  $m_1^1, \dots, m_p^1 \in C_b(\mathbb{R})$ ,  $m_1^{i,j}, \dots, m_p^{i,j} \in C_b(\mathbb{R})$  for

$i = 2, 3, j = 1, \dots, N_1$ ,  $z_1, \dots, z_p \in C_b(\mathbb{R}^M)$ , and  $0 \leq s_1 < \dots < s_p \leq t$ , we define  $F_2 : D_{E_1}([0, T]) \rightarrow \mathbb{R}_+$  as

$$\begin{aligned}
& F_2(\gamma, h, l_1, l_2, l_3, K) \\
&= F(\gamma^{N_1}, h^{N_1}) + \left| \left( l_{1,t}^{N_1} - 0 \right) \times m_1^1(l_{1,s_1}^{N_1}) \times \dots \times m_p^1(l_{1,s_p}^{N_1}) \right| + \\
&+ \sum_{i=2}^3 \sum_{j=1}^{N_1} \left| \left( l_{i,t}^{N_1,j} - 0 \right) \times m_1^{i,j}(l_{i,s_1}^{N_1,j}) \times \dots \times m_p^{i,j}(l_{i,s_p}^{N_1,j}) \right| + \\
&+ \sum_{x \in \mathcal{X}} \left| \left\{ K_t^{N_1}(x) - K_0^{N_1}(x) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \right. \right. \\
&\quad \left. \left. \left[ l_t^{N_1}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_t^{N_1}(B_{x,x'}^{2,j}(\theta)) \right] \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' l_t^{N_1}(B_x^{3,j}(\theta)) \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle l_t^{N_1}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right\} \right. \\
&\quad \left. \times z_1(K_{s_1}^{N_1}) \times \dots \times z_p(K_{s_p}^{N_1}) \right|
\end{aligned}$$

where  $F(\gamma^{N_1}, h^{N_1})$  is as given in equation (B.3) and  $l_{i,t}^{N_1,j}$  is the  $j$ -th element of the  $N_1$ -dimensional vector  $l_{i,t}^{N_1}$  for  $i = 2, 3$ . We now note that for any  $x \in \mathcal{X}$ , by equation (38),

$$\begin{aligned}
& K_t^{N_1, N_2}(x) - K_0^{N_1, N_2}(x) \\
&= - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \left[ l_t^{N_1, N_2}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_t^{N_1, N_2}(B_{x,x'}^{2,j}(\theta)) \right] \pi(dx', dy) ds \\
&\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) x x' l_t^{N_1, N_2}(B_x^{3,j}(\theta)) \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle l_t^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
&\quad + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&= (I) + (II) + (III) + (IV) + (V) + \Gamma_t^{N_1, N_2}(x) + N_2^\varphi M_t^{N_1, N_2}(x) + O(N_2^{-\gamma_2 + \varphi}),
\end{aligned}$$

where terms (I), (II), (III), (IV), (V) will be specified and analyzed as follows. We see that term (I) satisfies

$$\begin{aligned}
(I) &= \\
&= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left[ \left( y - h_s^{N_1}(x') \right) - \left( y - h_s^{N_1, N_2}(x') \right) \right] \\
&\quad \left[ l_t^{N_1, N_2} \left( B_{x, x'}^1(\theta) \right) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_t^{N_1, N_2} \left( B_{x, x'}^{2, j}(\theta) \right) \right] \pi(dx', dy) ds \\
&= \frac{1}{N_1^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left[ l_t^{N_1, N_2} \left( B_{x, x'}^1(\theta) \right) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_t^{N_1, N_2} \left( B_{x, x'}^{2, j}(\theta) \right) \right] \pi(dx', dy) ds \\
&= - \left( \Gamma_{1, t}^{N_1, N_2}(x) + \Gamma_{2, t}^{N_1, N_2}(x) \right).
\end{aligned}$$

Term (II) can be rearranged into

$$\begin{aligned}
(II) &= \\
&= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta), \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&= \frac{1}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta), \eta_0^{N_1} \right\rangle \pi(dx', dy) ds,
\end{aligned}$$

and by the Cauchy-Schwarz inequality, Lemmas C.1 and C.6, for any  $t \in [0, T]$ ,

$$\begin{aligned}
&\mathbb{E} \left( \left| \frac{1}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta), \eta_0^{N_1} \right\rangle \pi(dx', dy) ds \right| \right) \\
&\leq \frac{C}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left( \left| K_s^{N_1, N_2}(x') \right| \left| \left\langle B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta), \eta_0^{N_1} \right\rangle \right| \right) \pi(dx', dy) ds \\
&\leq \frac{C}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left( \left| K_s^{N_1, N_2}(x') \right|^2 \right)^{\frac{1}{2}} \mathbb{E} \left( \left| \left\langle B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta), \eta_0^{N_1} \right\rangle \right|^2 \right)^{\frac{1}{2}} \\
&\quad \pi(dx', dy) ds \\
&\leq \frac{C(T)}{N_2^\varphi},
\end{aligned} \tag{50}$$

where  $C(T) < \infty$  is some finite constant depending on  $T$ .

We discuss terms (III) and (IV) together. Since

$$\begin{aligned}
(III) &= \\
&= \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' l_t^{N_1, N_2} \left( B_x^{3, j}(\theta) \right) \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' l_t^{N_1, N_2} \left( B_x^{3, j}(\theta) \right) \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&= - \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' l_t^{N_1, N_2} \left( B_x^{3, j}(\theta) \right) l_t^{N_1, N_2} \left( B_{x'}^{3, j}(\theta) \right) \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_t^{N_1, N_2} \left( B_x^{3,j}(\theta) \right) \left\langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' l_t^{N_1, N_2} \left( B_x^{3,j}(\theta) \right) \left\langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} - \gamma_0^{N_1, N_2} \right\rangle \\
& \pi(dx', dy) ds,
\end{aligned}$$

and

$$\begin{aligned}
& (IV) \\
& = \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle l_t^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \pi(dx', dy) ds \\
& - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle l_t^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \pi(dx', dy) ds \\
& = \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle l_t^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle l_t^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \\
& \pi(dx', dy) ds,
\end{aligned}$$

one has

$$\begin{aligned}
& (III) + (IV) = \\
& = -\Gamma_{3,t}^{N_1, N_2}(x) + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' l_t^{N_1, N_2} \left( B_x^{3,j}(\theta) \right) \\
& \left\langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} - \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle l_t^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \\
& \pi(dx', dy) ds \\
& = -\Gamma_{3,t}^{N_1, N_2}(x) + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' l_t^{N_1, N_2} \\
& \left( B_x^{3,j}(\theta) \right) l_t^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \pi(dx', dy) ds \\
& - \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' l_t^{N_1, N_2} \left( B_x^{3,j}(\theta) \right) l_0^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \pi(dx', dy) ds \\
& + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' l_0^{N_1, N_2} \left( B_x^{3,j}(\theta) \right) l_t^{N_1, N_2} \left( B_{x'}^{3,j}(\theta) \right) \\
& \pi(dx', dy) ds. \tag{51}
\end{aligned}$$

Since by Lemmas B.1 and C.1,

$$\begin{aligned}
& \mathbb{E} \left( \left| \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' l_t^{N_1, N_2} (B_x^{3,j}(\theta)) l_t^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right| \right) \\
& \leq \mathbb{E} \left( \left| \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' l_t^{N_1, N_2} (B_x^{3,j}(\theta)) l_t^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right|^2 \right) \\
& \leq C(T) \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| l_t^{N_1, N_2} (B_x^{3,j}(\theta)) l_t^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \right|^2 \right] \pi(dx', dy) ds \\
& \leq C(T) \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| l_t^{N_1, N_2} (B_x^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| l_t^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{2}} \pi(dx', dy) ds \\
& \leq C(T), \\
& \mathbb{E} \left[ \left| \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) x x' l_0^{N_1, N_2} (B_x^{3,j}(\theta)) l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right| \right] \\
& \leq \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| (y - h_s^{N_1, N_2}(x')) x x' \right| \left| l_0^{N_1, N_2} (B_x^{3,j}(\theta)) \right| \left| l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \right| \right] \pi(dx', dy) ds \\
& \leq \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| (y - h_s^{N_1, N_2}(x')) x x' \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| l_0^{N_1, N_2} (B_x^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{4}} \mathbb{E} \left[ \left| l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{4}} \\
& \quad \pi(dx', dy) ds \\
& \leq C(T),
\end{aligned}$$

the expectation of the last three terms in (51) is bounded by  $O(N_2^{-\varphi})$ . Lastly, for term (V), we have

$$\begin{aligned}
& (V) \\
& = \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& \quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
& = \frac{1}{N_1 N_2^{2\varphi}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_0^{N_1, N_2} (B_x^{3,j}(\theta)) l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& \quad + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_0^{N_1, N_2} (B_x^{3,j}(\theta)) \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
& \quad + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds.
\end{aligned}$$

By the Cauchy-Schwarz inequality, Lemmas C.1 and C.6, for any  $t \in [0, T]$ ,

$$\begin{aligned}
& \mathbb{E} \left( \left| \frac{1}{N_1 N_2^{2\varphi}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_0^{N_1, N_2} (B_x^{3,j}(\theta)) l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right| \right) \\
& \leq \frac{1}{N_1 N_2^{2\varphi}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| K_s^{N_1, N_2}(x') l_0^{N_1, N_2} (x x' B_x^{3,j}(\theta)) l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \right| \right] \pi(dx', dy) ds \\
& \leq \frac{1}{N_1 N_2^{2\varphi}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| K_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| l_0^{N_1, N_2} (x x' B_x^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{4}} \\
& \quad \mathbb{E} \left[ \left| l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{4}} \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
&\leq \frac{C(T)}{N_2^{2\varphi}}, \\
&\mathbb{E} \left( \left| \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_0^{N_1, N_2} (B_x^{3,j}(\theta)) \langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \rangle \pi(dx', dy) ds \right| \right) \\
&\leq \mathbb{E} \left( \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \sup_{x'} \left| \langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \rangle \right| \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1, N_2}(x') l_0^{N_1, N_2} (B_x^{3,j}(\theta)) \right| \pi(dx', dy) ds \right) \\
&\leq \frac{C}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| K_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| l_0^{N_1, N_2} (B_x^{3,j}(\theta)) \right|^2 \right]^{\frac{1}{2}} \pi(dx', dy) ds \\
&\leq \frac{C(T)}{N_2^\varphi},
\end{aligned}$$

and similarly,

$$\begin{aligned}
&\mathbb{E} \left( \left| \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \langle B_x^{3,j}(\theta), \gamma_0^{N_1} \rangle l_0^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right| \right) \quad (52) \\
&\leq \frac{C(T)}{N_2^\varphi}.
\end{aligned}$$

By equations (50) to (52), the analysis in Sections B.3 and C.1, and Lemma C.5, we have

$$\begin{aligned}
&\mathbb{E}_{\pi^{N_1, N_2}} \left[ F_2(\gamma^{N_1}, h^{N_1}, l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1}) \right] \\
&= \mathbb{E}_{\pi^{N_1, N_2}} \left[ F(\gamma^{N_1}, h^{N_1}) \right] + \mathbb{E} \left[ \left| \left( l_{1,t}^{N_1, N_2} - 0 \right) \times \prod_{n=1}^p m_n^1(l_{1,s_n}^{N_1, N_2}) \right| \right] \\
&+ \sum_{i=2}^3 \sum_{j=1}^{N_1} \mathbb{E} \left[ \left| \left( l_{i,t}^{N_1, N_2, j} - 0 \right) \times \prod_{n=1}^p m_n^{i,j}(l_{i,s_n}^{N_1, N_2, j}) \right| \right] \\
&+ \sum_{x \in \mathcal{X}} \mathbb{E} \left\{ \left| \left( K_t^{N_1, N_2}(x) - K_0^{N_1, N_2}(x) \right. \right. \right. \\
&\quad \left. \left. \left. - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) l_t^{N_1, N_2} (B_{x,x'}^1(\theta)) \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) l_t^{N_1, N_2} (B_{x,x'}^{2,j}(\theta)) \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' l_t^{N_1, N_2} (B_x^{3,j}(\theta)) \langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) x x' \langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \rangle l_t^{N_1, N_2} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. \times \prod_{i=1}^p z_i(K_{s_i}^{N_1, N_2}) \right\} \right. \right. \\
&\leq C \left( \frac{1}{N_2^{1-\gamma_2}} \right) + C \left( \frac{1}{N_2^{1-\gamma_2-\varphi}} + \frac{1}{N_2^{1-\varphi}} + \frac{1}{N_2^{\frac{1}{2}-\varphi}} \right) + C \mathbb{E} \left[ \left| N_2^\varphi M_t^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}}
\end{aligned}$$

$$+ C \left( \frac{1}{N_2^{\gamma_2 - \varphi}} + \frac{1}{N_2^\varphi} \right) \\ \leq C \left( \frac{1}{N_2^{1 - \gamma_2 - \varphi}} + \frac{1}{N_2^{\frac{1}{2} - \varphi}} + \frac{1}{N_2^\varphi} \right).$$

Therefore,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}}[F_2(\gamma, h, l_1, l_2, l_3, K)] = 0$ . Since  $F(\cdot)$  is continuous and  $F(\gamma^{N_1, N_2}, h^{N_1, N_2})$  is uniformly bounded, together with analysis in Sections C.1 and C.2, we have that  $F_2(\cdot)$  is continuous and  $F_2(\gamma^{N_1, N_2}, h^{N_1, N_2}, l_1^{N_1, N_2}, l_2^{N_1, N_2}, l_3^{N_1, N_2}, K^{N_1, N_2})$  is uniformly bounded. Hence, by weak convergence we have

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}}[F_2(\gamma, h, l_1, l_2, l_3, K)] = \mathbb{E}_{\pi^{N_1}}[F_2(\gamma, h, l_1, l_2, l_3, K)] = 0.$$

We have shown that any limit point  $\pi^{N_1}$  of a convergence sequence must be a Dirac measure concentrated  $(\gamma^{N_1}, h^{N_1}, l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1})$ , which satisfies equation (36),  $l_i^{N_1} = 0$  for  $i = 1, 2, 3$ , and equation (16). Since the solutions to equations (36) and (16) are unique, the processes in consideration converges in distribution to  $(\gamma_0^{N_1}, h^{N_1}, 0, 0, 0, K^{N_1})$  by Prokhorov's theorem.

Case 2. When  $\gamma_2 \in [\frac{3}{4}, 1)$  and  $\varphi = 1 - \gamma_2$ , for any  $t \in [0, T]$ ,  $m_1^1, \dots, m_p^1 \in C_b(\mathbb{R})$ ,  $m_1^{i,j}, \dots, m_p^{i,j} \in C_b(\mathbb{R})$  for  $i = 2, 3, j = 1, \dots, N_1$ ,  $z_1, \dots, z_p \in C_b(\mathbb{R}^M)$ , and  $0 \leq s_1 < \dots < s_p \leq t$ , we define  $F_3(\gamma, h, l_1, l_2, l_3, K) : D_{E_1}([0, T]) \rightarrow \mathbb{R}_+$  as

$$\begin{aligned}
&= F(\gamma, h, l_1, l_2, l_3, K) \\
&= F(\gamma^{N_1}, h^{N_1}) + \left| F_\eta(l_{1,t}^{N_1}) \times m_1^1(l_{1,s_1}^{N_1}) \times \cdots \times m_p^1(l_{1,s_p}^{N_1}) \right| \\
&\quad + \sum_{i=2}^3 \sum_{j=1}^{N_1} \left| F_\eta(l_{i,t}^{N_1,j}) \times m_1^{i,j}(l_{i,s_1}^{N_1,j}) \times \cdots \times m_p^{i,j}(l_{i,s_p}^{N_1,j}) \right| \\
&\quad + \sum_{x \in \mathcal{X}} \left| \left\{ K_t^{N_1}(x) - K_0^{N_1}(x) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \right. \right. \\
&\quad \left. \left. \left[ l_t^{N_1}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_t^{N_1}(B_{x,x'}^{2,j}(\theta)) \right] \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' l_t^{N_1}(B_x^{3,j}(\theta)) \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle l_t^{N_1}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right. \right. \\
&\quad \left. \left. \times z_1(K_{s_1}^{N_1}) \times \cdots \times z_p(K_{s_p}^{N_1}) \right| \right\} \quad (53)
\end{aligned}$$

where  $F(\gamma^{N_1}, h^{N_1})$  is as given in equation (B.3),  $l_{i,t}^{N_1,j}$  is the  $j$ -th element of the  $N_1$ -dimensional vector  $l_{i,t}^{N_1}$  for  $i = 2, 3$ , and

$$\begin{aligned} F_\eta(l_t^{N_1}(f)) &= l_t^{N_1}(f) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \partial_c f(\theta) \sigma(Z^{2,N_1}(x')), \gamma_s^{N_1} \right\rangle \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle c\sigma'(Z^{2,N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_s^{N_1} \right\rangle \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\ &\quad \left\langle \left\langle c\sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \gamma_s^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_s^{N_1} \right\rangle \pi(dx', dy) ds. \end{aligned}$$

By equations (50) to (52), Lemmas C.4 and C.5, and the analysis in Section B.3, we obtain

$$\mathbb{E}_{\pi^{N_1, N_2}} [F_3(\gamma, h, l_1, l_2, l_3, K)] \leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\frac{1}{2}-\varphi}} + \frac{1}{N_2^{1-\varphi}} + \frac{1}{N_2^{\gamma_2-\varphi}} + \frac{1}{N_2^\varphi} \right).$$

Therefore,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_3(\gamma, h, l_1, l_2, l_3, K)] = 0$ . By analysis in Sections B.2, C.1 and C.2, we have that  $F_3(\cdot)$  is continuous and  $F_3(\gamma^{N_1, N_2}, h^{N_1, N_2}, l_1^{N_1, N_2}, l_2^{N_1, N_2}, l_3^{N_1, N_2}, K^{N_1, N_2})$  is uniformly bounded. Hence,

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_3(\gamma, h, l_1, l_2, l_3, K)] = \mathbb{E}_{\pi^{N_1}} [F_3(\gamma, h, l_1, l_2, l_3, K)] = 0.$$

We have shown that any limit point  $\pi^{N_1}$  of a convergence sequence must be a Dirac measure concentrated  $(\gamma^{N_1}, h^{N_1}, l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1}) \in D_{E_1}([0, T])$ , which satisfies equation (36), (42), and (17). Since the solutions to equations (36) and (17) are unique, by Prokhorov's theorem, the processes

$(\gamma^{N_1, N_2}, h^{N_1, N_2}, l_1^{N_1, N_2}, l_2^{N_1, N_2}, l_3^{N_1, N_2}, K^{N_1, N_2})$  converges in distribution to  $(\gamma^{N_1}, h^{N_1}, l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1})$ .

**Appendix D. Proof of Theorem 2.9.** For  $\gamma_2 \in (\frac{3}{4}, 1)$ ,  $\varphi = 1 - \gamma_2$ , we can further look at the fluctuation process  $\Psi_t^{N_1, N_2} = N_2^{\zeta-\varphi} (K_t^{N_1, N_2} - K_t^{N_1})$ , for  $\zeta > \varphi$ . The evolution of  $\Psi_t^{N_1, N_2}(x)$  can be written as

$$\begin{aligned} \Psi_t^{N_1, N_2}(x) &= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) N_2^{\zeta-\varphi} \left[ l_s^{N_1, N_2} (B_{x, x'}^1(\theta)) - l_s^{N_1} (B_{x, x'}^1(\theta)) \right] \pi(dx', dy) ds \\ &\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) N_2^{\zeta-\varphi} \left[ l_s^{N_1, N_2} (B_{x, x'}^{2,j}(\theta)) - l_s^{N_1} (B_{x, x'}^{2,j}(\theta)) \right] \pi(dx', dy) ds \\ &\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) N_2^{\zeta-\varphi} \left[ l_s^{N_1, N_2} (B_{x, x'}^{3,j}(\theta)) - l_s^{N_1} (B_{x, x'}^{3,j}(\theta)) \right] \\ &\quad \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\ &\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle x x' B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle N_2^{\zeta-\varphi} \\ &\quad \left[ l_s^{N_1, N_2} (B_{x'}^{3,j}(\theta)) - l_s^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds \end{aligned} \tag{54}$$

$$\begin{aligned}
& - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \left\langle B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
& - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
& + N_2^{\zeta - \varphi} \Gamma_t^{N_1, N_2}(x) + \Psi_0^{N_1, N_2}(x) + N_2^{\zeta} M_t^{N_1, N_2} + O(N_2^{-\gamma_2 + \zeta}),
\end{aligned}$$

where  $\Psi_0^{N_1, N_2}(x) = N_2^{1-\gamma_2 + \zeta} \left\langle c\sigma(Z^{2, N_1}(x)), \tilde{\gamma}_0^{N_1, N_2} \right\rangle$ , and  $\Gamma_t^{N_1, N_2}$  and  $M_t^{N_1, N_2}$  are as given in Sections C and B.1. We see that if  $\zeta \leq \gamma_2 - \frac{1}{2}$ , the last two remainder terms in equation (54) converge to zero as  $N_2 \rightarrow \infty$  by the similar analysis in Lemma C.5. In addition, if  $\zeta = \gamma_2 - \frac{1}{2}$ ,  $\Psi_0^{N_1, N_2}(x) = \left\langle c\sigma(Z^{2, N_1}(x)), \sqrt{N_2} \tilde{\gamma}_0^{N_1, N_2} \right\rangle \xrightarrow{d} \mathcal{G}^{N_1}(x)$  where  $\mathcal{G}^{N_1}(x)$  is the Gaussian random variable defined in (9). For any fixed  $f \in C_b^3(\mathbb{R}^{1+N_1(1+d)})$ , let  $L_t^{N_1, N_2}(f) = N_2^{\zeta - \varphi} [l_t^{N_1, N_2}(f) - l_t^{N_1}(f)]$ . Its the evolution can be written as

$$\begin{aligned}
L_t^{N_1, N_2}(f) &= N_2^{\zeta - \varphi} [l_t^{N_1, N_2}(f) - l_0^{N_1, N_2}(f) - l_t^{N_1}(f) + l_0^{N_1, N_2}(f)] \\
&= N_2^{\zeta - \varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad - N_2^{\zeta - \varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{N_2^{\zeta - \varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad - \frac{N_2^{\zeta - \varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad + \frac{N_2^{\zeta - \varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \\
&\quad \quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad - \frac{N_2^{\zeta - \varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1} \right\rangle \\
&\quad \quad \pi(dx', dy) ds \\
&\quad + N_2^{\zeta} \left\langle f, \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle + N_2^{\zeta - \varphi} M_{\eta, 1, t}^{N_1, N_2} + N_2^{\zeta - \varphi} M_{\eta, 2, t}^{N_1, N_2} + N_2^{\zeta - \varphi} M_{\eta, 3, t}^{N_1, N_2} + O(N_2^{-\gamma_2 + \zeta - \varphi}) \\
&= (I)_L + (II)_L + (III)_L + \Gamma_{L, t}^{N_1, N_2} + N_2^{\zeta} \left\langle f, \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle \\
&\quad + N_2^{\zeta - \varphi} (M_{\eta, 1, t}^{N_1, N_2} + M_{\eta, 2, t}^{N_1, N_2} + M_{\eta, 3, t}^{N_1, N_2}) + O(N_2^{-\gamma_2 + \zeta - \varphi}),
\end{aligned}$$

where

$$\begin{aligned}
(I)_L &= N_2^{\zeta - 2\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad - N_2^{\zeta - 2\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds,
\end{aligned}$$

(II)<sub>L</sub>

$$\begin{aligned}
&= \frac{N_2^{\zeta - 2\varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&\quad - \frac{N_2^{\zeta - 2\varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds,
\end{aligned}$$

$$\begin{aligned}
(III)_L &= \\
&= \frac{N_2^{\zeta-2\varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \left\langle c\sigma'(Z^{2,N_1}(x'))\sigma'(w^1 x')w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta)x', \eta_s^{N_1, N_2} \right\rangle \\
&\quad \pi(dx', dy) ds \\
&+ \frac{N_2^{\zeta-2\varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \left\langle c\sigma'(Z^{2,N_1}(x'))\sigma'(w^1 x')w^2, \eta_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta)x', \gamma_0^{N_1} \right\rangle \\
&\quad \pi(dx', dy) ds \\
&- \frac{N_2^{\zeta-2\varphi}}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle \left\langle c\sigma'(Z^{2,N_1}(x'))\sigma'(w^1 x')w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta)x', \gamma_0^{N_1} \right\rangle \\
&\quad \pi(dx', dy) ds,
\end{aligned}$$

and

$$\begin{aligned}
\Gamma_{L,t}^{N_1, N_2} &= \\
&= -\frac{1}{N_2^{2\varphi-\zeta}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle \partial_c f(\theta)\sigma(Z^{2,N_1}(x')), (\gamma_s^{N_1, N_2} - \gamma_0^{N_1}) \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1} N_2^{2\varphi-\zeta}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \\
&\quad \left\langle c\sigma'(Z^{2,N_1}(x'))\sigma(w^1 x') \cdot \partial_{w^2} f(\theta), (\gamma_s^{N_1, N_2} - \gamma_0^{N_1}) \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1} N_2^{2\varphi-\zeta}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle \left\langle c\sigma'(Z^{2,N_1}(x'))\sigma'(w^1 x')w^2, \gamma_0^{N_1} \right\rangle \cdot \right. \\
&\quad \left. \cdot \nabla_{w^1} f(\theta)x', (\gamma_s^{N_1, N_2} - \gamma_0^{N_1}) \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1} N_2^{2\varphi-\zeta}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle \left\langle c\sigma'(Z^{2,N_1}(x'))\sigma'(w^1 x')w^2, (\gamma_s^{N_1, N_2} - \gamma_0^{N_1}) \right\rangle \cdot \right. \\
&\quad \left. \cdot \nabla_{w^1} f(\theta)x', \gamma_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&+ \frac{1}{N_1^{1-\gamma_1} N_2^{2\varphi-\zeta}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle \left\langle c\sigma'(Z^{2,N_1}(x'))\sigma'(w^1 x')w^2, (\gamma_s^{N_1, N_2} - \gamma_0^{N_1}) \right\rangle \cdot \right. \\
&\quad \left. \cdot \nabla_{w^1} f(\theta)x', \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds
\end{aligned}$$

The following lemmas show compact containment and regularity of  $L_t^{N_1, N_2}(f)$  for fixed  $f \in C_b^3(\mathbb{R}^{1+N_1(1+d)})$ .

**Lemma D.1.** *When  $\zeta \leq 2 - 2\gamma_2$ , for any fixed  $f \in C_b^3(\mathbb{R}^{1+N_1(1+d)})$ , there exists a constant  $C < \infty$ , such that*

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{E} \left[ \left| L_t^{N_1, N_2}(f) \right|^2 \right] < C.$$

Thus, for any  $\epsilon > 0$ , there exist a compact interval  $U \subset \mathbb{R}$ , such that

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{P} \left( L_t^{N_1, N_2}(f) \notin U \right) < \epsilon.$$

*Proof.* By equation (55) and the Cauchy-Schartz inequality, we have

$$\begin{aligned}
&\left| L_t^{N_1, N_2}(f) \right|^2 \\
&\leq \frac{C}{N^{2(2\varphi-\zeta)}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| y - h_s^{N_1}(x') \right|^2 \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
& \times \left( \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| \left\langle \partial_c f(\theta) \sigma(Z^{2,N_1}(x')), \eta_s^{N_1, N_2} \right\rangle \right|^2 \pi(dx', dy) ds + \right. \\
& \quad \left. + \frac{1}{N_1^{2(1-\gamma_1)}} \left| \left\langle c\sigma'(Z^{2,N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \eta_s^{N_1, N_2} \right\rangle \right|^2 \pi(dx', dy) ds \right) \\
& + \frac{C}{N^{2(2\varphi-\zeta)}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1, N_2}(x') \right|^2 \pi(dx', dy) ds \\
& \times \left( \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| \left\langle \partial_c f(\theta) \sigma(Z^{2,N_1}(x')), \gamma_0^{N_1} \right\rangle \right|^2 \pi(dx', dy) ds \right. \\
& \quad \left. + \frac{1}{N_1^{2(1-\gamma_1)}} \left| \left\langle c\sigma'(Z^{2,N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_0^{N_1} \right\rangle \right|^2 \pi(dx', dy) ds \right) \\
& + \frac{C}{N_1^{2(1-\gamma_1)} N_2^{2(2\varphi-\zeta)}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| y - h_s^{N_1}(x') \right|^2 \pi(dx', dy) ds \\
& \times \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| \left\langle \left\langle c\sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_s^{N_1, N_2} \right\rangle \right|^2 \pi(dx', dy) ds \right. \\
& \quad \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| \left\langle \left\langle c\sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1} \right\rangle \right|^2 \pi(dx', dy) ds \right\} \\
& + \frac{C}{N_1^{2(1-\gamma_1)} N_2^{2(2\varphi-\zeta)}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| K_s^{N_1, N_2}(x') \right|^2 \pi(dx', dy) ds \\
& \times \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| \left\langle \left\langle c\sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1} \right\rangle \right|^2 \pi(dx', dy) ds \\
& + C \left\{ \left| \Gamma_{L,t}^{N_1, N_2} \right|^2 + \left| N_2^\zeta \left\langle f, \gamma_0^{N_1, N_2} - \gamma_0^{N_1} \right\rangle \right|^2 + \left| N_2^{\zeta-\varphi} M_{\eta,1,t}^{N_1, N_2} \right|^2 \right. \\
& \quad \left. + \left| N_2^{\zeta-\varphi} M_{\eta,2,t}^{N_1, N_2} \right|^2 + \left| N_2^{\zeta-\varphi} M_{\eta,3,t}^{N_1, N_2} \right|^2 \right\} \\
& + O(N_2^{-2+2\zeta})
\end{aligned}$$

When  $\zeta \leq 2\varphi = 2 - 2\gamma_2$ ,  $0 \leq t \leq T$ , the expectation of the first five terms and  $|\Gamma_{L,t}^{N_1, N_2}|^2$  are bounded by Assumption 2.1, Lemmas A.1, C.1 and C.6. Since  $\gamma_2 > \frac{3}{4}$ ,  $\zeta < \frac{1}{2}$  and  $\zeta + \gamma_2 - 1 \leq 1 - \gamma_2 < \frac{1}{2}$ , by similar analysis as in Section C, the remainder terms all converges to 0 as  $N_2 \rightarrow \infty$ . The result of the lemma follows.  $\square$

**Lemma D.2.** *When  $\zeta \leq 2 - 2\gamma_2$ , for any  $f \in C_b^3(\mathbb{R}^{1+N_1(1+d)})$ ,  $\delta \in (0, 1)$ , there is a constant  $C < \infty$  such that for  $0 \leq u \leq \delta$ ,  $0 \leq v \leq \delta \wedge t$ , and  $t \in [0, T]$ ,*

$$\begin{aligned}
& \mathbb{E} \left[ q \left( L_{t+u}^{N_1, N_2}(f), L_t^{N_1, N_2}(f) \right) q \left( L_t^{N_1, N_2}(f), L_{t-v}^{N_1, N_2}(f) \right) | \mathcal{F}_t^{N_1, N_2} \right] \\
& \leq \frac{C}{N_2^{2-2\gamma_2-\zeta}} \delta + \frac{C}{N_2^{2-\zeta-\gamma_2}}.
\end{aligned}$$

*Proof.* The proof is identical to that of Lemma C.2 of [45] and thus it is omitted from here.  $\square$

Denote  $\mathfrak{K}_t^{N_1, N_2} = (\gamma_t^{N_1, N_2}, h_t^{N_1, N_2}, l_{1,t}^{N_1, N_2}, l_{2,t}^{N_1, N_2}, l_{3,t}^{N_1, N_2}, K_t^{N_1, N_2})$ . In the next lemma, we prove the convergence of the processes  $(\mathfrak{K}_t^{N_1, N_2}, L_t^{N_1, N_2}(f))$  in distribution in the space  $D_{E_2}([0, T])$ , where we have set  $E_2 = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^{N_1} \times \mathbb{R}^{N_1} \times \mathbb{R}^M \times \mathbb{R}$ .

**Lemma D.3.** *When  $\gamma_2 \in (\frac{3}{4}, 1)$ ,  $\varphi = 1 - \gamma_2$  and  $\zeta \leq 2\varphi$ , for any fixed  $f \in C_b^3(\mathbb{R}^{1+N_1(1+d)})$ , the processes  $(\mathfrak{K}_t^{N_1, N_2}, L_t^{N_1, N_2}(f))$  converge in distribution in the*

space  $D_{E_2}([0, T])$  to  $(\mathfrak{K}_t^{N_1}, L_t^{N_1}(f))$ , where  $\mathfrak{K}_t^{N_1} = (\gamma_t^{N_1}, h_t^{N_1}, l_{1,t}^{N_1}, l_{2,t}^{N_1}, l_{3,t}^{N_1}, K_t^{N_1})$  satisfying equations (36), (42), and (17). When  $\zeta < 2\varphi$ ,  $L_t^{N_1}(f) = 0$ . When  $\zeta = 2\varphi$ ,  $L_t^{N_1}(f)$  satisfies (18).

*Proof.* Recall that  $\{\mathfrak{K}_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{E_1}([0, T])$ , where  $E_1 = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^{N_1} \times \mathbb{R}^{N_1} \times \mathbb{R}^M$ . By Lemmas D.1 and D.2,  $\{L_t^{N_1, N_2}(f)\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{\mathbb{R}}([0, T])$ . These implies that the probability measures of the family of processes  $\{\mathfrak{K}_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  and the probability measures of the family of processes  $\{L_t^{N_1, N_2}(f)\}_{N_2 \in \mathbb{N}}$  are tight. Therefore,  $\{\mathfrak{K}_t^{N_1, N_2}, L_t^{N_1, N_2}(f)\}_{N_2 \in \mathbb{N}}$  is tight. Hence,  $\{\mathfrak{K}_t^{N_1, N_2}, L_t^{N_1, N_2}(f)\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{E_2}([0, T])$ .

Denote  $\pi^{N_1, N_2} \in \mathcal{M}(D_{E_2}([0, T]))$  the probability measure corresponding to  $(\mathfrak{K}_t^{N_1, N_2}, L_t^{N_1, N_2}(f))$ . Relative compactness implies that there is a subsequence  $\pi^{N_1, N_{2k}}$  that converges weakly. One can show that any limit point  $\pi^{N_1}$  of a convergence subsequence  $\pi^{N_1, N_{2k}}$  is a Dirac measure concentrated on  $(\mathfrak{K}_t^{N_1}, L_t^{N_1}(f)) \in D_{E_2}([0, T])$ .

*Case 1.* When  $\zeta < 2\varphi$ , for any  $t \in [0, T]$ ,  $b_1, \dots, b_p \in C_b(\mathbb{R})$ , and  $0 \leq s_1 < \dots < s_p \leq t$ , we define  $F_4(\mathfrak{K}, L(f)) : D_{E_2}([0, T]) \rightarrow \mathbb{R}_+$  as

$$\begin{aligned} F_4(\mathfrak{K}, L(f)) \\ = F_3(\gamma^{N_1}, h^{N_1}, l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1}) + \left| \left( L_t^{N_1}(f) - 0 \right) \times b_1(L_{s_1}^{N_1}(f)) \times \dots \times b_p(L_{s_p}^{N_1}(f)) \right|, \end{aligned} \quad (55)$$

where  $F_3$  is as given in equation (53). By equation (55), Lemma C.3, and similar analysis as in Lemma D.1, we have

$$\begin{aligned} & \mathbb{E}_{\pi^{N_1, N_2}} [F_4(\mathfrak{K}, L(f))] \\ &= E_{\pi^{N_1, N_2}} [F_3(\mu, h, l_B, K)] + \mathbb{E} \left[ \left| \left( L_t^{N_1, N_2}(f) - 0 \right) \times \prod_{i=1}^p b_i(L_{s_i}^{N_1, N_2}(f)) \right| \right] \\ &\leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\frac{1}{2}-\varphi}} + \frac{1}{N_2^{1-\varphi}} + \frac{1}{N_2^{\gamma_2-\varphi}} \right) + \frac{C}{N_2^{2\varphi-\zeta}} + \frac{C}{N_1^{1-\gamma_2} N_2^{2\varphi-\zeta}} + \frac{C}{N_2^{\frac{1}{2}-\zeta}} \\ &\quad + \mathbb{E} \left[ \left| N_2^{\zeta+\gamma_2-1} M_{\eta, 1, t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \left| N_2^{\zeta+\gamma_2-1} M_{\eta, 2, t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} + \frac{C}{N_2^{1-\zeta}} \\ &\leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{2\varphi-\zeta}} + \frac{1}{N_2^{\frac{3}{2}-\zeta-\gamma_2}} \right). \end{aligned}$$

Therefore,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_4(\mathfrak{K}, L(f))] = 0$ . Since  $F_4(\cdot)$  is continuous and uniformly bounded,

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_4(\mathfrak{K}, L(f))] = \mathbb{E}_{\pi^{N_1}} [F_4(\mathfrak{K}, L(f))] = 0.$$

Since relative compactness implies that every subsequence  $\pi^{N_1, N_{2k}}$  has a further sub-subsequence that converges weakly. And we have show that any limit point  $\pi^{N_1}$  of a convergence sequence must be a Dirac measure concentrated  $(\mathfrak{K}_t^{N_1}, L_t^{N_1}(f)) \in D_{E_2}([0, T])$ , where  $L_t(f) = 0$ . Since the solutions to equations (36) and (17) are unique, by Prokhorov's theorem, the processes  $(\mathfrak{K}_t^{N_1, N_2}, L_t^{N_1, N_2}(f))$  converges in distribution to  $(\mathfrak{K}_t^{N_1}, 0)$ .

*Case 2.* When  $\zeta = 2\varphi$ , for any  $t \in [0, T]$ ,  $b_1, \dots, b_p \in C_b(\mathbb{R})$ , and  $0 \leq s_1 < \dots < s_p \leq t$ , we define  $F_4(\mathfrak{K}, L(f)) : D_{E_2}([0, T]) \rightarrow \mathbb{R}_+$  as

$$\begin{aligned}
& F_5(\mathfrak{K}, L(f)) \\
&= F_3(\gamma^{N_1}, h^{N_1}, l_1^{N_1}, l_2^{N_1}, l_3^{N_1}, K^{N_1}) \\
&+ \left| \left( L_t^{N_1}(f) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) l_s^{N_1} (\partial_c f(\theta) \sigma(Z^{2, N_1}(x'))) \pi(dx', dy) ds \right. \right. \\
&+ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) l_s^{N_1} (c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta)) \pi(dx', dy) ds \\
&+ \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) l_s^{N_1} \left( \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x' \right) \\
&\quad \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle l_s^{N_1} (c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2) \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1} \right\rangle \\
&\quad \pi(dx', dy) ds \\
&+ \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left\langle \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1} \right\rangle \\
&\quad \pi(dx', dy) ds \\
&\times b_1(L_{s_1}(f)) \times \dots \times b_p(L_{s_p}(f)) \Big|,
\end{aligned}$$

where  $F_3$  is as given in equation (53). We first note that by equation (55)

$$\begin{aligned}
& L_t^{N_1, N_2}(f) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) l_s^{N_1, N_2} (\partial_c f(\theta) \sigma(Z^{2, N_1}(x'))) \pi(dx', dy) ds \\
&+ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle \partial_c f(\theta) \sigma(Z^{2, N_1}(x')), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) l_s^{N_1, N_2} (c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta)) \pi(dx', dy) ds \\
&+ \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) l_s^{N_1, N_2} \\
&\quad \left( \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x' \right) \pi(dx', dy) ds \\
&- \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1, N_2}(x')) \\
&\quad \left\langle l_s^{N_1, N_2} (c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2) \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&+ \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \\
&\quad \left\langle \left\langle c\sigma'(Z^{2, N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&= \frac{1}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x')
\end{aligned}$$

$$\begin{aligned}
& \left\langle \partial_c f(\theta) \sigma(Z^{2,N_1}(x')) + \frac{1}{N_1^{1-\gamma_1}} c \sigma'(Z^{2,N_1}(x')) \sigma(w^1 x') \cdot \partial_{w^2} f(\theta), \eta_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_2^{2\varphi} N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_0^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& - \frac{1}{N_2^{2\varphi} N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_2^{2\varphi} N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_2^\varphi N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \gamma_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_2^\varphi N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_0^{N_1} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& - \frac{1}{N_2^\varphi N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_0^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_2^\varphi N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& - \frac{1}{N_2^\varphi N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \\
& \quad \left\langle \left\langle c \sigma'(Z^{2,N_1}(x')) \sigma'(w^1 x') w^2, \eta_s^{N_1, N_2} \right\rangle \cdot \nabla_{w^1} f(\theta) x', \eta_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + N_2^{\zeta - \frac{1}{2}} \left\langle f, \sqrt{N_2} (\gamma_0^{N_1, N_2} - \gamma_0^{N_1}) \right\rangle + N_2^{\zeta + \gamma_2 - 1} M_{\eta, 1, t}^{N_1, N_2} + N_2^{\zeta + \gamma_2 - 1} M_{\eta, 2, t}^{N_1, N_2} + O(N_2^{-1+\zeta})
\end{aligned}$$

By similar analysis as for equations (50) to (52), the expectation of the absolute value of the first nine terms above are bounded by  $O(N_2^{-\varphi})$ . Then by Lemma C.3, we have

$$\begin{aligned}
& \mathbb{E}_{\pi^{N_1, N_2}} [F_5(\mathfrak{K}, L(f))] \\
& \leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\frac{1}{2}-\varphi}} + \frac{1}{N_2^{1-\varphi}} + \frac{1}{N_2^{\gamma_2-\varphi}} \right) + \frac{C}{N_2^\varphi} + \frac{C}{N_2^{\frac{1}{2}-\zeta}} \\
& \quad + \mathbb{E} \left[ \left| N_2^{\zeta+\gamma_2-1} M_{\eta, 1, t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} + \mathbb{E} \left[ \left| N_2^{\zeta+\gamma_2-1} M_{\eta, 2, t}^{N_1, N_2} \right|^2 \right]^{\frac{1}{2}} + \frac{C}{N_2^{1-\zeta}} \\
& \leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\frac{1}{2}-\varphi}} + \frac{1}{N_2^{1-\varphi}} + \frac{1}{N_2^{\frac{1}{2}-\zeta}} + \frac{1}{N_2^{\frac{3}{2}-\zeta-\gamma_2}} + \frac{1}{N_2^{1-\zeta}} \right) \\
& \leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\frac{1}{2}-\zeta}} \right).
\end{aligned}$$

Therefore,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_5(\mathfrak{K}, L(f))] = 0$ . Since  $F_5(\cdot)$  is continuous and uniformly bounded,

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_5(\mathfrak{K}, L(f))] = \mathbb{E}_{\pi^{N_1}} [F_5(\mathfrak{K}, L(f))] = 0.$$

The result then follows.  $\square$

Moving back to the analysis of  $\Psi_t^{N_1, N_2}$ , we first show compact containment of  $\Psi_t^N$  in the next lemma.

**Lemma D.4.** *When  $\zeta \leq \min\{\gamma_2 - \frac{1}{2}, 2 - 2\gamma_2\}$ , there exist a constant  $C < \infty$ , such that*

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{E} \left[ \left| \Psi_t^{N_1, N_2}(x) \right|^2 \right] < C.$$

Thus, for any  $\epsilon > 0$ , there exist a compact subset  $U \subset \mathbb{R}^M$ , such that

$$\sup_{N_2 \in \mathbb{N}, 0 \leq t \leq T} \mathbb{P} \left( \Psi_t^{N_1, N_2} \notin U \right) < \epsilon.$$

*Proof.* In the proof below,  $C < \infty$  represents some positive constant, which may be different from line to line. We first rewrite the term  $N_2^{\zeta - \varphi} \Gamma_t^{N_1, N_2}(x) = N_2^{\zeta - \varphi} \Gamma_{1,t}^{N_1, N_2}(x) + N_2^{\zeta - \varphi} \Gamma_{2,t}^{N_1, N_2}(x) + N_2^{\zeta - \varphi} \Gamma_{3,t}^{N_1, N_2}(x)$  as

$$\begin{aligned} & N_2^{\zeta - \varphi} \Gamma_{1,t}^{N_1, N_2}(x) \\ &= -\frac{1}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \langle B_{x,x'}^1(\theta), \eta_s^{N_1, N_2} \rangle \pi(dx', dy) ds \\ &\quad - \frac{1}{N_2^{2\varphi - \zeta}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \langle B_{x,x'}^1(\theta), \eta_s^{N_1, N_2} \rangle \pi(dx', dy) ds, \end{aligned}$$

$$\begin{aligned} & N_2^{\zeta - \varphi} \Gamma_{2,t}^{N_1, N_2}(x) \\ &= -\frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \langle B_{x,x'}^{2,j}(\theta), \eta_s^{N_1, N_2} \rangle \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1 N_2^{2\varphi - \zeta}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \langle B_{x,x'}^{2,j}(\theta), \eta_s^{N_1, N_2} \rangle \pi(dx', dy) ds, \end{aligned}$$

$$\begin{aligned} & N_2^{\zeta - \varphi} \Gamma_{3,t}^{N_1, N_2}(x) \\ &= -\frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \rangle \langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} \rangle \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1 N_2^{2\varphi - \zeta}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' \langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \rangle \langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} \rangle \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \langle B_x^{3,j}(\theta), \gamma_0^{N_1} \rangle \langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \rangle \pi(dx', dy) ds \end{aligned}$$

$$\begin{aligned}
& - \frac{1}{N_1 N_2^{2\varphi-\zeta}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) x x' \left\langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds.
\end{aligned}$$

Since the above terms involves the term  $K_t^{N_1}(x)$ , we first look at the bound for  $K_t^{N_1}(x)$ . By the Cauchy-Schwarz inequality, equations (36), (42), (17) and the analysis in Lemma C.6, for any  $t \in [0, T]$ , we have

$$\begin{aligned}
|K_t^{N_1}(x)|^2 & \leq C t^2 + C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} |K_s^{N_1}(x')|^2 \pi(dx', dy) ds \\
& \leq C T^2 + \frac{C T}{M} \int_0^t \sum_{x' \in \mathcal{X}} |K_s^{N_1}(x')|^2 \pi(dx', dy) ds.
\end{aligned}$$

Summing over  $x \in \mathcal{X}$  on both sides gives

$$\sum_{x \in \mathcal{X}} |K_t^{N_1}(x)|^2 \leq C T^2 M + C T \int_0^t \sum_{x' \in \mathcal{X}} |K_s^{N_1}(x')|^2 \pi(dx', dy) ds.$$

By applying Grönwall's inequality, we have

$$\sup_{0 \leq t \leq T} \sum_{x \in \mathcal{X}} |K_t^{N_1}(x)|^2 \leq \sup_{0 \leq t \leq T} C T^2 M \exp(CTt) < C,$$

which implies that  $\sup_{0 \leq t \leq T} |K_t^{N_1}(x)|^2 < C$  for any  $x \in \mathcal{X}$ . Using this uniform bound for  $K_t^{N_1}(x)$  together with Lemma C.1, similar analysis as for equation (48) gives

$$\begin{aligned}
& \mathbb{E} \left[ \left| N_2^{\zeta-\varphi} \Gamma_t^{N_1, N_2}(x) \right|^2 \right] \\
& \leq C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ |\Psi_s^{N_1, N_2}(x')|^2 \right] \pi(dx', dy) ds + C t^2 + \frac{C t^2}{N_2^{2(2\varphi-\zeta)}}
\end{aligned}$$

By equations (44), (45), (54) and Lemmas C.5 and D.1, we have

$$\begin{aligned}
& \mathbb{E} \left[ \left| \Psi_t^{N_1, N_2}(x) \right|^2 \right] \\
& \leq C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| L_s^{N_1, N_1}(B_{x, x'}^1(\theta)) \right|^2 + \frac{1}{N_1} \sum_{j=1}^{N_1} \left| L_s^{N_1, N_1}(B_{x, x'}^{2,j}(\theta)) \right|^2 \right] \pi(dx', dy) ds \\
& + C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| L_s^{N_1, N_1}(B_x^{3,j}(\theta)) \right|^2 + \left| L_s^{N_1, N_1}(B_{x'}^{3,j}(\theta)) \right|^2 + \left| \Psi_s^{N_1, N_2}(x') \right|^2 \right] \\
& \quad \pi(dx', dy) ds \\
& + \mathbb{E} \left[ \left| N_2^{\zeta-\varphi} \Gamma_t^{N_1, N_2}(x) \right|^2 \right] + C \mathbb{E} \left[ \left| \Psi_0^{N_1, N_2}(x) \right|^2 \right] + C \mathbb{E} \left[ \left| N_2^\zeta M_t^{N_1, N_2}(x) \right|^2 \right] \\
& \leq C t^2 + C t \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| \Psi_s^{N_1, N_2}(x') \right|^2 \right] \pi(dx', dy) ds + \frac{C t^2}{N_2^{2(2\varphi-\zeta)}} + C + \frac{C}{N_2^{1-2\zeta}}
\end{aligned}$$

Summing over  $x \in \mathcal{X}$  on both sides gives

$$\sum_{x \in \mathcal{X}} \mathbb{E} \left[ |\Psi_s^{N_1, N_2}(x)|^2 \right] \leq CMT^2 + CT \int_0^t \sum_{x' \in \mathcal{X}} \mathbb{E} \left[ |\Psi_s^{N_1, N_2}(x')|^2 \right] \pi(dx', dy) ds.$$

By Grönwall's inequality, we get

$$\sup_{0 \leq t \leq T} \sum_{x \in \mathcal{X}} \mathbb{E} \left[ |\Psi_t^{N_1, N_2}(x)|^2 \right] \leq \sup_{0 \leq t \leq T} CT^2 M \exp(CTt) < C,$$

which implies that  $\sup_{0 \leq t \leq T} \mathbb{E} \left[ |\Psi_t^{N_1, N_2}(x)|^2 \right] < C$  for any  $x \in \mathcal{X}$ . The result of the lemma then follows.  $\square$

The next lemma establishes the regularity of the process  $\Psi_t^{N_1, N_2}$  in  $D_{\mathbb{R}^M}([0, T])$ . For the purpose of this lemma, we denote  $q(z_1, z_2) = \min\{\|z_1 - z_2\|_{l_1}, 1\}$  for  $z_1, z_2 \in \mathbb{R}^M$ . The proof of the lemma is similar to that for Lemma C.7, which we omit here.

**Lemma D.5.** *For any  $\delta \in (0, 1)$ , there is a constant  $C < \infty$  such that for  $0 \leq u \leq \delta$ ,  $0 \leq v \leq \delta \wedge t$ , and  $t \in [0, T]$ ,*

$$\mathbb{E} \left[ q \left( \Psi_{t+u}^{N_1, N_2}, \Psi_t^N \right) q \left( \Psi_t^{N_1, N_2}, \Psi_{t-v}^{N_1, N_2} \right) | \mathcal{F}_t^{N_1, N_2} \right] \leq C\delta + \frac{C}{N_2^{1-\zeta}}.$$

Combining these with our analysis of  $L_t^{N_1, N_2}(f)$ , we can now identify the limit for  $\Psi_t^{N_1, N_2}$ . We denote  $\mathfrak{L}_t^{N_1, N_2} = (\mathfrak{K}_t^{N_1, N_2}, L_{1,t}^{N_1, N_2}, L_{2,t}^{N_1, N_2}, L_{3,t}^{N_1, N_2})$ , where  $L_{1,t}^{N_1, N_2} = L_t^{N_1, N_2}(B_{x,x'}^1(\theta))$ ,  $L_{2,t}^{N_1, N_2}$  and  $L_{3,t}^{N_1, N_2}$  are  $N_1$ -dimensional vectors with  $j$ -th entry being  $L_t^{N_1, N_2}(B_{x,x'}^{2,j}(\theta))$  and  $L_t^{N_1, N_2}(B_x^{3,j}(\theta))$ , respectively. In the next lemma, we prove the convergence of the processes  $(\mathfrak{L}_t^{N_1, N_2}, \Psi_t^{N_1, N_2})$  in distribution in the space  $D_{E_3}([0, T])$ , where  $E_3 = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^{N_1} \times \mathbb{R}^{N_1} \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^{N_1} \times \mathbb{R}^{N_1} \times \mathbb{R}^M$ .

**Lemma D.6.** *When  $\gamma_2 \in (\frac{3}{4}, 1)$ ,  $\varphi = 1 - \gamma_2$  and  $\zeta \leq \gamma_2 - \frac{1}{2}$ , the processes  $(\mathfrak{L}_t^{N_1, N_2}, \Psi_t^{N_1, N_2})$  in distribution in the space  $D_{E_3}([0, T])$  to  $(\mathfrak{L}_t^{N_1}, \Psi_t^{N_1})$ . In particular,  $\mathfrak{L}_t^{N_1} = (\gamma_0^{N_1}, h_t^{N_1}, l_{1,t}^{N_1}, l_{2,t}^{N_1}, l_{3,t}^{N_1}, K_t^{N_1}, L_{1,t}^{N_1}, L_{2,t}^{N_1}, L_{3,t}^{N_1})$  satisfies equations (36), (42), and (17),  $L_{1,t}^{N_1}, L_{2,t}^{N_1}, L_{3,t}^{N_1}, \Psi_t^{N_1}$  satisfy either of the following case:*

*Case 1. When  $\gamma_2 \in (\frac{3}{4}, \frac{5}{6})$  and  $\zeta \leq \gamma_2 - \frac{1}{2}$ , or when  $\gamma_2 \in [\frac{5}{6}, 1)$  and  $\zeta < 2 - 2\gamma_2 \leq \gamma_2 - \frac{1}{2}$ , one has  $L_{1,t}^{N_1} = 0$ ,  $L_{2,t}^{N_1} = L_{3,t}^{N_1} = 0$  and  $\Psi_t^{N_1}$  satisfies (19).*

*Case 2. When  $\gamma_2 \in [\frac{5}{6}, 1)$  and  $\zeta = 2 - 2\gamma_2 \leq \gamma_2 - \frac{1}{2}$ ,  $L_{1,t}^{N_1}, L_{2,t}^{N_1}, L_{3,t}^{N_1}$  satisfy equation (18) and  $\Psi_t$  satisfies (20).*

*Proof.* By analysis in Lemma D.3,  $\{\mathfrak{L}_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{E_4}([0, T])$ , where we have set  $E_4 = \mathcal{M}(\mathbb{R}^{1+N_1(1+d)}) \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^{N_1} \times \mathbb{R}^{N_1} \times \mathbb{R}^M \times \mathbb{R} \times \mathbb{R}^{N_1} \times \mathbb{R}^{N_1}$ . By Lemmas D.4 and D.5,  $\{\Psi_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{\mathbb{R}^M}([0, T])$ . These implies that the probability measures of the family of processes  $\{\mathfrak{L}_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  and the probability measures of the family of processes  $\{\Psi_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  are tight. Therefore,  $\{\mathfrak{L}_t^{N_1, N_2}, \Psi_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is tight. Hence,  $\{\mathfrak{L}_t^{N_1, N_2}, \Psi_t^{N_1, N_2}\}_{N_2 \in \mathbb{N}}$  is relatively compact in  $D_{E_3}([0, T])$ .

Denote  $\pi^{N_1, N_2} \in \mathcal{M}(D_{E_3}([0, T]))$  the probability measure corresponding to  $(\mathfrak{L}_t^{N_1, N_2}, \Psi_t^{N_1, N_2})$ . Relative compactness implies that there is a subsequence  $\pi^{N_1, N_{2k}}$  that converges weakly. We now show that any limit point  $\pi^{N_1}$  of a convergence subsequence  $\pi^{N_1, N_{2k}}$  is a Dirac measure concentrated on  $(\mathfrak{L}_t^{N_1}, \Psi_t^{N_1}) \in D_{E_3}([0, T])$ .

*Case 1.* When  $\gamma_2 \in (\frac{3}{4}, \frac{5}{6})$  and  $\zeta \leq \gamma_2 - \frac{1}{2} < 2\varphi$ , or when  $\gamma_2 \in [\frac{5}{6}, 1)$  and  $\zeta < 2\varphi \leq \gamma_2 - \frac{1}{2}$ , for any  $t \in [0, T]$ ,  $b_1^{i,j}, \dots, b_p^{i,j} \in C_b(\mathbb{R})$ ,  $d_1, \dots, d_p \in C_b(\mathbb{R}^M)$ , and  $0 \leq s_1 < \dots < s_p \leq t$ , we define  $F_6(\mathfrak{L}, \Psi) : D_{E_3}([0, T]) \rightarrow \mathbb{R}_+$  as

$$\begin{aligned}
& F_6(\mathfrak{L}, \Psi) \\
&= F_4(\mathfrak{K}^{N_1}, L^{N_1}(B_{x,x'}^1(\theta))) + \sum_{i=2}^3 \sum_{j=1}^{N_1} \left| \left( L_{i,t}^{N_1,j} - 0 \right) \times b_1^{i,j}(L_{i,s_1}^{N_1,j}) \times \dots \times b_p^{i,j}(L_{i,s_p}^{N_1,j}) \right| \\
&+ \sum_{x \in \mathcal{X}} \left| \left( \Psi_t^{N_1}(x) - \Psi_0^{N_1}(x) + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1} \right\rangle \right. \right. \\
&\quad \left. \left. \pi(dx', dy) ds \right) \right. \\
&+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&- \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) \left[ L_s^{N_1}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1}((B_{x,x'}^{2,j}(\theta))) \right] \pi(dx', dy) ds \\
&- \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) L_s^{N_1}(B_x^{3,j}(\theta)) \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) \left\langle x x' B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle L_s^{N_1}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
&\quad \times \dots \times d_p(\Psi_{s_p}) \left. \right|,
\end{aligned}$$

where  $F_4(\mathfrak{K}, L(B_{x,x'}(c, w)))$  is as given in equation (55) and  $L_{i,t}^{N_1,j}$  is the  $j$ -th element of the  $N_1$ -dimensional vector  $L_{i,t}^{N_1}$  for  $i = 2, 3$ . Note that by equation (54), we have

$$\begin{aligned}
& \Psi_t^{N_1, N_2}(x) - \Psi_0^{N_1, N_2}(x) \\
&+ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&- \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left[ L_s^{N_1, N_2}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1, N_2}((B_{x,x'}^{2,j}(\theta))) \right] \pi(dx', dy) ds \\
&- \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) L_s^{N_1, N_2}(B_x^{3,j}(\theta)) \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left\langle x x' B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle L_s^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
&= (1)_\Psi + (2)_\Psi + (3)_\Psi + (4)_\Psi + (5)_\Psi + N_2^{\zeta - \varphi} \Gamma_t^{N_1, N_2}(x) + N_2^\zeta M_t^{N_1, N_2}(x) + O(N_2^{-\gamma_2 + \zeta}),
\end{aligned}$$

where

$$\begin{aligned}
(1)_\Psi &= \frac{1}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \eta_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds, \quad (56) \\
(2)_\Psi &= \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \eta_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&+ \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \eta_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds,
\end{aligned}$$

$$\begin{aligned}
(3)_\Psi &= \frac{1}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left[ L_s^{N_1, N_2}(B_{x, x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1, N_2}(B_{x, x'}^{2, j}(\theta)) \right] \pi(dx', dy) ds, \\
(4)_\Psi &= \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') L_s^{N_1, N_2}(B_x^{3, j}(\theta)) \left\langle x x' B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&\quad - \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) L_s^{N_1, N_2}(B_x^{3, j}(\theta)) \left\langle x x' B_{x'}^{3, j}(\theta), \eta_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds, \\
(5)_\Psi &= \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left\langle x x' B_x^{3, j}(\theta), \gamma_0^{N_1} \right\rangle L_s^{N_1, N_2}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds \\
&\quad - \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left\langle x x' B_x^{3, j}(\theta), \eta_0^{N_1, N_2} \right\rangle L_s^{N_1, N_2}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds.
\end{aligned} \tag{57}$$

We now analyze each of these five terms. By the Cauchy-Schwartz inequality, Lemmas C.1, C.6, D.1 and D.4, we have

$$\mathbb{E} [(1)_\Psi + (3)_\Psi] \leq \frac{C}{N_2^\varphi}. \tag{58}$$

For term  $(2)_\Psi$ , since

$$\begin{aligned}
&\frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \mathbb{E} \left[ \left| \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3, j}(\theta), \eta_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \right| \right] \\
&\leq \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \mathbb{E} \left[ \sup_{x'} \left| \left\langle x x' B_{x'}^{3, j}(\theta), \gamma_0^{N_1} \right\rangle \right| \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left| \Psi_s^{N_1, N_2}(x') \left\langle B_x^{3, j}(\theta), \eta_0^{N_1, N_2} \right\rangle \right| \pi(dx', dy) ds \right] \\
&\leq \frac{C}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| \Psi_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| \left\langle B_x^{3, j}(\theta), \eta_0^{N_1, N_2} \right\rangle \right|^2 \right]^{\frac{1}{2}} \pi(dx', dy) ds \\
&\leq \frac{C}{N_2^\varphi}
\end{aligned}$$

and similar bound can be obtained for the second term in  $(2)_\Psi$ , we have  $\mathbb{E} [(2)_\Psi] \leq C/N_2^\varphi$ . For term  $(4)_\Psi$ , we see that

$$\begin{aligned}
\mathbb{E} [(4)_\Psi] &\leq \frac{C}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| K_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| L_s^{N_1, N_2}(B_x^{3, j}(\theta)) \right|^2 \right]^{\frac{1}{2}} \pi(dx', dy) ds \\
&\quad + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| L_s^{N_1, N_2}(B_x^{3, j}(\theta)) \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| y - h_s^{N_1, N_2}(x') \right|^4 \right]^{\frac{1}{4}} \\
&\quad \mathbb{E} \left[ \left| \left\langle x x' B_{x'}^{3, j}(\theta), \eta_0^{N_1, N_2} \right\rangle \right|^4 \right]^{\frac{1}{4}} \pi(dx', dy) ds \\
&\leq \frac{C(T)}{N_2^\varphi}.
\end{aligned}$$

Similarly,  $\mathbb{E} [(5)_\Psi] \leq C(T)/N_2^\varphi$ . By the Cauchy-Schwartz inequality, Lemmas A.1, C.1, D.1, and D.4, we have

$$\begin{aligned}
&\mathbb{E} \left[ \left| N_2^{\zeta - \varphi} \Gamma_t^{N_1, N_2} \right| \right] \\
&\leq \frac{C}{N_2^\varphi} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| \Psi_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| l_s^{N_1, N_2}(B_{x, x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_s^{N_1, N_2}(B_{x, x'}^{2, j}(\theta)) \right|^2 \right]^{\frac{1}{2}} \\
&\quad \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
& + \frac{C}{N_2^{2\varphi-\zeta}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \mathbb{E} \left[ \left| l_s^N(B_{x,x'}(c, w)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_s^{N_1, N_2}(B_{x,x'}^{2,j}(\theta)) \right| \right] \pi(dx', dy) ds \\
& + \frac{C}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} x x' \mathbb{E} \left[ \left| \Psi_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| \left\langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right|^4 \right]^{\frac{1}{4}} \\
& \quad \mathbb{E} \left[ \left| \left\langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} \right\rangle \right|^2 \right]^{\frac{1}{4}} \pi(dx', dy) ds \\
& + \frac{C}{N_1 N_2^{2\varphi-\zeta}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} x x' \mathbb{E} \left[ \left| \left\langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| \left\langle B_{x'}^{3,j}(\theta), \gamma_s^{N_1, N_2} \right\rangle \right|^2 \right]^{\frac{1}{2}} \\
& \quad \pi(dx', dy) ds \\
& + \frac{C}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} x x' \mathbb{E} \left[ \left| \Psi_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right|^2 \right]^{\frac{1}{2}} \pi(dx', dy) ds \\
& + \frac{C}{N_1 N_2^{2\varphi-\zeta}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} x x' \mathbb{E} \left[ \left| \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right|^2 \right]^{\frac{1}{2}} \pi(dx', dy) ds \\
& + \frac{C}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \left( \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} x x' \mathbb{E} \left[ \left| \left\langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right|^4 \right]^{\frac{1}{2}} \right. \\
& \quad \left. \mathbb{E} \left[ \left| \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \right|^2 \right]^{\frac{1}{2}} \pi(dx', dy) ds \right)^{\frac{1}{2}} \\
& \leq C \left( \frac{1}{N_2^\varphi} + \frac{1}{N_2^{2\varphi-\zeta}} \right). \tag{59}
\end{aligned}$$

Putting everything together, by equation (54), Lemmas Lemmas C.1, C.6, D.1, D.3, D.4, and the analysis in Section C, we have

$$\begin{aligned}
& \mathbb{E}_{\pi^{N_1, N_2}} [F_6(\mathcal{L}, \Psi)] \\
& = \mathbb{E}_{\pi^{N_1, N_2}} \left[ F_4(\mathfrak{K}^{N_1}, L^{N_1}(B_{x,x'}^1(\theta))) \right] \\
& \quad + \sum_{i=2}^3 \sum_{j=1}^{N_1} \mathbb{E} \left[ \left| \left( L_{i,t}^{N_1, N_2, j} - 0 \right) \times b_1^{i,j}(L_{i,s_1}^{N_1, N_2, j}) \times \dots \times b_p^{i,j}(L_{i,s_p}^{N_1, N_2, j}) \right| \right] \\
& \quad + \sum_{x \in \mathcal{X}} \mathbb{E} \left[ \left| \left( \Psi_t^{N_1, N_2}(x) - \Psi_0^{N_1, N_2}(x) \right. \right. \right. \\
& \quad \left. \left. \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \right) \pi(dx', dy) ds \right. \right. \\
& \quad \left. \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \right. \right. \\
& \quad \left. \left. \pi(dx', dy) ds \right| \right. \\
& \quad - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left[ L_s^{N_1, N_2}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1, N_2}((B_{x,x'}^{2,j}(\theta))) \right] \\
& \quad \pi(dx', dy) ds \\
& \quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) L_s^{N_1, N_2}(B_x^{3,j}(\theta)) \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
& -\frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left\langle xx' B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle L_s^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& \quad \times \cdots \times d_p(\Psi_{s_p}^{N_1, N_2}) \Big] \\
& \leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{2\varphi-\zeta}} \right) + C \left( \frac{1}{N_1^\varphi} + \frac{1}{N_1^{2\varphi-\zeta}} \right) + C \mathbb{E} \left[ \left| N_2^\zeta M_t^{N_1, N_2} \right| \right] + O(N_2^{-\gamma_2+\zeta}) \\
& \leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{2\varphi-\zeta}} \right).
\end{aligned}$$

Therefore,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_6(\mathfrak{L}, \Psi)] = 0$ . Since  $F_6(\cdot)$  is continuous and uniformly bounded,

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_6(\mathfrak{L}, \Psi)] = \mathbb{E}_{\pi^{N_1}} [F_6(\mathfrak{L}, \Psi)] = 0.$$

Since relative compactness implies that every subsequence  $\pi^{N_1, N_{2k}}$  has a further sub-subsequence that converges weakly. And we have show that any limit point  $\pi^{N_1}$  of a convergence sequence must be a Dirac measure concentrated  $(\mathfrak{L}^{N_1}, \Psi^{N_1}) \in D_{E_3}([0, T])$ . In particular,  $K_t^{N_1}$  satisfies (17),  $L_{1,t}^{N_1} = 0$ ,  $L_{2,t}^{N_1} = L_{3,t}^{N_1} = 0$  and  $\Psi_t^{N_1}$  satisfies equation (19). Since the solutions to equations (36), (17) and (19) are unique, by Prokhorov's theorem, the processes  $(\mathfrak{L}_t^{N_1, N_2}, \Psi_t^{N_1, N_2})$  converges in distribution to  $(\mathfrak{L}_t^{N_1}, \Psi_t^{N_1})$ .

*Case 2.* When  $\gamma_2 \in [\frac{5}{6}, 1)$  and  $\zeta = 2 - 2\gamma_2 = 2\varphi$ , for any  $t \in [0, T]$ ,  $b_1^{i,j}, \dots, b_p^{i,j} \in C_b(\mathbb{R})$ ,  $d_1, \dots, d_p \in C_b(\mathbb{R}^M)$ , and  $0 \leq s_1 < \dots < s_p \leq t$ , we define  $F_7(\mathfrak{L}, \Psi) : D_{E_3}([0, T]) \rightarrow \mathbb{R}_+$  as

$$\begin{aligned}
F_7(\mathfrak{L}, \Psi) &= F_5(\mathfrak{K}, L(B_{x,x'}^1(\theta))) + \sum_{i=2}^3 \sum_{j=1}^{N_1} \left| F_L(L_{i,t}^{N_1,j}) \times b_1^{i,j}(L_{i,s_1}^{N_1,j}) \times \cdots \times b_p^{i,j}(L_{i,s_p}^{N_1,j}) \right| \\
&+ \sum_{x \in \mathcal{X}} \left| \left( \Psi_t^{N_1}(x) - \Psi_0^{N_1}(x) + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1} \right\rangle \right. \right. \\
&\quad \left. \left. \pi(dx', dy) ds \right) \right. \\
&+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1}(x') xx' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&- \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left[ L_s^{N_1}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1}((B_{x,x'}^{2,j}(\theta))) \right] \pi(dx', dy) ds \\
&- \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) L_s^{N_1}(B_x^{3,j}(\theta)) \left\langle xx' B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
&- \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - h_s^{N_1}(x')) \left\langle xx' B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle L_s^{N_1}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
&+ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') \left[ l_s^{N_1}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_s^{N_1}(B_{x,x'}^{2,j}(\theta)) \right] \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' l_s^{N_1} (B_x^{3,j}(\theta)) \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1} \right\rangle l_s^{N_1} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& \times d_1(\Psi_{s_1}^{N_1}) \times \cdots \times d_p(\Psi_{s_p}^{N_1}) \Big|,
\end{aligned}$$

where  $F_5(\mathfrak{K}, L(B_{x,x'}(c, w)))$  is as given in equation (56),  $L_{i,t}^{N_1,j}$  is the  $j$ -th element of the  $N_1$ -dimensional vector  $L_{i,t}^{N_1}$  for  $i = 2, 3$ , and  $F_L(f)$  is equal to  $L_t^{N_1}(f)$  minus the right-hand side of (18). Note that by equation (54),

$$\begin{aligned}
& \Psi_t^{N_1, N_2}(x) - \Psi_0^{N_1, N_2}(x) \\
& + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \left\langle B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left[ L_s^{N_1, N_2}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1, N_2}((B_{x,x'}^{2,j}(\theta))) \right] \pi(dx', dy) ds \\
& - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) L_s^{N_1, N_2}(B_x^{3,j}(\theta)) \left\langle x x' B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left\langle x x' B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle L_s^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left[ l_s^{N_1, N_2}(B_{x,x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_s^{N_1}(B_{x,x'}^{2,j}(\theta)) \right] \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_s^{N_1, N_2}(B_x^{3,j}(\theta)) \left\langle B_{x'}^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3,j}(\theta), \gamma_0^{N_1, N_2} \right\rangle l_s^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& = \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_s^{N_1, N_2}(B_x^{3,j}(\theta)) l_0^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_0^{N_1, N_2}(B_x^{3,j}(\theta)) l_s^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& - \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_s^{N_1, N_2}(B_x^{3,j}(\theta)) l_s^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \\
& + \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1}(x') \right) x x' \left\langle B_x^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3,j}(\theta), \eta_s^{N_1, N_2} \right\rangle \pi(dx', dy) ds \\
& + (1)_\Psi + (2)_\Psi + (3)_\Psi + (4)_\Psi + (5)_\Psi + N_2^\zeta M_t^{N_1, N_2}(x) + O(N_2^{-\gamma_2 + \zeta}),
\end{aligned} \tag{60}$$

where  $(1)_\Psi$  to  $(5)_\Psi$  are given in (56) to (57). By Lemmas C.1 and C.6,

$$\begin{aligned} & \mathbb{E} \left[ \left| \frac{1}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_s^{N_1, N_2}(B_x^{3,j}(\theta)) l_0^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right| \right] \\ & \leq \frac{C}{N_1 N_2^\varphi} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \\ & \quad \mathbb{E} \left[ \left| K_s^{N_1, N_2}(x') \right|^2 \right]^{\frac{1}{2}} \mathbb{E} \left[ \left| l_s^{N_1, N_2}(B_x^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{4}} \mathbb{E} \left[ \left| l_0^{N_1, N_2}(B_{x'}^{3,j}(\theta)) \right|^4 \right]^{\frac{1}{4}} \pi(dx', dy) ds. \\ & \leq \frac{C}{N_2^\varphi} \end{aligned}$$

Similarly, the expectation of the absolute value of the first three terms on the right-hand side of (60) are bounded by  $O(N_2^{-\varphi})$ . The analysis for the forth term and  $(1)_\Psi$  to  $(5)_\Psi$  are given in (58) to (59).

Therefore, we have

$$\begin{aligned} & \mathbb{E}_{\pi^{N_1, N_2}} [F_7(\mathfrak{L}, \Psi)] \\ & = \mathbb{E}_{\pi^{N_1, N_2}} \left[ F_5(\mathfrak{K}, L(B_{x, x'}^1(\theta))) \right] \\ & \quad + \sum_{i=2}^3 \sum_{j=1}^{N_1} \mathbb{E} \left[ \left| F_L(L_{i, t}^{N_1, N_2, j}) \times b_1^{i, j}(L_{i, s_1}^{N_1, N_2, j}) \times \cdots \times b_p^{i, j}(L_{i, s_p}^{N_1, N_2, j}) \right| \right] \\ & \quad + \sum_{x \in \mathcal{X}} \mathbb{E} \left[ \left| \left( \Psi_t^{N_1, N_2}(x) - \Psi_0^{N_1, N_2}(x) \right. \right. \right. \\ & \quad \left. \left. \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') \left\langle B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \right) \right. \right. \\ & \quad \left. \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \Psi_s^{N_1, N_2}(x') x x' \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \right. \right. \\ & \quad \left. \left. - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left[ L_s^{N_1, N_2}(B_{x, x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} L_s^{N_1, N_2}((B_{x, x'}^{2, j}(\theta))) \right] \pi(dx', dy) ds \right. \right. \\ & \quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) L_s^{N_1, N_2}(B_x^{3, j}(\theta)) \left\langle x x' B_{x'}^{3, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \right. \right. \\ & \quad \left. \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - h_s^{N_1, N_2}(x') \right) \left\langle x x' B_x^{3, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle L_s^{N_1, N_2}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds \right. \right. \\ & \quad \left. \left. + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') \left[ l_s^{N_1, N_2}(B_{x, x'}^1(\theta)) + \frac{1}{N_1} \sum_{j=1}^{N_1} l_s^{N_1, N_2}(B_{x, x'}^{2, j}(\theta)) \right] \pi(dx', dy) ds \right. \right. \\ & \quad \left. \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' l_s^{N_1, N_2}(B_x^{3, j}(\theta)) \left\langle B_{x'}^{3, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle \pi(dx', dy) ds \right. \right. \\ & \quad \left. \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} K_s^{N_1, N_2}(x') x x' \left\langle B_x^{3, j}(\theta), \gamma_0^{N_1, N_2} \right\rangle l_s^{N_1, N_2}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds \right. \right. \\ & \quad \left. \left. \times d_1(\Psi_{s_1}^{N_1, N_2}) \times \cdots \times d_p(\Psi_{s_p}^{N_1, N_2}) \right] \right] \\ & \leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\frac{1}{2}-\zeta}} \right) + \frac{C}{N_2^\varphi} + C \mathbb{E} \left[ N_2^\zeta M_t^{N_1, N_2}(x) \right] + \frac{C}{N_2^{\gamma_2-\zeta}} \end{aligned}$$

$$\leq C \left( \frac{1}{N_2^{1-\gamma_2}} + \frac{1}{N_2^{\gamma_2-\zeta}} \right).$$

Hence,  $\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_7(\mathcal{L}, \Psi)] = 0$ . Since  $F_7(\cdot)$  is continuous and uniformly bounded,

$$\lim_{N_2 \rightarrow \infty} \mathbb{E}_{\pi^{N_1, N_2}} [F_7(\mathcal{L}, \Psi)] = \mathbb{E}_{\pi^{N_1}} [F_7(\mathcal{L}, \Psi)] = 0.$$

The result then follows by Prokhorov's theorem.  $\square$

**Appendix E. Derivation of the asymptotic expansion of  $h_t^{N_1, N_2}$  for  $\gamma_2 \in (1/2, 1)$ .** The goal of this section is to provide an inductive argument to derive the asymptotic expansion for  $\langle f, \gamma_t^{N_1, N_2} \rangle$  and  $h_t^{N_1, N_2}$  as  $N_2 \rightarrow \infty$  as claimed in (21) and (22) respectively.

Let  $\nu \in \mathbb{N}$  and let  $\mathcal{G}^{N_1}(x)$  be the Gaussian random variable defined in Section 2. Then, when  $\gamma_2 \in \left[ \frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2} \right)$ , we obtain that for any fixed  $f \in C_b^\infty(\mathbb{R}^{1+N_1(1+d)})$ , as  $N_2 \rightarrow \infty$ , we have the expansion given by (21) where for  $n \geq 3$ ,

$$\begin{aligned} l_{n,t}^{N_1}(f) &= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - Q_{0,s}^{N_1}(x')) \left[ l_{n-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_1}} l_{n-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\ &\quad + \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - Q_{0,s}^{N_1}(x')) \left[ \sum_{k=0}^{n-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{n-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \pi(dx', dy) ds \\ &\quad - \sum_{m=1}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') \left[ l_{m-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_2}} l_{m-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\ &\quad - \sum_{m=1}^{n-1} \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') \left[ \sum_{k=0}^{m-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{m-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \pi(dx', dy) ds, \end{aligned} \tag{61}$$

where

$$\begin{aligned} C_x^{f,1}(\theta) &= \partial_c f(\theta) \sigma(Z^{2,N_1}(x)), \\ C_x^{f,2}(\theta) &= c \sigma'(Z^{2,N_1}(x)) \sigma(w^1 x) \cdot \partial_{w^2} f(\theta), \\ C_x^3(\theta) &= c \sigma'(Z^{2,N_1}(x)) \sigma'(w^1 x) w^2. \end{aligned} \tag{62}$$

As  $N_2 \rightarrow \infty$  and when  $\gamma_2 \in \left[ \frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2} \right]$ , we have the asymptotic expansion (22) for  $h_t^{N_1, N_2}(x)$ . The terms on the right hand side of the asymptotic expansion (22) satisfy the deterministic evolution equations (63), (64) and (65).

$$\begin{aligned} Q_{n,t}^{N_1}(x) &= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - Q_{0,s}^{N_1}(x')) l_{n,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\ &\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} (y - Q_{0,s}^{N_1}(x')) x x' \left[ \sum_{k=0}^n l_{k,s}^{N_1} \left( B_x^{3,j}(\theta) \right) l_{n-k,s}^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \right] \pi(dx', dy) ds \\ &\quad - \sum_{m=0}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') l_{m,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \sum_{m=0}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') x x' \left[ \sum_{k=0}^m l_{k,s}^{N_1} \left( B_x^{3,j}(\theta) \right) l_{m-k,s}^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \right] \pi(dx', dy) ds, \end{aligned} \tag{63}$$

When  $\gamma_2 \in \left(\frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2}\right)$ ,

$$\begin{aligned} & Q_{\nu,t}^{N_1}(x) \\ &= \mathcal{G}(x) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu,s}^{N_1}(x') l_{0,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu,s}^{N_1}(x') x x' l_{0,s}^{N_1} \left( B_x^{3,j}(\theta) \right) l_{0,s}^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \pi(dx', dy) ds. \end{aligned} \quad (64)$$

and when  $\gamma_2 = \frac{2\nu+1}{2\nu+2}$ ,

$$\begin{aligned} & Q_{\nu,t}^{N_1}(x) \\ &= \mathcal{G}^{N_1}(x) + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) l_{\nu,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\ &\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) x x' \left[ \sum_{k=0}^n l_{k,s}^{N_1} \left( B_x^{3,j}(\theta) \right) l_{\nu-k,s}^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \right] \pi(dx', dy) ds \\ &\quad - \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') l_{m,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') x x' \left[ \sum_{k=0}^m l_{k,s}^{N_1} \left( B_x^{3,j}(\theta) \right) l_{m-k,s}^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \right] \pi(dx', dy) ds. \end{aligned} \quad (65)$$

It is interesting to note that the approach that is presented in this section also recovers the rigorously derived formulas for  $\nu = 1$  and  $\nu = 2$  as presented in the main theoretical results of Section 2. Below we focus on presenting the argument for the case  $\nu > 2$ .

**E.1. General  $\nu > 2$  case.** To find an expression for  $Q_{\nu,t}^{N_1}$  for any  $\nu > 2$ , we use an inductive argument. Assuming  $Q_{0,t}^{N_1}(x) = h_t^{N_1}(x)$  and  $l_{0,t}^{N_1}(f) = \langle f, \gamma_0^{N_1} \rangle$ , we have already rigorously shown that the statement holds for  $\nu = 1$  and  $\nu = 2$ . For  $n = 3, \dots, \nu - 1$ , we will assume that  $Q_{n,t}^{N_1}$  and  $l_{n,t}^{N_1}(f)$  satisfy the following deterministic evolution equations,

$$\begin{aligned} & Q_{n,t}^{N_1}(x) \\ &= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) l_{n,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\ &\quad + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) x x' \left[ \sum_{k=0}^n l_{k,s}^{N_1} \left( B_x^{3,j}(\theta) \right) l_{n-k,s}^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \right] \pi(dx', dy) ds \\ &\quad - \sum_{m=0}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') l_{m,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\ &\quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \sum_{m=0}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') x x' \left[ \sum_{k=0}^m l_{k,s}^{N_1} \left( B_x^{3,j}(\theta) \right) l_{m-k,s}^{N_1} \left( B_{x'}^{3,j}(\theta) \right) \right] \pi(dx', dy) ds, \end{aligned}$$

and

$$\begin{aligned}
& l_{n,t}^{N_1}(f) \\
&= \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) \left[ l_{n-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_1}} l_{n-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\
&+ \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) \left[ \sum_{k=0}^{n-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{n-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \pi(dx', dy) ds \\
&- \sum_{m=1}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') \left[ l_{m-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_2}} l_{m-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\
&- \sum_{m=1}^{n-1} \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') \left[ \sum_{k=0}^{m-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{m-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \pi(dx', dy) ds
\end{aligned}$$

We now derive the formulas for  $Q_{\nu,t}^{N_1}$  and  $l_{\nu,t}^{N_1}(f)$  for any  $\nu \in \mathbb{N}$ .

- When  $\gamma_2 \in \left( \frac{2\nu-1}{2\nu}, \frac{2\nu+1}{2\nu+2} \right)$ , plugging equations (22) and (21) into the left hand side of equation (28) gives (the symbol  $\approx$  is used to ignore the remainder terms in (21) and (22))

$$\begin{aligned}
& h_t^{N_1, N_2}(x) - h_0^{N_1, N_2}(x) \\
&\approx \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - \sum_{k=1}^{\nu-1} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k,s}^{N_1}(x') - \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} Q_{\nu,s}^{N_1}(x') \right) \\
&\quad \times \sum_{k=0}^{\nu-1} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\
&+ \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - \sum_{k=0}^{\nu-1} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k,s}^{N_1}(x') - \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} Q_{\nu,s}^{N_1}(x') \right) x x' \\
&\quad \times \left[ \sum_{k=0}^{\nu-1} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k,s}^{N_1} (B_x^{3,j}(\theta)) \right] \left[ \sum_{k=0}^{\nu-1} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds \\
&= \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) l_{0,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \right. \\
&\quad \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) x x' l_{0,s}^{N_1} (B_x^{3,j}(\theta)) l_{0,s}^{N_1} (B_{x'}^{3,j}(\theta)) \pi(dx', dy) ds \right\} \\
&+ \sum_{n=1}^{\nu-1} \frac{1}{N_2^{n(1-\gamma_2)}} \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) l_{n,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \right. \\
&\quad \left. - \sum_{m=0}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') l_{m,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \right. \\
&\quad \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) x x' \left[ \sum_{k=0}^n l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{n-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds \right. \\
&\quad \left. - \frac{1}{N_1} \sum_{j=1}^{N_1} \sum_{m=0}^{n-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{n-m,s}^{N_1}(x') x x' \left[ \sum_{k=0}^m l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{m-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds \right\} \\
&- \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu,s}^{N_1}(x') l_{0,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \right\}
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu, s}^{N_1}(x') x x' l_{0, s}^{N_1}(B_x^{3, j}(\theta)) l_{0, s}^{N_1}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds \Big\} + O(N_2^{-\Omega_2}) \\
& = \sum_{n=0}^{\nu-1} \frac{1}{N_2^{n(1-\gamma_2)}} Q_{n, t}^{N_1}(x) - \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu, s}^{N_1}(x') l_{0, s}^{N_1} \left( B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta) \right) \right. \\
& \quad \left. \pi(dx', dy) ds \right\} + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu, s}^{N_1}(x') x x' l_{0, s}^{N_1}(B_x^{3, j}(\theta)) l_{0, s}^{N_1}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds \Big\} + O(N^{-\Omega_\nu}),
\end{aligned}$$

for some  $\Omega_\nu > \gamma_2 - \frac{1}{2}$ . Adding  $h_0^{N_1, N_2}(x)$  and subtracting  $\sum_{k=0}^{\nu-1} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k, t}^{N_1}(x)$  on both sides, we have

$$\begin{aligned}
& \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} Q_{\nu, t}^{N_1}(x) \\
& = \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} (N_2^{\gamma_2-\frac{1}{2}} h_0^{N_1, N_2}(x)) \\
& \quad - \frac{1}{N_2^{\gamma_2-\frac{1}{2}}} \left\{ \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu, s}^{N_1}(x') l_{0, s}^{N_1} \left( B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta) \right) \pi(dx', dy) ds \right. \\
& \quad \left. + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu, s}^{N_1}(x') x x' l_{0, s}^{N_1}(B_x^{3, j}(\theta)) l_{0, s}^{N_1}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds \right\}.
\end{aligned}$$

Since  $N_2^{\gamma_2-\frac{1}{2}} h_0^{N_1, N_2}(x)$  converges in distribution to the Gaussian random variable  $\mathcal{G}^{N_1}(x)$  defined in (9), we have an expression for  $Q_{\nu, t}^{N_1}$ :

$$\begin{aligned}
Q_{\nu, t}^{N_1}(x) & = \mathcal{G}(x) - \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu, s}^{N_1}(x') l_{0, s}^{N_1} \left( B_{x, x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x, x'}^{2, j}(\theta) \right) \pi(dx', dy) ds \\
& \quad - \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu, s}^{N_1}(x') x x' l_{0, s}^{N_1}(B_x^{3, j}(\theta)) l_{0, s}^{N_1}(B_{x'}^{3, j}(\theta)) \pi(dx', dy) ds.
\end{aligned}$$

which coincides with (64).

- When  $\gamma_2 \geq \frac{2\nu+1}{2\nu+2}$ , we first derive an expression for  $l_t^\nu(f)$  by plugging (22) and (21) into equation (30),

$$\begin{aligned}
& \langle f, \gamma_t^{N_1, N_2} \rangle - \langle f, \gamma_0^{N_1, N_2} \rangle \\
& \approx \frac{1}{N_2^{1-\gamma_2}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k, s}^{N_1}(x') - O(N_2^{-(\nu+1)(1-\gamma_2)}) \right) \\
& \quad \times \left[ \sum_k^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k, s}^{N_1}(C_{x'}^{f, 1}(\theta)) + O(N_2^{-(\nu+1)(1-\gamma_2)}) \right] \pi(dx', dy) ds \\
& \quad + \frac{1}{N_1^{1-\gamma_1} N_2^{1-\gamma_2}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k, s}^{N_1}(x') - O(N_2^{-(\nu+1)(1-\gamma_2)}) \right) \\
& \quad \times \left[ \sum_k^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k, s}^{N_1}(C_{x'}^{f, 2}(\theta)) + O(N_2^{-(\nu+1)(1-\gamma_2)}) \right] \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N_1^{1-\gamma_1} N_2^{1-\gamma_2}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k,s}^{N_1}(x') - O(N_2^{-(\nu+1)(1-\gamma_2)}) \right) \\
& \quad \times \left[ \sum_k^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k,s}^{N_1}(C_{x'}^{f,3}(\theta)) + O(N_2^{-(\nu+1)(1-\gamma_2)}) \right] \\
& \quad \cdot \left[ \sum_k^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} \frac{1}{k(N_2^{1-\gamma_2})} l_{k,s}^{N_1}(\nabla_{w^1} f(\theta)x') + O(N_2^{-(\nu+1)(1-\gamma_2)}) \right] \pi(dx', dy) ds \\
& = \sum_{n=1}^{\nu-1} \frac{1}{N_2^{n(1-\gamma_2)}} l_{n,t}^{N_1}(f) \\
& + \frac{1}{N_2^{\nu(1-\gamma_2)}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) \left[ l_{\nu-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_2}} l_{\nu-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\
& - \frac{1}{N_2^{\nu(1-\gamma_2)}} \sum_{m=1}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') \left[ l_{m-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_2}} l_{m-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\
& + \frac{1}{N_1^{1-\gamma_1} N_2^{\nu(1-\gamma_2)}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) \left[ \sum_{k=0}^{\nu-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{\nu-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \\
& \quad \pi(dx', dy) ds \\
& - \frac{1}{N_1^{1-\gamma_1} N_2^{\nu(1-\gamma_2)}} \sum_{m=1}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') \left[ \sum_{k=0}^{m-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{m-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \\
& \quad \pi(dx', dy) ds \\
& + O(N_2^{-(\nu+1)(1-\gamma_2)})
\end{aligned}$$

Subtracting  $\langle f, \gamma_0^{N_1} \rangle + \sum_{n=1}^{\nu-1} \frac{1}{N_2^{n(1-\gamma_2)}} l_{n,t}^{N_1}(f)$ , multiplying  $N_2^{\nu(1-\gamma_2)}$  on both sides of the above equation, and using the fact that  $N_2^{\nu(1-\gamma_2)} (\langle f, \gamma_0^{N_1, N_2} \rangle - \langle f, \gamma_0^{N_1} \rangle)$  converges to 0 in distribution when  $\gamma_2 \geq \frac{2\nu+1}{2\nu+2}$ , we can get the following evolution equation for  $l_{\nu,t}^{N_1}(f)$ ,

$$\begin{aligned}
& l_{\nu,t}^{N_1}(f) \\
& = \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) \left[ l_{\nu-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_2}} l_{\nu-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\
& - \sum_{m=1}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') \left[ l_{m-1,s}^{N_1}(C_{x'}^{f,1}(\theta)) + \frac{1}{N_1^{1-\gamma_2}} l_{m-1,s}^{N_1}(C_{x'}^{f,2}(\theta)) \right] \pi(dx', dy) ds \\
& + \frac{1}{N_1^{1-\gamma_1}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) \left[ \sum_{k=0}^{\nu-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{\nu-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \pi(dx', dy) ds \\
& - \frac{1}{N_1^{1-\gamma_1}} \sum_{m=1}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') \left[ \sum_{k=0}^{m-1} l_{k,s}^{N_1}(C_{x'}^3(\theta)) \cdot l_{m-1-k,s}^{N_1}(\nabla_{w^1} f(\theta)x') \right] \pi(dx', dy) ds,
\end{aligned}$$

which concludes the inductive step for  $l_{\nu,t}^{N_1}(f)$ .

Next, we derive  $Q_{\nu,t}^{N_1}$  by plugging equations (22) and (21) into the left hand side of equation (28):

$$\begin{aligned}
& h_t^{N_1, N_2}(x) - h_0^{N_1, N_2}(x) \\
& \approx \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k,s}^{N_1}(x') - O(N_2^{(\nu+1)(1-\gamma_2)}) \right)
\end{aligned}$$

$$\begin{aligned}
& \times \left[ \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) + O(N_2^{-(\nu+1)(1-\gamma_2)}) \right] \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} Q_{k,s}^{N_1}(x') - O(N_2^{-(\nu+1)(1-\gamma_2)}) \right) x x' \\
& \times \left[ \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k,s}^{N_1} (B_x^{3,j}(\theta)) + O(N_2^{-(\nu+1)(1-\gamma_2)}) \right] \\
& \times \left[ \sum_{k=0}^{\nu} \frac{1}{N_2^{k(1-\gamma_2)}} l_{k,s}^{N_1} (B_{x'}^{3,j}(\theta)) + O(N_2^{-(\nu+1)(1-\gamma_2)}) \right] \pi(dx', dy) ds \\
& = \sum_{n=0}^{\nu-1} \frac{1}{N_2^{n(1-\gamma_2)}} Q_{n,t}^{N_1}(x) \\
& + \frac{1}{N_2^{\nu(1-\gamma_2)}} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) l_{\nu,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\
& + \frac{1}{N_1 N_2^{\nu(1-\gamma_2)}} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) x x' \left[ \sum_{k=0}^{\nu} l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{\nu-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \\
& \pi(dx', dy) ds \\
& - \frac{1}{N_2^{\nu(1-\gamma_2)}} \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') l_{m,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\
& - \frac{1}{N_1 N_2^{\nu(1-\gamma_2)}} \sum_{j=1}^{N_1} \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') x x' \left[ \sum_{k=0}^m l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{m-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \\
& \pi(dx', dy) ds \\
& + O(N^{-\Omega_{\nu+1}}),
\end{aligned}$$

where  $\Omega_{\nu+1} > (\nu+1)(1-\gamma_2)$ . Following the same idea as earlier, when  $\gamma_2 = \frac{2\nu+1}{2\nu+2}$ , we note that  $\nu(1-\gamma_2) = \gamma_2 - \frac{1}{2} = \frac{\nu}{2\nu+2}$ , we can obtain an expression for  $Q_{\nu,t}^{N_1}$  (which coincides with (65)):

$$\begin{aligned}
& Q_{\nu,t}^{N_1}(x) \\
& = \mathcal{G}^{N_1}(x) + \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) l_{\nu,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) x x' \left[ \sum_{k=0}^n l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{\nu-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds \\
& - \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') l_{m,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\
& - \frac{1}{N_1} \sum_{j=1}^{N_1} \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') x x' \left[ \sum_{k=0}^m l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{m-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds,
\end{aligned}$$

where  $\mathcal{G}(x)$  is the Gaussian random variable. And when  $\gamma_2 > \frac{2\nu+1}{2\nu+2}$ ,  $Q_{\nu,t}^{N_1}$  is driven by the deterministic equation

$$\begin{aligned}
& Q_{\nu,t}^{N_1}(x) \\
& = \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) l_{\nu,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds
\end{aligned}$$

$$\begin{aligned}
& + \frac{1}{N_1} \sum_{j=1}^{N_1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} \left( y - Q_{0,s}^{N_1}(x') \right) x x' \left[ \sum_{k=0}^n l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{\nu-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds \\
& - \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') l_{m,s}^{N_1} \left( B_{x,x'}^1(\theta) + \frac{1}{N_1} \sum_{j=1}^{N_1} B_{x,x'}^{2,j}(\theta) \right) \pi(dx', dy) ds \\
& - \frac{1}{N_1} \sum_{j=1}^{N_1} \sum_{m=0}^{\nu-1} \int_0^t \int_{\mathcal{X} \times \mathcal{Y}} Q_{\nu-m,s}^{N_1}(x') x x' \left[ \sum_{k=0}^m l_{k,s}^{N_1} (B_x^{3,j}(\theta)) l_{m-k,s}^{N_1} (B_{x'}^{3,j}(\theta)) \right] \pi(dx', dy) ds,
\end{aligned}$$

This concludes the inductive step for the derivation of  $Q_{\nu,t}^{N_1}(x)$ .

**Acknowledgments.** We would like to thank the three reviewers for their careful and constructive review of the article.

## REFERENCES

- [1] B. Alipanahi, A. Delong, M. T. Weirauch and B. J. Frey, [Predicting the sequence specificities of DNA-and RNA-binding proteins by deep learning](#), *Nature Biotechnology*, **33** (2015), 831-838.
- [2] D. Araújo, R. I. Oliveira and D. Yukimura, A mean-field limit for certain deep neural networks, 2019, [arXiv:1906.00193](#).
- [3] S. Ö. Arik, M. Chrzanowski, A. Coates, G. Diamos, A. Gibiansky, Y. Kang, X. Li, J. Miller, A. Ng, J. Raiman and S. Sengputa, Deep voice: Real-time neural text-to-speech, [arXiv:1702.07825](#), 2017.
- [4] A. R. Barron, [Approximation and estimation bounds for artificial neural networks](#), *Machine Learning*, **14** (1994), 115-133.
- [5] P. Bartlett, D. Foster and M. Telgarsky, Spectrally-normalized margin bounds for neural networks, *Advances in Neural Information Processing Systems*, (2017), 6241-6250.
- [6] M. Bojarski, D. Del Test, D. Dworakowski, B. Firnier, B. Flepp, P. Goyal, L. Jackel, M. Monfort, U. Muller, J. Zhang, X. Zhang, J. Zhao and K. Zieba, End to end learning for self-driving cars, [arXiv:1604.07316](#), 2016.
- [7] L. Chizat and F. Bach, On the global convergence of gradient descent for over-parameterized models using optimal transport, *Advances in Neural Information Processing Systems (NeurIPS)*, (2018), 3040-3050.
- [8] S. Du, J. Lee, H. Li, L. Wang and X. Zhai, Gradient Descent finds global minima of deep neural networks, *Proceedings of the 36th International Conference on Machine Learning*, Long Beach, California, PMLR 97, 2019.
- [9] S. Du, X. Zhai, B. Poczos and A. Singh, Gradient Descent provably optimizes over-parameterized neural networks, *ICLR*, 2019.
- [10] A. Esteva, B. Kuprel, R. A. Novoa, J. Ko, S. M. Swetter, H. M. Blau and S. Thrun, [Dermatologist-level classification of skin cancer with deep neural networks](#), *Nature*, **542** (2017), 115-118.
- [11] S. N. Ethier and T. G. Kurtz, *Markov Processes: Characterization and Convergence*, 1986, Wiley, New York.
- [12] M. Geiger, A. Jacot, S. Spigler, F. Gabriel, L. Sagun, S. d'Ascoli, G. Biroli, C. Hongler and M. Wyart, [Scaling description of generalization with number of parameters in deep learning](#), *J. Stat. Mech. Theory Exp.*, (2020), 023401, 23 pp. [arXiv:1901.01608](#), 2019
- [13] X. Glorot and Y. Bengio, Understanding the difficulty of training deep feedforward neural networks, *Proceedings of the Thirteenth International Conference on Artificial Intelligence and Statistics*, (2010), 249-256.
- [14] I. Goodfellow, Y. Bengio and A. Courville, *Deep Learning*, Cambridge: MIT Press, 2016.
- [15] S. Gu, E. Holly, T. Lillicrap and S. Levine, [Deep reinforcement learning for robotic manipulation with asynchronous off-policy updates](#), *IEEE Conference on Robotics and Automation*, (2017), 3389-3396.
- [16] K. Hornik, [Approximation capabilities of multilayer feedforward networks](#), *Neural Networks*, **4** (1991), 251-257.
- [17] K. Hornik, M. Stinchcombe and H. White, [Multilayer feedforward networks are universal approximators](#), *Neural Networks*, **2** (1989), 359-366.

- [18] J. Huang and H. T. Yau, Dynamics of deep neural networks and neural tangent hierarchy, In *International Conference on Machine Learning, PMLR*, (2020), 4542-4551.
- [19] Y. Ito, **Nonlinearity creates linear independence**, *Advances in Computational Mathematics*, **5** (1996), 189-203.
- [20] A. Jacot, F. Gabriel and C. Hongler, Neural tangent kernel: Convergence and generalization in neural networks, *32nd Conference on Neural Information Processing Systems (NeurIPS 2018)*, Montreal, Canada.
- [21] A. Krizhevsky, Learning multiple layers of features from tiny images, *Technical Report*, 2009.
- [22] C. Kuan and K. Hornik, **Convergence of learning algorithms with constant learning rates**, *IEEE Transactions on Neural Networks*, **2** (1991), 484-489.
- [23] H. J. Kushner and G. G. Yin, *Stochastic Approximation and Recursive Algorithms and Applications*, Springer-Verlag, New York, 2003.
- [24] Y. LeCun, Y. Bengio and G. Hinton, **Deep learning**, *Nature*, **521** (2015), 436-444.
- [25] Y. LeCun, L. Bottou, Y. Bengio and P. Haffner, **Gradient-based learning applied to document recognition**, *Proceedings of the IEEE*, **86** (1998), 2278-2324.
- [26] Y. Leviathan and Y. Matias, Google duplex: An AI system for accomplishing real-world tasks over the phone, *Google*, 2018.
- [27] J. Ling, A. Kurzawski and J. Templeton, **Reynolds averaged turbulence modelling using deep neural networks with embedded invariance**, *Journal of Fluid Mechanics*, **807** (2016), 155-166.
- [28] J. Ling, R. Jones and J. Templeton, **Machine learning strategies for systems with invariance properties**, *Journal of Computational Physics*, **318** (2016), 22-35.
- [29] S. Mallat, **Understanding deep convolutional neural networks**, *Philosophical Transactions of the Royal Society A*, **374** (2016), 20150203.
- [30] S. Mei, A. Montanari and P.-M. Nguyen, **A mean field view of the landscape of two-layer neural networks**, *Proceedings of the National Academy of Sciences*, **115** (2018), E7665-E7671.
- [31] O. Moynot and M. Samuelides, **Large deviations and mean-field theory for asymmetric random recurrent neural networks**, *Probability Theory and Related Fields*, **123** (2002), 41-75.
- [32] B. Neal, S. Mittal, A. Baratin, V. Tantia, M. Scicluna, S. Lacoste-Julien and I. Mitliagkas, A modern take on the bias-variance tradeoff in neural networks, [arXiv:1810.08591](https://arxiv.org/abs/1810.08591), 2018.
- [33] P.-M. Nguyen, Mean field limit of the learning dynamics of multilayer neural networks, 2019, [arXiv:1902.02880](https://arxiv.org/abs/1902.02880).
- [34] H. A. Pierson and M. S. Gashler, **Deep learning in robotics: A review of recent research**, *Advanced Robotics*, **31** (2017), 821-835.
- [35] G. M. Rotskoff and E. Vanden-Eijnden, Neural networks as interacting particle systems: Asymptotic convexity of the loss landscape and universal scaling of the approximation error, [arXiv:1805.00915](https://arxiv.org/abs/1805.00915), 2018.
- [36] J. Sirignano and R. Cont, **Universal features of price formation in financial markets: Perspectives from deep Learning**, *Quantitative Finance*, **19** (2019), 1449-1459.
- [37] J. Sirignano, A. Sadhwani and K. Giesecke, Deep learning for mortgage risk, [arXiv:1607.02470](https://arxiv.org/abs/1607.02470), 2016.
- [38] J. Sirignano and K. Spiliopoulos, **DGM: A deep learning algorithm for solving partial differential equations**, *Journal of Computational Physics*, **375** (2018), 1339-1364.
- [39] J. Sirignano and K. Spiliopoulos, **Mean field analysis of neural networks: A law of large numbers**, *SIAM Journal on Applied Mathematics*, **80** (2020), 725-752.
- [40] J. Sirignano and K. Spiliopoulos, **Mean field analysis of neural networks: A central limit theorem**, *Stochastic Processes and their Applications*, **130** (2020), 1820-1852.
- [41] J. Sirignano and K. Spiliopoulos, **Mean field analysis of deep neural networks**, *Mathematics of Operations Research*, **47** (2021), 120-152.
- [42] J. Sirignano and K. Spiliopoulos, **Asymptotics of reinforcement learning with neural networks**, *Stochastic Systems*, **12** (2022), 2-29.
- [43] Y. Taigman, M. Yang, M. Ranzato and L. Wolf, **Deepface: Closing the gap to human-level performance in face verification**, In *Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition*, (2014), 1701-1708.
- [44] M. Telgarsky, Benefits of depth in neural networks, [arXiv:1602.04485](https://arxiv.org/abs/1602.04485), 2016.
- [45] J. Yu and K. Spiliopoulos, **Normalization effects on shallow neural networks and related asymptotic expansions**, *AIMS Journal on Foundations of Data Science*, **3** (2021), 151-200.
- [46] Y. Zhang, W. Chan and N. Jaitly, **Very deep convolutional networks for end-to-end speech recognition**, In *IEEE International Conference on Acoustics, Speech, and Signal Processing*, (2017), 4845-4849.

[47] D. Zou, Y. Cao, D. Zhou and Q. Gu, Stochastic gradient descent optimizes over-parameterized deep ReLU networks, [arXiv:1811.08888](https://arxiv.org/abs/1811.08888), 2018.

Received September 2022; revised February 2023; early access March 2023.