The high-dimensional asymptotics of first order methods with random data

Michael Celentano* Chen Cheng† Andrea Montanari‡

December 14, 2021

Abstract

We study a class of deterministic flows in $\mathbb{R}^{d\times k}$, parametrized by a random matrix $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ with i.i.d. centered subgaussian entries. We characterize the asymptotic behavior of these flows over bounded time horizons, in the high-dimensional limit in which $n, d \to \infty$ with k fixed and converging aspect ratios $n/d \to \delta$. The asymptotic characterization we prove is in terms of a system of a nonlinear stochastic process in k dimensions, whose parameters are determined by a fixed point condition. This type of characterization is known in physics as dynamical mean field theory. Rigorous results of this type have been obtained in the past for a few spin glass models. Our proof is based on time discretization and a reduction to certain iterative schemes known as approximate message passing (AMP) algorithms, as opposed to earlier work that was based on large deviations theory and stochastic processes theory. The new approach allows for a more elementary proof and implies that the high-dimensional behavior of the flow is universal with respect to the distribution of the entries of $\boldsymbol{X}$.

As specific applications, we obtain high-dimensional characterizations of gradient flow in some classical models from statistics and machine learning, under a random design assumption.

Contents

*Department of Statistics, University of California Berkeley
†Department of Statistics, Stanford University
‡Department of Electrical Engineering and Department of Statistics, Stanford University

1 Introduction

Understanding the behavior of gradient descent dynamics in non-convex random energy landscapes is a central problem in a number of disciplines, ranging from statistical physics to applied mathematics, machine learning and statistics. Consider to be concrete, the problem of solving n nonlinear equalities or inequalities in unknown vector $\boldsymbol{\theta} \in \mathbb{R}^d$:

$$\text{System of equalities:} \quad \varphi(\langle \boldsymbol{x}_i, \boldsymbol{\theta}\rangle) = y_i\,, \quad \forall i \le n\,, \tag{1}$$

$$\text{System of inequalities:} \quad \varphi(\langle \boldsymbol{x}_i, \boldsymbol{\theta}\rangle) \le y_i\,, \quad \forall i \le n\,. \tag{2}$$

Here $\varphi : \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ is a known function, $\boldsymbol{x}_i \in \mathbb{R}^d$, $y_i \in \mathbb{R}$ are known data of the problem, and $\langle \boldsymbol{u}, \boldsymbol{v}\rangle$ denotes the standard scalar product of $\boldsymbol{u}, \boldsymbol{v} \in \mathbb{R}^d$. We are given y_i and $\boldsymbol{x}_i$, $i \le n$, and the function φ, and would like to solve either Problem (1) or Problem (2) for $\boldsymbol{\theta} \in \mathbb{R}^d$.

An interesting way to explore the space of solutions and near-solutions of these problems is to consider the following gradient flow in $\mathbb{R}^d$:

$$\frac{\mathrm{d}\boldsymbol{\theta}^t}{\mathrm{d}t} = -\nabla \mathcal{L}_n(\boldsymbol{\theta}^t)\,, \tag{3}$$

$$\mathcal{L}_n(\boldsymbol{\theta}) := \frac{d}{n}\sum_{i=1}^n \mathsf{Loss}(\langle \boldsymbol{x}_i, \boldsymbol{\theta}\rangle; y_i)\,. \tag{4}$$

We could take, for instance, $\mathsf{Loss}(z; y) = (\varphi(z) - y)^2$ for the system of equalities (1) and $\mathsf{Loss}(z; y) = (\varphi(z) - y)_+^2$ for the system of inequalities (2).

Apart from being a purely theoretical tool, gradient flow can be discretized to yield a gradient descent algorithm

$$\boldsymbol{\theta}^{k+1} = \boldsymbol{\theta}^k - \eta \nabla \mathcal{L}_n(\boldsymbol{\theta}^k)\,, \tag{5}$$

where η is a stepsize parameter. For small enough η, this algorithm should closely track the gradient flow. In practice, this is to be initialized at an uninformative position, e.g. $\boldsymbol{\theta}^0 = 0$ or $\boldsymbol{\theta}^0 \sim \mathsf{N}(0, c_0 \boldsymbol{I}_d)$. This or closely related flows have been studied for a number of applications, including linear regression [SGB94], phase retrieval [Fie82, CC17], generalized linear models [SHN+18]. We refer to Section 2 for further pointers to the literature.

Taking the derivative in Eq. (3) and adopting the vector notation, the gradient flow reads

$$\frac{\mathrm{d}\boldsymbol{\theta}^t}{\mathrm{d}t} = -\frac{d}{n} \boldsymbol{X}^{\mathsf{T}} \mathsf{Loss}'(\boldsymbol{X}\boldsymbol{\theta}^t; \boldsymbol{y})\,, \tag{6}$$

where $\boldsymbol{X} \in \mathbb{R}^{n\times d}$ is the matrix whose i-th row is given by vector $\boldsymbol{x}_i$, $\boldsymbol{y} = (y_1, \dots, y_n)^{\mathsf{T}}$. We adopt the convention of applying Loss' to vectors componentwise, i.e. $\mathsf{Loss}'(\boldsymbol{z}; \boldsymbol{y}) \in \mathbb{R}^n$ is the vector whose i-th entries is $\mathsf{Loss}(z_i, y_i)$.

Similar flows are studied in statistical physics. For instance, in the Hopfield model of associative memories the memory retrieval dynamics is closely related to the following flow

$$\frac{\mathrm{d}\boldsymbol{\sigma}^t}{\mathrm{d}t} = -\nabla \mathcal{H}(\boldsymbol{\sigma}^t)\,, \tag{7}$$

$$\mathcal{H}(\boldsymbol{\sigma}) := \frac{1}{2} \sum_{i=1}^{n} \langle \boldsymbol{x}_i, \boldsymbol{\sigma} \rangle^2 + V(\boldsymbol{\sigma})\,, \tag{8}$$

where $V(\boldsymbol{\sigma}) := \sum_{i=1}^d V(\sigma_i)$, and $\mathcal{H}(\boldsymbol{\sigma})$ is the system's Hamiltonian. This type of dynamics was studied in the context of the Sherrington-Kirkpatrick model [SZ82], the spherical p-spin glass model [CHS93, CK93], and the Hopfield model [RSZ88].

All of these dynamics can be modified by introducing a noise term, thus yielding a Langevin dynamics. For instance, Eq. (3) can be modified to $\mathrm{d}\boldsymbol{\theta}^t = -\nabla \mathcal{L}(\boldsymbol{\theta}^t)\mathrm{d}t + \alpha \mathrm{d}\boldsymbol{B}^t$, with $(\boldsymbol{B}_t)_{t\ge 0}$ a standard d-dimensional Brownian motion. We believe that this generalization can be treated using our approach, but our formal results are established for the 'zero temperature' case $\alpha = 0$.

In this paper we will be interested in the behavior of the flow (3) (and indeed a significant generalization of the latter), when the matrix $\boldsymbol{X}$ is random, with i.i.d. centered subgaussian entries[1]. We will focus on the high-dimensional limit $d \to \infty$ under a proportional asymptotics, namely $n/d \to \delta \in (0, \infty)$.

The mainstream approach to the analysis of the flow (6) in statistics and applied mathematics is to study the landscape of the cost function $\mathcal{L}_n(\boldsymbol{\theta})$. One then relates the properties of gradient flow to such landscape properties (e.g., the absence of 'bad' local minima) via a deterministic argument. The drawback of this approach is that one cannot account for situations in which, for instance, bad local minima exist but a typical initialization does not fall in their basin of attraction.

In contrast, within statistical physics, there exists a well established approach to the analysis of gradient or Langevin flows for spin glass Hamiltonians. One takes the limit $n, d \to \infty$ at $n/d = \delta$ and $t \le T$ fixed and uses a non-rigorous argument to derive an asymptotic characterization. This characterization can be written as a fixed point condition for the two points correlation functions (the limit below is understood in almost sure sense)

$$C_\theta(t, s) := \lim_{d\to\infty} \frac{1}{d} \langle \boldsymbol{\theta}^t, \boldsymbol{\theta}^s \rangle\,, \tag{9}$$

and a response function as well (see below for definitions). This characterization holds over time horizons of order one as $n, d \to \infty$, and is often referred in physics as 'dynamical mean field theory' (DMFT).

At first sight, studying gradient flow or similar flows over time a time horizon $T = O(1)$ as $n, d \to \infty$ might seem to have little use. Instead it turns out that, in many problems of interest, a non-trivial evolution

[1] Our main theorem can be proved in a slightly stronger form for the case of Gaussian entries, because of available theorems in the literature.

takes place on this time scales, and a near optimum is achieved. Examples from the literature will be provided in the next section. (An important role is of course played by the scaling of the cost function in Eq. (4)). We will also prove that gradient descent, as defined in Eq. (5), closely tracks gradient flow (and the same happens for more general flows) when η is small, but still $O(1)$ as $n, d \to \infty$. This means the theory developed here concerns algorithms whose complexity is of the order of $T/\eta = O(1)$ matrix–vector multiplications.

DMFT asymptotics have been proved in the past for Langevin dynamics on several spin glass models [AG95, AG97, ADG01, ADG06]. These proofs were based either on a large deviations argument or on stochastic processes and weak convergence theory. Over the last few years, physicists have applied the DMFT approach to analyze gradient flow algorithms in several problems from high-dimensional statistics and machine learning (see next section for some pointers). While a DMFT characterization was not proven for these applications, several insights were extracted from the analysis of DMFT systems. The present paper aims at filling this gap.

Namely, we report contributions in several directions:

Asymptotic characterization. We prove an asymptotic DMFT characterization of a class of flows including (4) as a special case. Our setting includes cases in which $\boldsymbol{\theta}^t \in \mathbb{R}^{d\times k}$ is a matrix with a fixed number k of columns (with k independent of d, n). Further, the flow can depend on time, and the function Loss' in Eq. (6) is replaced by a general function $\ell : \mathbb{R}^{k+1} \to \mathbb{R}^k$.

These generalizations allows us to cover a range of applications of interest such as generalized linear models, or shallow neural networks with a constant number of hidden neurons. The characterization we obtain determines, among other things, the asymptotic distribution of trajectories of single coordinates:

$$\frac{1}{d}\sum_{i=1}^{d} \delta_{(\boldsymbol{\theta}_i)_0^T} \Rightarrow \mathrm{P}_T \tag{10}$$

where $(\boldsymbol{\theta}_i)_0^T$ is the trajectory of coordinate i (seen as a function from $[0,T]$ to $\mathbb{R}^k$), and $\Rightarrow$ denotes weak convergence in the space of probability distributions over $C([0,T],\mathbb{R}^k)$.

Proof technique. We introduce a new proof technique that is based on a 3 step procedure: (1) Discretize time; (2) Show that the discrete-time flow can be obtained by applying a simple change of variables to the iterates of an approximate message passing (AMP) algorithm; (3) Apply an existing asymptotic characterization of AMP algorithms, known as 'state evolution'.

Given the special structure of AMP algorithms, and the wealth of rigorous results about them, we believe that this approach is simpler than earlier ones. The only technical challenge is to prove that the time step can be taken to zero the DMFT. We show this by establishing a contraction property in a suitable function space. This contraction property has other useful consequences: among them, it implies existence and uniqueness of solutions of the asymptotic dynamics.

Beyond gradient flow. In applications, gradient flow is only one among other algorithms that we might be interested in. These algorithms need not to be gradient flows with respect to a cost function. For instance, it is known that, among first order methods for statistical estimation (algorithms that proceed by successive multiplication by $\boldsymbol{X}$ or $\boldsymbol{X}^{\mathsf{T}}$), Bayes AMP achieves optimal statistical accuracy, under suitable assumptions [CMW20b].

The proof by discretization and reduction to AMP makes transparent the relation between various algorithms, and in particular the fact that each of these algorithms can be viewed as AMP plus some post-processing.

Non-vanishing step size. As a byproduct of our analysis, the asymptotic characterization does not apply only to the continuous time flow, but also to its discretization (e.g. gradient descent (5)) for any stepsize $\eta = \eta_n \to 0$ as $n, d \to \infty$. This follows from the fact that our proof technique is based on time discretization.

For the case of non-vanishing stepsize, our analysis still gives an asymptotic characterization (with discrete time).

Universality. As our approach leverages available results on the analysis of AMP algorithms, we inherit the generality of those results. In particular, we can establish universality of the $n, d \to \infty$ limit with respect to the distribution of the entries X_{ij} of $\boldsymbol{X}$.

Stationary points. In the case of gradient flows, we study a certain class of stationary points of the DMFT dynamics.

We show that these stationary points are in correspondence with stationary points of an infinite-dimensional variational principle which arises from Gordon's Gaussian comparison inequality. This correspondence holds both for convex and non-convex optimization problems. For convex problems, the infinite-dimensional variational principle is also convex, and typically has a unique minimizer, whence DMFT has a unique stationary point.

In physics language, these stationary points correspond to 'replica symmetric solutions'. A subset of them should describe the long-time asymptotics of the flow. However, we leave for future work the study of how and when these stationary points do actually control the long-time asymptotics. Statistical physics predicts that other types of asymptotic behaviors are possible as well in non-convex problems [CK93].

The rest of the paper is organized as follows. We briefly survey related work in Section 2. We then state our general results in Section 3. We specialize the general result to a few cases of interest in Section 4. We finally present our proofs in Sections 5 and 6, with several technical lemmas deferred to the appendices.

2 Related work

As mentioned in the introduction, DMFT was used by physicists for a long time to characterize the high-dimensional behavior of Langevin dynamics in mean field spin glasses [SZ81, SZ82]. The asymptotic characterization is given, as in our paper, by a correlation and response function. In some cases, these functions are determined by a set of integral-differential equations [CHS93, CK93]. More often, they solve a fixed point condition that is given in terms of a one-dimensional stochastic process with correlated noise and memory [SZ82, CK08, ABC20].

Over the last few years physicists applied the same techniques to several problems in high-dimensional statistics and machine-learning. They studied the behavior of gradient flow learning and extracted useful insights from the DMF characterization. An incomplete list of examples includes tensor principal component analysis [MBC+20], max margin linear classification [ABUZ18, MKUZ20], Gaussian mixture models [MKUZ20].

Some of these papers compare Langevin learning to Bayes optimal AMP, by solving numerically the corresponding high-dimensional characterizations. They observe that Bayes AMP achieves superior accuracy and provide physics-based explanations for this phenomenon. Our analysis (alongside the results of [CMW20b]) provides a simple rigorous explanation of this observation. Langevin (as gradient flow and indeed any first order method) is equivalent to a specific AMP plus post-processing. Bayes AMP is the optimal AMP algorithm in Bayesian estimation problems.

DMFT characterizations for the Langevin dynamics of the Sherrington-Kirkpatrick (SK) model were first proved by Ben Arous and Guionnet [AG95, AG97, Gui97] (who considered continuous spins and Langevin dynamics) and by Grunwald [Gru96] who instead considered Ising spins and Glauber dynamics. Spherical spin glasses (whereby the vector $\boldsymbol{\theta}$ lies on a sphere) were studied in [ADG01] in the case of quadratic cost functions. With respect to all other cases discussed here, the example of spherical spin grasses with quadratic activations is significantly simpler. In this case, the solutions to the flow can be written explicitly. The case of spherical spin glasses with general polynomial interactions (the so called p-spin model) was studied in [ADG06]. This paper proved the DMFT equations using a concentration technique and Girsanov formula, and leveraging in a crucial way the fact that the energy function is a Gaussian process.

We notice in passing that all of the above approaches use in an important way the fact that the process studied is a non-degenerate diffusion, e.g. Langevin dynamics at non-zero temperature. In contrast, we focus on the degenerate case of deterministic flow. We believe it is possible to apply our proof technique to non-zero temperature, by constructing the Brownian noise as a deterministic function of the vector $\boldsymbol{z}$. We also note that the models we treat are analogous to the SK model in that the asymptotic characterization is given in terms of a stochastic process.

Recently the mathematical study of DMFT asymptotics has attracted renewed interest to address the question of universality with respect to the distribution of the underlying randomness (the matrix $\boldsymbol{X}$ in our case). Dembo and Gheissari [DG21] prove universality for a class of diffusions parametrized by a random matrix. Finally, Dembo, Lubetzky and Zeitouni [DLZ19] prove universality for a version of the SK model Langevin dynamics in which the symmetric interaction matrix is replaced by an asymmetric matrix with independent entries. This additional independence allows to use a direct approaches based on Girsanov formula.

Our proof technique is based on a reduction to AMP, and hence allows us to leverage the wealth of results proved in that context. While most of these results [BM11, JM13, BMN20] were proven for Gaussian randomness, using a technique first introduced in [Bol14], universality results were proven in [BLM15, CL21]. In particular, we exploit the result of Chen and Lam [CL21] and deduce universality for a large class of flows.

Finally, we believe the same technique should be applicable to prove universality in other cases as well, e.g. for Langevin dynamics in the SK model which is not covered by our main theorem.

3 Main results

3.1 Setting

The general flow. Let $\ell : \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}^k$, $(r, z, t) \mapsto \ell_t(r; z)$ be a Lipschitz function, and $\Lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$, $t \mapsto \Lambda^t$ be a bounded matrix-valued function.

We then consider the general flow over $\mathbb{R}^{d\times k}$, defined via the following ordinary differential equation, denoted by $\mathfrak{F}(\boldsymbol{\theta}^0, \boldsymbol{z}, \Lambda, \ell)$,

$$\frac{\mathrm{d}\boldsymbol{\theta}^t}{\mathrm{d}t} = -\boldsymbol{\theta}^t\Lambda^t - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t; \boldsymbol{z})\,, \tag{11}$$

with initial condition $\boldsymbol{\theta}^0 \in \mathbb{R}^{d\times k}$. Here we follow the convention of applying functions to matrices row-wise. In particular $\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t; \boldsymbol{z}) \in \mathbb{R}^{d\times k}$ is the matrix with rows

$$\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t; \boldsymbol{z}) = \begin{bmatrix} \ell_t(\boldsymbol{x}_1^\mathsf{T}\boldsymbol{\theta}^t; z_1) \\ \ell_t(\boldsymbol{x}_2^\mathsf{T}\boldsymbol{\theta}^t; z_2) \\ \vdots \\ \ell_t(\boldsymbol{x}_d^\mathsf{T}\boldsymbol{\theta}^t; z_d) \end{bmatrix},$$

where we recall that $\boldsymbol{x}_i$ is the i-th row of $\boldsymbol{X}$. Notice that this setting generalizes the gradient flow equation (3) in a few ways, apart from the fact that $\boldsymbol{\theta}^t$ can now have k columns. First, the function ℓ_t can now depend on the additional argument z as well as on the time t; second, $\ell_t(\,\cdot\,; z)$ is not necessarily the gradient of a cost function; third, the additional term $-\boldsymbol{\theta}^t\Lambda^t$ allows us to include constraints on the norm of $\boldsymbol{\theta}^t$, or regularization terms.

In Section 4 we will illustrate how the additional flexibility introduced here allows to capture applications in statistics and machine learning.

Notational conventions. We use boldface symbols for matrices or vectors whose dimensions diverge, e.g. $\boldsymbol{\theta}^t$, $\boldsymbol{X}$ and so on. Also, we generally use upper case letters for matrices and lower case letters for vectors, with the exception of $n \times k$ or $d \times k$ matrices such as $\boldsymbol{\theta}^t$. We use $\|u\|_2$ to denote the ℓ_2 norm of a vector u. We also use $\|M\|$ and $\|M\|_F$ to denote the operator norm and Frobenius norm of a matrix M. Further notations can be found in Appendix A.

Dynamical Mean Field Theory. Given random variables $(\theta^0, z) \in \mathbb{R}^k \times \mathbb{R}$, a matrix valued function $\Lambda : \mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$ and a Lipschitz function $\ell : \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}_{\geq 0} \to \mathbb{R}^k$, we consider the following system of equations for the unknown deterministic functions $\Gamma : \mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$, $R_\theta, R_\ell, C_\theta, C_\ell : \mathbb{R}_{\geq 0}\times\mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$ and stochastic processes $\theta, r : \mathbb{R}_{\geq 0} \to \mathbb{R}^k$, which we will denote by $\mathfrak{S} := \mathfrak{S}(\theta^0, z, \delta, \Lambda, \ell)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^t = -(\Lambda^t + \Gamma^t)\theta^t - \int_0^t R_\ell(t,s)\theta^s \mathrm{d}s + u^t\,, \qquad u \sim \mathsf{GP}(0, C_\ell/\delta)\,, \tag{12a}$$

$$r^t = -\frac{1}{\delta}\int_0^t R_\theta(t,s)\ell_s(r^s;z)\mathrm{d}s + w^t\,, \qquad w \sim \mathsf{GP}(0, C_\theta)\,, \tag{12b}$$

$$R_\theta(t,s) = \mathbb{E}\left[\frac{\partial\theta^t}{\partial u^s}\right]\,, \qquad 0 \leq s \leq t < \infty\,, \tag{12c}$$

$$R_\ell(t,s) = \mathbb{E}\left[\frac{\partial \ell_t(r^t;z)}{\partial w^s}\right]\,, \qquad 0 \leq s < t < \infty\,, \tag{12d}$$

$$\Gamma^t = \mathbb{E}\left[\nabla_r \ell_t(r^t;z)\right]\,, \tag{12e}$$

$$C_\theta(t,s) = \mathbb{E}\left[\theta^t {\theta^s}^{\mathsf{T}}\right]\,, \qquad 0 \leq s,t < \infty\,, \tag{12f}$$

$$C_\ell(t,s) = \mathbb{E}\left[\ell_t(r^t;z)\ell_s(r^s;z)^{\mathsf{T}}\right]\,, \qquad 0 \leq s,t < \infty\,. \tag{12g}$$

Here the notation $u \sim \mathsf{GP}(0, C_\ell/\delta)$, $w \sim \mathsf{GP}(0, C_\theta)$ means that u, w are independent centered Gaussian processes with covariance kernels C_ℓ/δ and C_θ. We set $R_\theta(t,s) = R_\ell(t,s) = 0$ for $t < s$. On regions R_θ, R_ℓ,

The quantities $\partial\theta^t/\partial u^s$ in Eq. (12c) and $\partial\ell_t(r^t;z)/\partial w^s$ in Eq. (12d) are stochastic processes defined via the following equations:

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\theta^t}{\partial u^s} = -(\Lambda^t + \Gamma^t)\frac{\partial\theta^t}{\partial u^s} - \int_s^t R_\ell(t,s')\frac{\partial\theta^{s'}}{\partial u^s}\mathrm{d}s'\,, \qquad 0 \leq s \leq t < \infty\,, \tag{13a}$$

$$\frac{\partial\ell_t(r^t;z)}{\partial w^s} = \nabla_r\ell_t(r^t;z)\cdot\left(-\frac{1}{\delta}\int_s^t R_\theta(t,s')\frac{\partial \ell_{s'}(r^{s'};z)}{\partial w^s}\mathrm{d}s' - \frac{1}{\delta}R_\theta(t,s)\nabla_r\ell_s(r^s;z)\right)\,, \quad 0 \leq s < t < \infty\,, \tag{13b}$$

with boundary condition $\partial\theta^t/\partial u^t = I$. Note that the first one is a deterministic integral-differential equation and therefore $\partial\theta^t/\partial u^s$ is a deterministic function. On the other hand, $\partial\ell_t(r^t;z)/\partial w^s$ is in general a stochastic process because r^t is.

The rationale for the notations $\partial\theta^t/\partial u^s$ and $\partial\ell_t(r^t;z)/\partial w^s$ is that Eqs. (13) can be heuristically derived as equations for such functional derivatives. However, we will not need to prove that these are the actual functional derivatives.

We set $\partial\theta^t/\partial u^s = 0$ if $s > t$ and $\partial\ell_t(r^t;z)/\partial w^s = 0$ if $s \geq t$.

Remark 3.1. Since $\partial\theta^t/\partial u^s$ is a deterministic function, the expectation operator in Eq. (12c) can be removed. However, this is not immediately clear from Eq. (12c) alone and also to keep uniformity in our definitions, we retain the expectation operator in Eq. (12c).

3.2 Statement of main results

Assumption 1. *The entries $\boldsymbol{X} = (X_{ij})_{i\leq n, j\leq d}$ are given by $X_{ij} = \overline{X}_{ij}/\sqrt{d}$, where $(\overline{X}_{ij})_{i,j\geq 1}$ is a collection of i.i.d. random variables with distribution independent of n, d, such that $\mathbb{E}\{\overline{X}_{ij}\} = 0$, $\mathbb{E}\{\overline{X}_{ij}^2\} = 1$, and $\|\overline{X}_{ij}\|_{\psi_2} \leq C$ for a constant C (here $\|\cdot\|_{\psi_2}$ denotes the sub-Gaussian norm).*

The function $\ell_t(r;z)$ is Lipschitz continuous with Lipschitz continuous Jacobian in t and r. Further, these Lipschitz constants are bounded uniformly over $t \in [0,T]$ and $z \in \mathbb{R}$. Namely there exists some $M_\ell \in \mathbb{R}_{\geq 0}$ such that, for all z, all $r_1, r_2 \in \mathbb{R}^k$ and $t_1, t_2 \in [0,T]$, we have

$$\left\|\ell_{t_1}(r_1;z) - \ell_{t_2}(r_2;z)\right\|_2 \leq M_\ell(\|r_1 - r_2\|_2 + |t_1 - t_2|)\,, \tag{14a}$$

$$\left\|D\ell_{t_1}(r_1;z) - D\ell_{t_2}(r_2;z)\right\| \leq M_\ell(\|r_1 - r_2\|_2 + |t_1 - t_2|)\,, \tag{14b}$$

where

$$D\ell_t(r;z) = \left[\nabla_r\ell_t(r;z) \quad \tfrac{\mathrm{d}}{\mathrm{d}t}\ell_t(r;z)\right].$$

In addition, Λ^t is Lipschitz continuous and symmetric. There exists some $M_\Lambda \in \mathbb{R}_{\ge 0}$ such that $\|\Lambda^t\| \le M_\Lambda$ for all $t \in [0,T]$, and for all $t_1, t_2 \in [0,T]$

$$\left\|\Lambda^{t_1} - \Lambda^{t_2}\right\| \le M_\Lambda |t_1 - t_2|\,. \tag{15}$$

Our first result establishes existence and uniqueness of solutions of the DMFT system $\mathfrak{S}(\theta^0, z, \delta, \Lambda, \ell)$.

Theorem 1. *Under either Assumption 1, suppose the random variables $(\theta^0, z) \in \mathbb{R}^k \times \mathbb{R}$ satisfy*

$$M_{\theta^0,z} := \max\left\{\mathbb{E}\left[\left\|\theta^0\right\|_2^2\right], \sup_{t\in\mathbb{R}_{\ge 0}} \mathbb{E}\left[\|\ell_t(0;z)\|_2^2\right]\right\} < \infty. \tag{16}$$

Given any functions $\Lambda : \mathbb{R}_{\ge 0} \to \mathbb{R}^{k\times k}$ and $\ell : \mathbb{R}^k \times \mathbb{R} \times \mathbb{R}_{\ge 0} \to \mathbb{R}^k$, there exists a solution $(\theta, r, R_\theta, R_\ell, C_\theta, C_\ell)$ for the DMFT system $\mathfrak{S} := \mathfrak{S}(\theta^0, z, \delta, \Lambda, \ell)$ defined through Eqs. (12a) to (12g) and Eqs. (13a) to (13b). The solution is also unique among all (C_θ, R_θ) that are bounded in all compact sets in $\mathbb{R}^2_{\ge 0}$. There further exists nondecreasing functions $\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}, \Phi_{C_\ell} : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ that satisfy

$$\|R_\theta(t,s)\| \le \Phi_{R_\theta}(t-s)\,, \qquad \|R_\ell(t,s)\| \le \Phi_{R_\ell}(t-s)\,, \qquad \forall 0 \le s \le t < \infty\,, \tag{17a}$$

$$\|C_\theta(t,t)\| \le \Phi_{C_\theta}(t)\,, \qquad \|C_\ell(t,t)\| \le \Phi_{C_\ell}(t)\,, \qquad \forall 0 \le t < \infty\,, \tag{17b}$$

$$\left\|\Gamma^t\right\| \le M_\ell\,, \qquad \forall 0 \le t < \infty\,. \tag{17c}$$

Further, the process $(\theta^t)_{t\in[0,T]}$ has continuous sample paths. Finally, the functions $\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}, \Phi_{C_\ell}$ are such that there exists $\lambda := \lambda(\theta^0, z, \delta, M_{\theta^0,z}, M_\ell, M_\Lambda) > 0$ such that

$$\lim_{t\to\infty} e^{-\lambda t} \max\left\{\Phi_{R_\theta}(t), \Phi_{R_\ell}(t), \Phi_{C_\theta}(t), \Phi_{C_\ell}(t)\right\} = 0. \tag{18}$$

The proof and the definitions of $\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}, \Phi_{C_\ell}$ are presented in Section 5, with most technical details deferred to the appendices.

We next prove that the original flow converges —in a suitable sense— to the unique solution of the DMFT system in the proportional asymptotics $n, d \to \infty$, with $n/d \to \delta$.

Theorem 2. *Under Assumption 1, further assume that $n, d \to \infty$ with $n/d \to \delta \in (0, \infty)$. Let $\boldsymbol{z}, \boldsymbol{\theta}^0$ be independent of $\boldsymbol{X}$, and assume that the empirical distributions $\widehat{\mu}_{\theta^0} := d^{-1}\sum_{i=1}^d \delta_{\theta_i^0}$ and $\widehat{\mu}_z := n^{-1}\sum_{i=1}^n \delta_{z_i}$ converge weakly to μ_{θ^0} and μ_z, $\mathbb{E}_{\widehat{\mu}_{\theta^0}}[\|\theta^0\|^2] \to \mathbb{E}_{\mu_{\theta^0}}[\|\theta^0\|^2] < \infty$ and $\mathbb{E}_{\widehat{\mu}_z}[\|z\|^2] \to \mathbb{E}_{\mu_z}[\|z\|^2] < \infty$. Let θ_0^T be the stochastic process that solves $\mathfrak{S}$. Finally, define $\boldsymbol{r}^t := \boldsymbol{X}\boldsymbol{\theta}^t \in \mathbb{R}^{n\times k}$, $t \ge 0$.*

Then, for any distance $d_{\rm W}$ that metrizes weak convergence in $C([0,T], \mathbb{R}^k)$ (for instance $d_{\rm W} = d_{\rm BL}$ the bounded Lipschitz distance), we have

$$\operatorname*{p-lim}_{n,d\to\infty} d_{\rm W}\Big(\frac{1}{d}\sum_{i=1}^d \delta_{(\theta_i)_0^T}, \mathrm{P}_{\theta_0^T}\Big) = 0\,, \tag{19}$$

$$\operatorname*{p-lim}_{n,d\to\infty} d_{\rm W}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{z_i,(r_i)_0^T}, \mathrm{P}_{z,r_0^T}\Big) = 0\,. \tag{20}$$

Here $\operatorname{p-lim}_{n,d\to\infty}$ denotes convergence in probability, $\mathrm{P}_{\theta_0^T}$ denotes the law of $\theta_0^T := (\theta^t)_{0\le t\le T}$, and P_{z,r_0^T} denotes the joint law of z and $r_0^T := (r^t)_{0\le t\le T}$.

The proof of ths theorem is presented in Section 6.

Remark 3.2. Concretely, the convergence in Theorem 2 implies the following. For any $L \in \mathbb{N}$, any times $0 \le t_1 < \cdots \le t_L$, and any bounded continuous functions $\psi : (\mathbb{R}^k)^L \to \mathbb{R}$, mapping $(x_1, \dots, x_L) \mapsto \psi(x_1, \dots, x_L)$, for $x_i \in \mathbb{R}^k$, and $\tilde{\psi} : (\mathbb{R}^k)^L \times \mathbb{R} \to \mathbb{R}$, we have:

$$\operatorname*{p-lim}_{n,d\to\infty} \frac{1}{d}\sum_{i=1}^d \psi(\theta_i^{t_1}, \dots, \theta_i^{t_L}) = \mathbb{E}\big\{\psi(\theta^{t_1}, \dots, \theta^{t_L})\big\}\,, \tag{21}$$

$$\text{p-lim}_{n,d\to\infty} \frac{1}{n}\sum_{i=1}^{n} \tilde{\psi}(r_i^{t_1},\dots,r_i^{t_L},z) = \mathbb{E}\{\tilde{\psi}(r^{t_1},\dots,r^{t_L},z)\}\,. \tag{22}$$

The expectation on the right-hand side is with respect to the processes $(\theta^t)_{t\geq 0}$, $(r^t)_{t\geq 0}$ defined by the DMFT system.

In the case the matrix $\boldsymbol{X}$ has i.i.d. Gaussian entries, the same proof of Section 6 implies a somewhat stronger statement by leveraging the results of [JM13]. Namely, Eq. (21) holds for any continuous functions with at most quadratic growth $|\psi(x)| \leq C(1+\|x\|_2^2)$, $|\tilde{\psi}(x)| \leq C(1+\|x\|_2^2)$.

4 Applications

In this section, we apply Theorem 2 to prove DMFT characterization of gradient flow for generalized linear models and shallow neural networks with a constant number of hidden neurons.

Next, we define a notion of stationary-point solutions of the DMFT system $\mathfrak{S}(\theta^0, z, \delta, \Lambda, \ell)$. We show that these stationary points are characterized by a system of five nonlinear equations, and that they are in correspondence with stationary points of a certain infinite-dimensional variational principle. This variational principle also emerges in the study of global minimizers of the risk via Gordon's comparison inequality. As an illustration, we discuss the case the case of logistic regression.

We next define two classes of applications that are covered by our general results, Theorem 1 and 2.

Generalized linear models The statistician observes n iid pairs $(y_i, \boldsymbol{x}_i)$ where $y_i = \varphi(\langle \boldsymbol{x}_i, \boldsymbol{\theta}^*\rangle, z_i)$ and z_i is noise drawn independently of $\boldsymbol{x}_i$. The goal is to estimate or recover the planted signal $\boldsymbol{\theta}^*$. Ridge regularized empirical risk minimization attempts to minimize the objective

$$\mathcal{L}_n(\boldsymbol{\theta}) := \frac{d}{n}\sum_{i=1}^{n} \mathsf{L}_0(\langle \boldsymbol{x}_i, \boldsymbol{\theta}\rangle; y_i) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2. \tag{23}$$

When $\varphi(r,z) = r + z$, we recover linear regression. When $\varphi(r,z) = \mathbf{1}\{r+z\geq 0\}$ and $z \sim$ Logistic, we recover logistic regression. When $\varphi(r,z) = |r|^2 + z$, we recover a model of noisy phase-retrieval. Alternative choices of φ recover several other popular regression and classification models.

Shallow neural networks with constant number of hidden units For a fixed-constant k, a two-layer neural network with width k and activation $\sigma : \mathbb{R}\to\mathbb{R}$ is the function class containing functions of the form

$$f_{\boldsymbol{\theta}}(\boldsymbol{x}) := \sum_{a=1}^{k} \alpha_a \sigma(\langle \boldsymbol{x}, \boldsymbol{\theta}^{(a)}\rangle), \tag{24}$$

where $\boldsymbol{\theta} \in \mathbb{R}^{d\times k}$ has columns $\{\boldsymbol{\theta}^{(a)}\}_{a\in[k]}$. Given training data $\{(y_i, \boldsymbol{x}_i)\}_{i\in[n]}$, the statistician may fit a neural network by gradient descent on the objective

$$\mathcal{L}_n(\boldsymbol{\theta}) := \frac{d}{n}\sum_{i=1}^{n} \mathsf{L}_0\left(\sum_{a=1}^{k} \alpha_a \sigma(\langle \boldsymbol{x}_i, \boldsymbol{\theta}^{(a)}\rangle); y_i\right) + \frac{\lambda}{2}\sum_{a=1}^{k}\|\boldsymbol{\theta}^{(a)}\|_2^2. \tag{25}$$

An interesting data model assumes that the response depends on a low-dimensional projection of the data, for instance $y_i = \varphi\big((\boldsymbol{\theta}^*)^{\mathsf{T}}\boldsymbol{x}_i; z_i\big)$ where $\varphi : \mathbb{R}^{k+1}\to\mathbb{R}$ is an unknown function, and $\boldsymbol{\theta}^* \in \mathbb{R}^{d\times k}$ is an unknown matrix.

4.1 Flow with planted signal

At first glance, it may appear that Theorem 2 does not apply to gradient flow in the examples above because $\boldsymbol{y}$ is not independent of $\boldsymbol{X}$, whereas the noise $\boldsymbol{z}$ in Theorem 2 is independent of $\boldsymbol{X}$. In fact, gradient flow in these examples can be represented as a cross-section of a flow of the form (11) on a higher dimensional space, so that its DMFT is a instance of Theorem 2.

First note that the cost functions above can be written as

$$\mathcal{L}_n(\boldsymbol{\theta}) := \frac{d}{n}\sum_{i=1}^{n} \mathsf{L}(\boldsymbol{\theta}^{\mathsf{T}}\boldsymbol{x}_i, (\boldsymbol{\theta}^*)^{\mathsf{T}}\boldsymbol{x}_i; z_i) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|_2^2\,, \tag{26}$$

for a suitable function $\mathsf{L} : (\mathbb{R}^k)^2 \to \mathbb{R}$. For instance, in the generalized linear model, we set $k = 1$ and $\mathsf{L}(r, w; z) = \mathsf{L}_0(r, \varphi(w, z))$.

It is convenient to consider a more generalized (non-gradient, time-dependent) flow of the form

$$\frac{\mathrm{d}\boldsymbol{\theta}^t}{\mathrm{d}t} = -\boldsymbol{\theta}^t\Lambda^t - \frac{1}{\delta}\boldsymbol{X}^{\mathsf{T}}\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t, \boldsymbol{X}\boldsymbol{\theta}^*; \boldsymbol{z})\,. \tag{27}$$

We recover gradient flow with respect to the general cost (26) by setting $\Lambda^t = \lambda I_k$ and $\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t, \boldsymbol{X}\boldsymbol{\theta}^*; \boldsymbol{z})_i = \nabla_r\mathsf{L}(r, w; z)$. We can put Eq. (27) into the form of flow (11) by concatenating $\boldsymbol{\theta}^*$ to the iterates $\boldsymbol{\theta}^t$ and considering the flow

$$\frac{\mathrm{d}(\boldsymbol{\theta}^t, \boldsymbol{\theta}^*)}{\mathrm{d}t} = -(\boldsymbol{\theta}^t, \boldsymbol{\theta}^*)\begin{pmatrix}\Lambda^t & 0\\ 0 & 0\end{pmatrix} - \frac{1}{\delta}\boldsymbol{X}^{\mathsf{T}}\begin{pmatrix} | & | \\ \boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t, \boldsymbol{X}\boldsymbol{\theta}^*; \boldsymbol{z}) & 0 \\ | & | \end{pmatrix}. \tag{28}$$

Indeed, the final k-columns on the right-hand side are 0, so that $\boldsymbol{\theta}^*$ does not change along the trajectory. Theorem 2 can be applied directly to this flow. Doing so leads to certain simplifications which allow us to represent the asymptotic characterization as a k-dimensional rather than $2k$-dimensional process. Here we present this characterization as a corollary to Theorem 2.

Also, notice that there is no loss of generality in assuming that $\boldsymbol{\theta}^t$ and $\boldsymbol{\theta}^*$ have the same number of columns. Indeed, we can always accommodate cases in which the number of columns is different by adding some zero columns, and redefining ℓ_t accordingly.

We require the following assumption on the flow defined in Eq. (27), defined in terms of a function $\ell_t : (\mathbb{R}^k)^2 \times \mathbb{R} \to \mathbb{R}^k$.

Assumption 2. *The same conditions of Assumption 1 are required on the random matrix $\boldsymbol{X}$.*

The function $(r, w^*, z) \mapsto \ell_t(r, w^*; z)$ is assumed to be Lipschitz continuous with Lipschitz continuous Jacobian in t, r, and w^*. Further, these Lipschitz constants are bounded uniformly over $t \in [0, T]$ and $z \in \mathbb{R}$. Explicitly, there exists $M_\ell \in \mathbb{R}_{\geq 0}$ such that, for all z, all $r_1, r_2 \in \mathbb{R}^k$ and $t_1, t_2 \in [0, T]$, we have

$$\left\|\ell_{t_1}(r_1, w_1^*; z) - \ell_{t_2}(r_2, w_2^*; z)\right\|_2 \leq M_\ell(\|r_1 - r_2\|_2 + \|w_1^* - w_2^*\|_2 + |t_1 - t_2|)\,, \tag{29a}$$

$$\left\|D\ell_{t_1}(r_1, w_1^*; z) - D\ell_{t_2}(r_2, w_2^*; z)\right\| \leq M_\ell(\|r_1 - r_2\|_2 + \|w_1^* - w_2^*\|_2 + |t_1 - t_2|)\,, \tag{29b}$$

where

$$D\ell_t(r, w^*; z) = \begin{bmatrix}\nabla_r\ell_t(r; z) & \nabla_{w^*}\ell_t(r, w^*; z) & \frac{\mathrm{d}}{\mathrm{d}t}\ell_t(r; z)\end{bmatrix}.$$

In addition, Λ^t is Lipschitz continuous and symmetric. Explicitly, there exists $M_\Lambda \in \mathbb{R}_{\geq 0}$ such that $\|\Lambda^t\| \leq M_\Lambda$ for all $t \in [0, T]$, and for all $t_1, t_2 \in [0, T]$

$$\left\|\Lambda^{t_1} - \Lambda^{t_2}\right\| \leq M_\Lambda|t_1 - t_2|\,. \tag{30}$$

Given random variables $(\theta^0, \theta^*, z) \in (\mathbb{R}^k)^2 \times \mathbb{R}$, we consider the following system of equations $\mathfrak{S} := \mathfrak{S}(\theta^0, \theta^*, z, \delta, \Lambda^t, \ell_t)$ for unknown deterministic functions $\Lambda^t : \mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$, $R_\ell, C_\theta : (\mathbb{R}_{\geq 0} \cup \{*\}) \times (\mathbb{R}_{\geq 0} \cup \{*\}) \to \mathbb{R}^{k\times k}$, $R_\theta, C_\ell : \mathbb{R}_{\geq 0} \times \mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$, and stochastic processes $\theta : (\mathbb{R}_{\geq 0} \cup \{*\}) \to \mathbb{R}^k$, $r : \mathbb{R}_{\geq 0} \to \mathbb{R}^k$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^t = -(\Lambda^t + \Gamma^t)\theta^t - \int_0^t R_\ell(t, s)\theta^s\mathrm{d}s - R_\ell(t, *)\theta^* + u^t\,, \qquad u^t \sim \mathsf{GP}(0, C_\ell/\delta)\,, \tag{31a}$$

$$r^t = -\frac{1}{\delta}\int_0^t R_\theta(t, s)\ell_s(r^s, w^*; z)\mathrm{d}s + w^t\,, \qquad w^t \sim \mathsf{GP}(0, C_\theta)\,, \tag{31b}$$

$$R_\theta(t, s) = \mathbb{E}\left[\frac{\partial\theta^t}{\partial u^s}\right], \qquad 0 \leq s \leq t < \infty\,, \tag{31c}$$

$$R_\ell(t,s) = \mathbb{E}\left[\frac{\partial \ell_t(r^t, w^*; z)}{\partial w^s}\right], \qquad 0 \le s < t < \infty\,, \tag{31d}$$

$$R_\ell(t,*) = \mathbb{E}\left[\frac{\partial \ell_t(r^t, w^*; z)}{\partial w^*}\right], \tag{31e}$$

$$\Gamma^t = \mathbb{E}\left[\nabla_r \ell_t(r^t, w^*; z)\right], \tag{31f}$$

$$C_\theta(t,s) = \mathbb{E}\left[\theta^t \theta^{s\mathsf{T}}\right], \qquad 0 \le s \le t < \infty \text{ or } s = *\,, \tag{31g}$$

$$C_\ell(t,s) = \mathbb{E}\left[\ell_t(r^t, w^*; z)\ell_s(r^s, w^*; z)^{\mathsf{T}}\right], \qquad 0 \le s \le t < \infty\,. \tag{31h}$$

Here u^t, w^t are independent centered Gaussian processes with covariance kernels C_ℓ/δ and C_θ, and $y = \varphi(w^*, z)$, with $w^* \sim \mathsf{N}(0, \mathbb{E}[\theta^*(\theta^*)^{\mathsf{T}}])$. As before, we have $C_\theta(s,t) = C_\theta(t,s)$, $C_\ell(s,t) = C_\ell(t,s)$, and $R_\theta(t,s) = R_\ell(t,s) = 0$ for $t < s$.

The quantities $\partial\theta^t/\partial u^s$, $\partial\ell(r^t, w^*; z)/\partial w^s$, and $\partial\ell(r^t, w^*; z)/\partial w^*$ are uniquely defined via the following integral-differential equations

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\theta^t}{\partial u^s} = -(\Lambda^t + \Gamma^t)\frac{\partial\theta^t}{\partial u^s} - \int_s^t R_\ell(t,s')\frac{\partial\theta^{s'}}{\partial u^s}\mathrm{d}s'\,, \qquad 0 \le s \le t < \infty\,, \tag{32a}$$

$$\frac{\partial\ell(r^t, w^*; z)}{\partial w^s} = \nabla_r \ell(r^t, w^*; y)\cdot\left(-\frac{1}{\delta}\int_s^t R_\theta(t,s')\frac{\partial\ell(r^{s'}, w^*; z)}{\partial w^s}\mathrm{d}s' - \frac{1}{\delta}R_\theta(t,s)\nabla_r\ell(r^s, w^*; z)\right), \quad 0 \le s < t < \infty\,, \tag{32b}$$

$$\frac{\partial\ell(r^t, w^*; z)}{\partial w^*} = -\frac{1}{\delta}\nabla_r\ell(r^t, w^*; y)\int_0^t R_\theta(t,s')\frac{\partial\ell(r^{s'}, w^*; z)}{\partial w^*}\mathrm{d}s' + \nabla_{w^*}\ell(r^t, w^*; z)\,, \qquad 0 \le s < t < \infty\,, \tag{32c}$$

with boundary condition $\partial\theta^t/\partial u^t = I$. Similarly, we set $\partial\theta^t/\partial u^s = 0$ if $s > t$ and $\partial\ell(r^t; z)/\partial w^s = 0$ if $s \ge t$. As before, our notations point to the fact that these quantities are functional derivatives, although we do not prove it formally (and we do not need to).

Corollary 4.1. *Under Assumption 2, suppose the random variables* $(\theta^0, \theta^*, z) \in (\mathbb{R}^k)^2 \times \mathbb{R}$ *satisfy*

$$M_{\theta^*,\theta^*,z} := \max\left\{\mathbb{E}\left[\|\theta^0\|_2^2\right], \mathbb{E}\left[\|\theta^*\|_2^2\right], \sup_{t\in\mathbb{R}_{\ge 0}}\mathbb{E}\left[\|\ell_t(0,0;z)\|_2^2\right]\right\} < \infty. \tag{33}$$

Then the system of equations $\mathfrak{S}(\theta^0, \theta^*, z, \delta, \Lambda^t, \ell_t)$ *defined in Eqs. (31) and (32) has a unique solution.*

Moreover, assume that $n, d \to \infty$ *with* $n/d \to \delta \in (0,\infty)$*. Let* $\boldsymbol{z}, \boldsymbol{\theta}^0, \boldsymbol{\theta}^*$ *be independent of* $\boldsymbol{X}$*, and assume that the empirical distributions* $\widehat{\mu}_{\theta^0,\theta^*} := d^{-1}\sum_{i=1}^d \delta_{(\theta_i^0,\theta_i^*)}$ *and* $\widehat{\mu}_z := n^{-1}\sum_{i=1}^n \delta_{z_i}$ *converge weakly to* μ_{θ^0,θ^*} *and* μ_z*,* $\mathbb{E}_{\widehat{\mu}_{\theta^0,\theta^*}}[\|\theta^0\|^2 + \|\theta^*\|^2] \to \mathbb{E}_{\mu_{\theta^0,\theta^*}}[\|\theta^0\|^2 + \|\theta^*\|^2] < \infty$ *and* $\mathbb{E}_{\widehat{\mu}_z}[\|z\|^2] \to \mathbb{E}_{\mu_z}[\|z\|^2] < \infty$*. Let* θ_0^T *be the stochastic processes that solve the system* $\mathfrak{S}$ *of Eq. (31). Finally, define* $\boldsymbol{r}^t := \boldsymbol{X}\boldsymbol{\theta}^t \in \mathbb{R}^{n\times k}$*,* $t \ge 0$*.*

Then, for any distance d_{W} *that metrizes weak convergence in* $C([0,T], \mathbb{R}^{k_0})$*, for some fixed* k_0 *(for instance* $d_{\mathrm{W}} = d_{\mathrm{BL}}$ *the bounded Lipschitz distance), we have*

$$\operatorname*{p-lim}_{n,d\to\infty} d_{\mathrm{W}}\Big(\frac{1}{d}\sum_{i=1}^d \delta_{\theta_i^*,(\theta_i)_0^T}, \mathrm{P}_{\theta^*,\theta_0^T}\Big) = 0\,, \tag{34}$$

$$\operatorname*{p-lim}_{n,d\to\infty} d_{\mathrm{W}}\Big(\frac{1}{n}\sum_{i=1}^n \delta_{y_i,(r_i)_0^T}, \mathrm{P}_{\varphi(w^*,z),r_0^T}\Big) = 0\,. \tag{35}$$

Corollary 4.1 is proved in Appendix D.

4.2 Stationary point solutions

We consider the case that ℓ_t, Λ^t do not depend on time t, so that we omit the subscripts and instead write ℓ, Λ. We characterize the case that the state evolution Eq. (31) converges exponentially fast as $t \to \infty$, and provide a system of non-linear equations which must be satisfied by its limit.

Definition 4.2. *We say that the DMFT system* $\mathfrak{S} := \mathfrak{S}(\theta^0, \theta^*, z, \delta, \Lambda, \ell)$ *given in Eq.* (31) *and* (32) converges exponentially *if there exists deterministic constants* $C, c > 0$ *and functions* $\Gamma, R_\ell^* \in \mathbb{R}^{k\times k}$, $R_\ell, R_\theta : \mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$, *random functions* $\widehat{R}_\ell : \mathbb{R}_{\geq 0} \to \mathbb{R}^{k\times k}$, *random matrix* $\widehat{R}_\ell^* \in \mathbb{R}^{k\times k}$, *and random variables* $\theta^\infty, r^\infty, u^\infty, w^\infty \in \mathbb{R}^k$ *such that*

$$\begin{gathered}\|r^t - r^\infty\|_{L^2} \leq Ce^{-ct}, \qquad \|\theta^t - \theta^\infty\|_{L^2} \leq Ce^{-ct}, \qquad \|u^t - u^\infty\|_{L^2} \leq Ce^{-ct}, \qquad \|w^t - w^\infty\|_{L^2} \leq Ce^{-ct},\\ \|R_\ell(\cdot) - R_\ell(t, t-\cdot)\|_\infty \to 0, \qquad \|R_\theta(\cdot) - R_\theta(t, t-\cdot)\|_\infty \to 0, \qquad R_\ell(t,*) \to R_\ell^*, \qquad \Gamma^t \to \Gamma^\infty,\\ \frac{\partial \ell(r^{t+s}, w^*; z)}{\partial w^t} \xrightarrow{L_2} \widehat{R}_\ell(t), \qquad \frac{\partial \ell(r^t, w^*; z)}{\partial w^*} \xrightarrow{L_2} \widehat{R}_\ell^*,\end{gathered} \tag{36}$$

where all limits are taken with s *fixed and* $t \to \infty$, *and further, almost surely,*

$$R_\ell(s), R_\ell(t+s, t), R_\theta(s), R_\theta(t+s,t), \widehat{R}_\ell(s) \leq Ce^{-cs}\,. \tag{37}$$

Notice the abuse of notation in the last definition. For instance we use the same notation for $C_\theta(t, t+s)$ and its limit $C_\theta(s) := \lim_{t\to\infty} C_\theta(t, t+s)$. This should not cause confusion in what follows.

The next theorem describes some properties that must be satisfied by such a limit.

Theorem 3. *Assume that the DMFT system* $\mathfrak{S} := \mathfrak{S}(\theta^0, \theta^*, z, \delta, \Lambda, \ell)$ *given in Eq.* (31) *and* (32) converges exponentially. *Then there exist matrices* $R_\ell^\infty, R_\theta^\infty, C_\ell^\infty \in \mathbb{R}^{k\times k}$ *and* $C_\theta^\infty \in \mathbb{R}^{2k\times 2k}$ *such that*

$$\begin{aligned} 0 &= -(\Lambda + R_\ell^\infty)\theta^\infty - R_\ell^*\theta^* + u^\infty, \quad r^\infty = -\frac{1}{\delta} R_\theta^\infty \ell(r^\infty, w^*; z) + w^\infty,\\ C_\theta &= \mathbb{E}[(\theta^{\infty\mathsf{T}}, \theta^{*\mathsf{T}})^\mathsf{T}(\theta^{\infty\mathsf{T}}, \theta^{*\mathsf{T}})]\,, \qquad C_\ell = \mathbb{E}[\ell(r^\infty, w^*; z)\ell(r^\infty, w^*; z)^\mathsf{T}]\,,\end{aligned} \tag{38}$$

where $u^\infty \sim \mathsf{N}(0, C_\ell/\delta)$ *and* $(w^\infty, w^*) \sim \mathsf{N}(0, C_\theta)$. *Here* $R_\ell^*, \theta^\infty, u^\infty, r^\infty, w^\infty$ *are the same as those which appear in Definition 4.2. Moreover,* $R_\ell^\infty, R_\theta^\infty, R_\ell^*$ *satisfy*

$$\begin{aligned} R_\ell^\infty &= \delta\mathbb{E}\Big[\Big(I_k - \Big(I_k + \frac{1}{\delta}\nabla_r \ell(r^\infty, w^*; z)R_\theta^\infty\Big)^{-1}\Big)R_\theta^{-1}\Big], \qquad (R_\theta^\infty)^{-1} = \Lambda + R_\ell^\infty,\\ R_\ell^* &= \mathbb{E}\Big[\Big(I_k + \frac{1}{\delta}\nabla_r \ell(r^\infty, w^*; z)R_\theta^\infty\Big)^{-1}\nabla_{w^*}\ell(r^\infty, w^*; z)\Big],\end{aligned} \tag{39}$$

assuming all inverted matrices are invertible.

We prove Theorem 3 in Appendix D.

In words, if the DMFT dynamics converges exponentially fast to a stationary point, then the latter must be associated to a solution of the system of nonlinear equations in Eq. (39). Let us emphasize that, so far, we assumed $\ell_t(r, w; z) = \ell(r, w; z)$ to be independent of time, but not necessarily that it is the gradient of a cost function.

4.3 Variational principle for stationary points

In this section we return to the case in which ℓ_t is the gradient of a smooth loss function L, namely $\ell_t(r, w; z) = \nabla_r \mathsf{L}(r, w; z)$, and hence the flow we are studying is gradient flow with respect to the cost (26). For the sake of simplicity, we will focus on the case $k = 1$, although the main result of this section, Proposition 4.4, can be generalized to $k > 1$ as well.

It is natural to conjecture that the stationary point solutions described in Theorem 3 correspond to certain stationary points of the cost function $\mathcal{L}_n(\boldsymbol{\theta})$ of Eq. (26). Namely they should correspond to stationary points of $\mathcal{L}_n(\boldsymbol{\theta})$ that can be reached rapidly by gradient flow when started from a random initialization. We present here a variational principle pointing at such a correspondence.

The properties of global minimizers of cost functions of the form $\mathcal{L}_n(\boldsymbol{\theta})$ have attracted considerable attention in the last few years. A few pointers to this literature are [Kar13, TOH15, DM16]. In order to motivate our variational principle, we restate a lower bound that follows from Gordon's Gaussian comparison inequality and was developed in this context starting with [Sto13, TOH15, TAH18].

Proposition 4.3. *Assume $(X_{ij})_{i\le n,j\le d} \sim_{iid} \mathsf{N}(0,1/d)$, and let $\boldsymbol{h}\sim \mathsf{N}(0,\boldsymbol{I}_n)$, $\boldsymbol{g}\sim\mathsf{N}(0,\boldsymbol{I}_d)$ be independent normal vectors. Let $\boldsymbol{\Pi}_{\boldsymbol{\theta}^*}^{\perp} := \boldsymbol{I}_d - \boldsymbol{\theta}^*(\boldsymbol{\theta}^*)^{\mathsf{T}}/\|\boldsymbol{\theta}^*\|_2^2$ be the projector orthogonal to $\boldsymbol{\theta}^*$, and define $\mathcal{F}_n:\mathbb{R}^n\times\mathbb{R}^n\times\mathbb{R}^d\to\mathbb{R}$ as follows*

$$\mathcal{F}_n(\boldsymbol{\xi},\boldsymbol{r},\boldsymbol{\theta}) := \frac{1}{\sqrt{d}}\langle\boldsymbol{\xi},\boldsymbol{h}\rangle\|\boldsymbol{\Pi}_{\boldsymbol{\theta}^*}^{\perp}\boldsymbol{\theta}\|_2 - \frac{1}{\sqrt{d}}\|\boldsymbol{\xi}\|_2\langle\boldsymbol{g},\boldsymbol{\Pi}_{\boldsymbol{\theta}^*}^{\perp}\boldsymbol{\theta}\rangle + \frac{\langle\boldsymbol{\xi},\boldsymbol{w}^*\rangle\langle\boldsymbol{\theta},\boldsymbol{\theta}^*\rangle}{\|\boldsymbol{\theta}^*\|^2} - \langle\boldsymbol{r},\xi\rangle + \frac{d}{n}\sum_{i=1}^n \mathsf{L}(r_i,w_i^*;z_i) + \frac{\lambda}{2}\|\boldsymbol{\theta}\|^2. \tag{40}$$

Then, for any $x\in\mathbb{R}$

$$\mathbb{P}\Big(\min_{\boldsymbol{\theta}\in\mathbb{R}^d}\mathcal{L}_n(\boldsymbol{\theta})\le x\Big) \le 2\mathbb{P}\Big(\min_{(\boldsymbol{\theta},\boldsymbol{r})\in\mathbb{R}^d\times\mathbb{R}^n}\max_{\boldsymbol{\xi}\in\mathbb{R}^n}\mathcal{F}_n(\boldsymbol{\xi},\boldsymbol{r},\boldsymbol{\theta})\Big)\,. \tag{41}$$

Informally, Proposition 4.3 states that $\min_{\boldsymbol{\theta},\boldsymbol{r}}\max_{\boldsymbol{\xi}}\mathcal{F}_n(\boldsymbol{\xi},\boldsymbol{r},\boldsymbol{\theta})$ is roughly a stochastic lower bound on the minimum risk $\min_{\boldsymbol{\theta}\in\mathbb{R}^d}\mathcal{L}_n(\boldsymbol{\theta})$.

Following [MRSY19], we consider an asymptotic version of the function $\mathcal{F}_n$ of Eq. (40). Consider the probability space $(\Omega,\mathcal{F},\mathbb{P}) = ([0,1],\mathcal{B},\mathsf{Unif})$, and let independent random variables $\theta^*\sim\mu_{\theta^*}$, $w^*\sim\mathsf{N}(0,\mathbb{E}[(\theta^*)^2])$, and $g,h\sim\mathsf{N}(0,1)$ be defined on this space (here μ_{θ^*} is the law of the entries of $\boldsymbol{\theta}^*$). Given square integrable random variables ξ,r,θ, we define:

$$\mathcal{F}(\xi,r,\theta) := \langle\xi,h\rangle\|\Pi_{\theta^*}^{\perp}\theta\| - \frac{\|\xi\|}{\delta^{1/2}}\langle g,\theta\rangle + \frac{\langle\xi,w^*\rangle\langle\theta,\theta^*\rangle}{\|\theta^*\|^2} - \langle r,\xi\rangle + \mathbb{E}[\mathsf{L}(r,w^*;z)] + \frac{\lambda}{2}\|\theta\|^2, \tag{42}$$

where the norms, inner products, and projection operators all occur on $L_2([0,1],\mathcal{B},\mathsf{Unif})$. We will not establish a general formal relationship between $\mathcal{F}$ and $\mathcal{F}_n$, although the intuitive relation is clear. We refer to [MRSY19, CMW20a] for examples in which this relation was made precise.

We emphasize that both Proposition 4.3 and the definition of Eq. (42) hold for non-convex L. The main result of this section is that the stationary points of $\mathcal{F}$ are in one-to-one correspondence with the solutions of Eqs. (38), (39).

Proposition 4.4. *Assume $k=1$ and $\ell(r,w^*;z)=\mathsf{L}'(r,w^*;z)$ for a smooth function L (where L' denotes derivative with respect to the first argument). Then the following hold:*

- *Let $(\xi^\infty,r^\infty,\theta^\infty)$ be a stationary point of $\mathcal{F}$. Define $R_\theta^\infty := \delta^{1/2}\langle g,\theta^\infty\rangle/\|\xi^\infty\|$ and assume that $1+(R_\theta^\infty/\delta)\partial_r\ell(r^\infty,w^*;z)>0$ almost surely and $\mathbb{E}[(1+(R_\theta^\infty/\delta)\partial_r\ell(r^\infty,w^*;z))^{-1}]<\infty$. Then there exists unique random variables w^∞,u^∞ and real numbers R_ℓ^∞, R_ℓ^*, C_ℓ, C_θ, such that these form a solution of Eqs. (38), (39).*
- *Viceversa, let $(r^\infty,\theta^\infty,u^\infty,w^\infty,C_\theta,C_\ell,R_\ell^\infty,R_\ell^*,R_\theta^\infty)$ solve Eqs. (38), (39). We further assume that $1+(R_\theta^\infty/\delta)\partial_r\ell(r^\infty,w^*;z)>0$ almost surely. Then defining $\xi^\infty := \ell(r^\infty,w^*;z)$, the triple $(\xi^\infty,r^\infty,\theta^\infty)$ is a stationary point of $\mathcal{F}$.*

Remark 4.1. If L is convex, then the condition $\inf_r\{1+(R_\theta^\infty/\delta)\partial_r\ell(\,\cdot\,,w^*;z)\}>0$ is satisfied everywhere since $\partial_r\ell(\,\cdot\,,w^*;z)>0$. Moreover, the condition $\mathbb{E}[(1+(R_\theta^\infty/\delta)\partial_r\ell(r^\infty,w^*;z))^{-1}]$ is implied by $\inf_r\{1+(R_\theta^\infty/\delta)\partial_r\ell(\,\cdot\,,w^*;z)\}>\epsilon$ for some $\epsilon>0$, which can be viewed as a type of uniform control on the non-convexity of L. In particular, it states that $r\mapsto(1/2)r^2+(R_\theta^\infty/\delta)\mathsf{L}(r,w^*;z)$ is ϵ-strongly convex.

Proof of Proposition 4.4. First, we show that the stationary points of $\mathcal{F}$ solve Eqs. (38) and (39) under the choices

$$\begin{gathered} w^\infty = \|\Pi_{\theta^*}^{\perp}\theta^\infty\|h + \frac{\langle\theta^\infty,\theta^*\rangle}{\|\theta^*\|^2}w^*, \qquad u^\infty = \frac{\|\xi^\infty\|}{\delta^{1/2}}g,\\ R_\ell^\infty = \frac{\langle\xi^\infty,h\rangle}{\|\Pi_{\theta^*}^{\perp}\theta^\infty\|}, \qquad R_\ell^* = \frac{\langle\xi^\infty,w^*\rangle}{\|\theta^*\|^2} - \frac{\langle\xi^\infty,h\rangle\langle\theta^*,\theta^\infty\rangle}{\|\theta^*\|^2\|\Pi_{\theta^*}^{\perp}\theta^\infty\|}, \qquad R_\theta^\infty = \frac{\delta^{1/2}\langle g,\theta^\infty\rangle}{\|\xi^\infty\|},\\ C_\theta = \mathbb{E}[(\theta^{\infty\mathsf{T}},\theta^{*\mathsf{T}})^{\mathsf{T}}(\theta^{\infty\mathsf{T}},\theta^{*\mathsf{T}})]\,, \qquad C_\ell = \mathbb{E}[\ell(r^\infty,w^*;z)\ell(r^\infty,w^*;z)^{\mathsf{T}}]. \end{gathered} \tag{43}$$

Indeed, taking derivatives and doing some rearranging, we find that the stationary points of $\mathcal{F}$ are exactly those triplets $(\xi^\infty, r^\infty, \theta^\infty)$ which satisfy

$$\begin{aligned}\Big(\frac{\langle \xi^\infty, w^*\rangle}{\|\theta^*\|^2} - \frac{\langle \xi^\infty, h\rangle\langle\theta^\infty,\theta^*\rangle}{\|\Pi_{\theta^*}^\perp\theta^\infty\|\|\theta^*\|^2}\Big)\theta^* + \frac{\langle \xi^\infty, h\rangle}{\|\Pi_{\theta^*}^\perp\theta^\infty\|}\theta^\infty - \frac{\|\xi^\infty\|}{\delta^{1/2}}g + \lambda\theta^\infty = 0,\\ \|\Pi_{\theta^*}^\perp\theta^\infty\|h - \frac{\langle g,\theta^\infty\rangle}{\delta^{1/2}\|\xi^\infty\|}\xi^\infty + \frac{\langle\theta^\infty,\theta^*\rangle}{\|\theta^*\|^2}w^* - r^\infty = 0,\\ -\xi^\infty + \ell(r^\infty, w^*; z) = 0.\end{aligned} \tag{44}$$

Under the definitions Eq. (43), the first line of (44) implies the first equation in (38), and the second two lines of (44) imply the second inequation in the first line of (38). Moreover, with C_θ, C_ℓ defined as in (43), w^∞ and u^∞ as defined in Eq. (43) satisfy $w^\infty \sim \mathsf{N}(0, C_\theta)$ and $u^\infty \sim \mathsf{N}(0, C_\ell/\delta)$ (where we have used, from (44), that $\xi^\infty = \ell(r^\infty, w^*; z)$).

Note that $1 + (R_\theta/\delta)\partial\ell(\cdot, w^*; z) > 0$ implies that $\eta(r^\infty; w^*, z) := r^\infty + (R_\theta^\infty/\delta)\ell(r^\infty, w^*; z)$ is a strictly increasing function of r^∞. It is also continuous. Thus, $\eta(\,\cdot\,; w^*, z)$ has am inverse. Note that $r^\infty = \eta^{-1}(w^\infty; w^*, z)$. It remains only to check that $R_\ell^\infty, R_\ell^*, R_\theta^\infty$ satisfy (39). Using (43) and that $r^\infty = \eta^{-1}(w^\infty; w^*, z)$, we rewrite the first equation in Eq. (44) as

$$w^\infty - \frac{R_\theta^\infty}{\delta}\ell(\eta^{-1}(w^\infty; w^*, z), w^*; z) - \eta^{-1}(w^\infty; w^*, z) = 0. \tag{45}$$

Taking the derivative with respect to w^∞ and rearranging, we conclude that almost surely

$$\partial_{w^\infty}\eta^{-1}(w^\infty; w^*, z) = \Big(1 + (R_\theta^\infty/\delta)\partial_r\ell(r^\infty, w^*; z)\Big)^{-1}. \tag{46}$$

In particular, η^{-1} is differentiable almost everywhere. Further, it derivative is integrable. Thus, we may apply Gaussian integration by parts. Using Eq. (43), we get

$$\begin{aligned}R_\ell^\infty &= \frac{\langle h, \ell(\eta^{-1}(w^\infty; w^*, z), w^*; z)\rangle}{\|\Pi_{\theta^*}^\perp\theta^\infty\|} = \mathbb{E}[\partial_r\ell(r^\infty, w^*; z)\partial_{w^\infty}\eta^{-1}(w^\infty; w^*, z)]\\ &= \frac{\delta}{R_\theta^\infty}\mathbb{E}\Big[1 - \Big(1 + (R_\theta^\infty/\delta)\partial_r\ell(r^\infty, w^*; z)\Big)^{-1}\Big],\end{aligned} \tag{47}$$

in agreement with Eq. (39). Likewise, taking the derivative with respect to w^* and rearranging gives

$$\partial_{w^*}\eta^{-1}(w^\infty, w^*; z) = -\frac{(R_\theta/\delta)\partial_{w^*}\ell(r^\infty, w^*; z)}{1 + (R_\theta/\delta)\partial_r\ell(r^\infty, w^*; z)}, \tag{48}$$

which is bounded above by a w^*, z-dependent constant. Using Eq. (43), we can check that the Gaussian variable $w^* - \langle\theta^*, \theta^\infty\rangle/\|\Pi_{\theta^*}^\perp\theta^\infty\|$ is uncorrelated and hence independent of w^∞. Thus, using Eq. (43) and Gaussian integration by parts, we get

$$\begin{aligned}R_\ell^* &= \frac{1}{\|\theta^*\|^2}\|w^*\|^2\mathbb{E}\big[\partial_r\ell(r^\infty, w^*; z)\partial_{w^*}\eta^{-1}(w^\infty; w^*, z) + \partial_{w^*}\ell(r^\infty, w^*; z)\big]\\ &= \mathbb{E}\Big[\Big(1 + (R_\theta/\delta)\partial_r\ell(r^\infty, w^*; z)\Big)^{-1}\partial_{w^*}\ell(r^\infty, w^*; z)\Big],\end{aligned} \tag{49}$$

in agreement with Eq. (39). Finally, one can directly compute using Eq. (44)

$$R_\theta^\infty = \frac{\delta^{1/2}}{\|\xi^\infty\|}\Big\langle g, \frac{(\|\xi^\infty\|/\delta^{1/2})}{R_\ell^\infty + \lambda}g\Big\rangle = \big(R_\ell^\infty + \lambda\big)^{-1}, \tag{50}$$

in agreement with Eq. (39).

We now show the second part of the proposition. Consider a solution to Eqs. (38) and (39) which satisfies $1 + (R_\theta^\infty/\delta)\partial_r\ell(r, w^*; z) > 0$ almost surely. First, observe that the equations for C_θ and C_ℓ in Eq. (43) hold by Eq. (38). Then we can view the first two Eqs. (43) as definitions for g, h and the second equation of Eq. (44) as the definition of ξ^∞, so that these equations hold by definition. Further, because $\inf_r\{1 + (R_\theta^\infty/\delta)\partial_r \ell(r, w^*; z)\} > 0$ almost surely, we have that $r^\infty = \eta^{-1}(w^\infty; w^*, z)$ by the same argument as above, where η^{-1} is almost surely differentiable with derivative (48). Moreover, by Eq. (39) and (48), we see that $\eta^{-1}(w^\infty; w^*, z)$ is integrable. Then, using Gaussian integration by parts and the same computations we performed above, we can check that the equations for R_ℓ^∞, R_ℓ^*, and R_θ^∞ in Eq. (43) also hold. Finally, the first and last equations of Eq. (44) are then equivalent to the first two equations of Eq. (38).

The proof is complete. □

4.4 An example: logistic regression

The previous section shows that the equations (38) and (39) are equivalent to those derived from Gordon's inequality [TOH15]. Alternative derivations of these fixed-point equations use tools from random matrix theory or Approximate Message Passing theory [Kar13, DM16, SC19]. After an appropriate change of variables, the derived equations are equivalent to (38) and (39), though the appropriate change of variables and the calculations translating one form of the fixed-point equations to another are often non-trivial. Here, we provide the explicit change of variables for the example of logistic regression which recovers the fixed-point equations appearing in [SC19]. In Appendix D.3, we provide the calculations which use this change of variables to derive Eq. (5) of [SC19] from Eqs. (38) and (39) of the current work.

In logistic regression, we take $\varphi(r, z) = 2 \cdot \mathbf{1}\{r + z \geq 0\} - 1$ and $z \sim$ Logistic. Let $\rho(t) = \log(1 + e^t)$, and note that $\mathcal{L}(r, w^*; z) = \rho(-yr)$, where $y = \varphi(w^*, z)$. In this case, $\ell(r, w^*; z) = -y\rho'(-yr) = -\varphi(w^*, z)\rho'(-\varphi(w^*, z))$, and $\ell'(r, w^*; z) = \rho''(-yr) = \rho''(-\varphi(w^*, z)r)$. Here we have $\lambda = 0$. The change of variables (with the notation of [SC19] on the right) is

$$
\begin{gathered}
(C_\theta)_{22} \to \gamma^2, \qquad (C_\theta)_{12}/(C_\theta)_{22} = -R_\ell^*/R_\ell^\infty \to \alpha, \\
\delta((C_\theta)_{11} - (C_\theta)_{12}(C_\theta)_{22}^{-1}(C_\theta)_{21}) = C_\ell/(R_\ell^\infty)^2 \to \sigma^2 \\
\delta^{-1} \to \kappa, \qquad R_\theta/\delta = 1/(\delta R_\ell^\infty) \to \lambda, \\
w^* \to Q_1, \qquad w^\infty \to -Q_2.
\end{gathered}
\tag{51}
$$

We refer the reader to Appendix D.3 for the calculations which derive Eq. (5) of [SC19] from Eqs. (38) and (39) of the current work.

5 Proof of Theorem 1

The proof is divided into the following parts.

I. Define the auxiliary real-valued functions $\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}$ and Φ_{C_ℓ}.

II. Construct a metric space $\mathcal{S} := \mathcal{S}(\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}, \Phi_{C_\ell}, T)$ for the function triplet (C_ℓ, R_ℓ, Γ) when $t, s \in [0, T]$ and also a space $\overline{\mathcal{S}}$ for (C_θ, R_θ). Show that in these spaces, the stochastic processes θ^t, r^t and functional derivatives $\partial\theta^t/\partial u^s, \partial\ell_t(r^t; z)/\partial w^s$ are uniquely defined.

III. Define a transformation $\mathcal{T} : \mathcal{S} \to \mathcal{S}$ such that for any solution of $\mathfrak{S}$, (C_ℓ, R_ℓ, Γ) must be a fixed-point of $\mathcal{T}$. We then show $\mathcal{T}$ is a contraction mapping. By Banach fixed-point theorem it then follows that $\mathcal{T}$ has a unique fixed-point. Finally, we recover the unique solution of $\mathfrak{S}$ from the fixed-point.

We first introduce some norms of functions that will be helpful throughout the proof. For any real-valued function $f(t)$ on $[0, T]$, $\lambda \geq 0$ and $T \in [0, \infty)$, we define the following norms

$$
|||f|||_{\lambda,T} := \int_0^T e^{-\lambda t}|f(t)|\mathrm{d}t\,, \tag{52}
$$

$$
\|f\|_{\lambda,T} := \sup_{0\leq t\leq T} e^{-\lambda t}|f(t)|\,. \tag{53}
$$

Suppose $f : \mathbb{R}_{\geq 0} \to \mathbb{R}^{k_1\times k_2}$ is vector-valued or matrix-valued, we use the same notation to represent

$$
|||f|||_{\lambda,T} := \int_0^T e^{-\lambda t}\,\|f(t)\|\,\mathrm{d}t\,, \tag{54}
$$

$$\|f\|_{\lambda,T} := \sup_{0\le t\le T} e^{-\lambda t} \|f(t)\| \,, \tag{55}$$

and $|||f|||_{\lambda,\infty} = \lim_{T\to\infty} |||f|||_{\lambda,T}, \|f\|_{\lambda,\infty} = \lim_{T\to\infty} \|f\|_{\lambda,T}$.

Part I: The auxiliary real-valued functions. Given constants $\alpha \in \mathbb{R}^{10}, \beta \in \mathbb{R}^2$, consider the deterministic integral-differential system of equations $\overline{\mathfrak{S}} := \overline{\mathfrak{S}}(\alpha;\beta)$ $f_i : \mathbb{R}_{\ge 0} \to \mathbb{R}_{\ge 0}$ for $i = 1,2,3,4$,

$$\frac{\mathrm{d}}{\mathrm{d}t} f_1(t) = \alpha_1 f_1(t) + \alpha_2 f_2(t)\,, \tag{56a}$$

$$f_2(t) = \alpha_3 f_1(t) + \alpha_4 \int_0^t f_1(t-s) f_2(s)\mathrm{d}s\,, \tag{56b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{f_3(t)} = \sqrt{\alpha_5 f_3(t) + \alpha_6 f_4(t) + \alpha_7 \int_0^t (t-s+1)^2 f_2(t-s)^2 f_3(s)\mathrm{d}s}\,, \tag{56c}$$

$$f_4(t) = \alpha_8 + \alpha_9 f_3(t) + \alpha_{10} \int_0^t (t-s+1)^2 f_1(t-s)^2 f_4(s)\mathrm{d}s\,, \tag{56d}$$

with boundary conditions $f_1(t) = \beta_1, f_3(t) = \beta_2$. The following lemma shows there exists nondecreasing functions that solve $\overline{\mathfrak{S}}(\alpha;\beta)$ when $\alpha > 0$ and $\beta > 0$. The proof is postponed to Appendix B.1.1.

Lemma 5.1. *Suppose $\alpha \in \mathbb{R}^{10}, \beta \in \mathbb{R}^2$ have strictly positive coordinates. Then there exists nondecreasing functions (f_1, f_2, f_3, f_4) solving the system $\overline{\mathfrak{S}}(\alpha;\beta)$ defined by Eqs. (56a) to (56d) when $t \in [0,\infty)$, and there exists some $\lambda := \lambda(\alpha,\beta) > 0$ such that*

$$\lim_{t\to\infty} e^{-\lambda t} \max\{f_1(t), f_2(t), f_3(t), f_4(t)\} = 0\,. \tag{57}$$

The functions f_i are unique in the space $L^1_{\mathrm{loc}}(\mathbb{R}_{\ge 0})$ of locally integrable functions.

We consider the following system for nonnegative functions $\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}$ and Φ_{C_ℓ} on $[0,\infty)$.

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{R_\theta}(t) = (M_\Lambda + M_\ell)\Phi_{R_\theta}(t) + \int_0^t \Phi_{R_\ell}(t-s)\Phi_{R_\theta}(s)\mathrm{d}s\,, \tag{58a}$$

$$\Phi_{R_\ell}(t) = \frac{M_\ell}{\delta}\cdot\left\{M_\ell \Phi_{R_\theta}(t) + \int_0^t \Phi_{R_\theta}(t-s)\Phi_{R_\ell}(s)\mathrm{d}s\right\}\,, \tag{58b}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\Phi_{C_\theta}(t)} = \sqrt{3\cdot\left\{(M_\Lambda + M_\ell)^2\Phi_{C_\theta}(t) + \frac{k}{\delta}\Phi_{C_\ell}(t) + \int_0^t (t-s+1)^2 \Phi_{R_\ell}(t-s)^2 \Phi_{C_\theta}(s)\mathrm{d}s\right\}}\,, \tag{58c}$$

$$\Phi_{C_\ell}(t) = 3\cdot\left\{M_{\theta^0,z} + kM_\ell^2\Phi_{C_\theta}(t) + \frac{M_\ell^2}{\delta^2}\int_0^t (t-s+1)^2\Phi_{R_\theta}(t-s)^2\Phi_{C_\ell}(s)\mathrm{d}s\right\}\,, \tag{58d}$$

Applying Lemma 5.1 shows that for any given $\Phi_{R_\theta}(0) > 0, \Phi_{C_\theta}(0) > 0$, the above system has a unique solution in the space of locally integrable functions on $\mathbb{R}_{\ge 0}$.

Part II: The function space $\mathcal{S}$, $\overline{\mathcal{S}}$. We next define a function space in which we solve $\mathfrak{S}$.

Definition 5.2 (The function triplet spaces $\mathcal{S}$ and $\mathcal{S}_{\mathrm{cont}}$)**.** *For $T > 0$, denote by $X = (C_\ell, R_\ell, \Gamma)$ the function triplet $C_\ell, R_\ell : [0,T]^2 \to \mathbb{R}^{k\times k}$ and $\Gamma^t : [0,T] \to \mathbb{R}^{k\times k}$. Define the space*

$$\mathcal{S} := \mathcal{S}(\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}, \Phi_{C_\ell}, T) \tag{59}$$

of all X such that the following hold

1. *$C_\ell(t,s)$ is a covariance kernel (satisfying in particular $C_\ell(t,s)^\top = C_\ell(s,t)$), whose diagonal is dominated as follows*

$$\|C_\ell(t,t)\| \le \Phi_{C_\ell}(t)\,, \tag{60}$$

for $t \in [0,T]$ *and*

$$C_\ell(0,0) = \mathbb{E}\left[\ell_0(r^0;z)\ell_0(r^0;z)^\top\right], \quad r^0 \sim \mathsf{N}\left(0, \mathbb{E}\left[\theta^0 {\theta^0}^\top\right]\right). \tag{61}$$

$C_\ell(t,s)$ *is continuous for* $s \le t$ *and* $s,t \in [0,T]\backslash P$ *where* P *is a finite set. Moreover, for any* $s \le t$ *such that* C_ℓ *is continuous in* $[s,t]^2$

$$\|C_\ell(t,t) - C_\ell(t,s) - C_\ell(s,t) + C_\ell(s,s)\| \le M_{\mathcal{S}}(t-s)^2. \tag{62}$$

where

$$M_{\mathcal{S}} := 3M_\ell \Bigg\{ \left(\frac{1}{\delta^2}(T+1)^2\Phi_{R_\theta}(T)^2 + \frac{3k^2}{\delta}\right)\Phi_{C_\ell}(T) + 3k\left((M_\Lambda + M_\ell)^2 + T(T+1)^2\Phi_{R_\ell}(T)^2\right)\Phi_{C_\theta}(T) + 1\Bigg\}. \tag{63}$$

2. $R_\ell(t,s)$ *is measurable and* $R_\ell(t,s) = 0$ *when* $t \le s$*. Further for any* $s \le t$ *and* $s,t \in [0,T]$

$$\|R_\ell(t,s)\| \le \Phi_{R_\ell}(t-s). \tag{64}$$

Γ^t *is measurable in* $[0,T]$ *such that*

$$\left\|\Gamma^t\right\| \le M_\ell, \tag{65}$$

and

$$\Gamma_0 = \mathbb{E}\left[\nabla_r \ell(r^0, z)\right], \quad r^0 \sim \mathsf{N}\left(0, \mathbb{E}\left[\theta^0 {\theta^0}^\top\right]\right). \tag{66}$$

Moreover, we define the space $\mathcal{S}_{\text{cont}} \subset \mathcal{S}$ *of all* X *such that* $P = \emptyset$ *in the first condition and for all* $s, s' \in [0,t]$,

$$\|C_\ell(t,s) - C_\ell(t,s')\| \le \sqrt{\Phi_{C_\ell}(T)M_{\mathcal{S}}} \cdot |s - s'|. \tag{67}$$

Next we consider the function pairs (C_θ, R_θ) when $t,s \in [0,T]$ and $C_\theta, R_\theta : [0,T]^2 \to \mathbb{R}^{k\times k}$.

Definition 5.3 (The function pair spaces $\overline{\mathcal{S}}$ and $\overline{\mathcal{S}}_{\text{cont}}$)**.** *Denote by* $Y = (C_\theta, R_\theta)$*, we consider the space space*

$$\overline{\mathcal{S}} := \overline{\mathcal{S}}(\Phi_{R_\theta}, \Phi_{R_\ell}, \Phi_{C_\theta}, \Phi_{C_\ell}, T) \tag{68}$$

for all Y *such that*

1. $C_\theta(t,s)$ *is a covariance kernel (satisfying in particular* $C_\theta(t,s)^\top = C_\theta(s,t)$*), whose diagonal is dominated in spectral norm as*

$$\|C_\theta(t,t)\| \le \Phi_{C_\theta}(t), \tag{69}$$

for $t \in [0,T]$ *and*

$$C_\theta(0,0) = \mathbb{E}\left[\theta^0 {\theta^0}^\top\right]. \tag{70}$$

$C_\theta(t,s)$ *is continuous for all* $s \le t$ *and* $s \le t$ *and* $s,t \in [0,T]\backslash P$ *where* P *is a finite set. Moreover, for any* $s \le t$ *such that* $\overline{C}_\theta(t,s)$ *is continuous in* $[s,t]^2$ *with*

$$\|C_\theta(t,t) - C_\theta(t,s) - C_\theta(s,t) + C_\theta(s,s)\| \le M_{\overline{\mathcal{S}}}(t-s)^2, \tag{71}$$

where

$$M_{\overline{\mathcal{S}}} := 3\left\{\left[(M_\Lambda + M_\ell)^2 + T(T+1)^2\Phi_{R_\ell}(T)^2\right]\Phi_{C_\theta}(T) + \frac{k}{\delta}\Phi_{C_\ell}(T)\right\}. \tag{72}$$

2. $R_\theta(t,s)$ is measurable and $R_\ell(t,s)=0$ when $t<s$. It is also satisfies for any $s\le t$ and $s,t\in[0,T]$

$$\|R_\theta(t,s)\| \le \Phi_{R_\theta}(t-s)\,. \tag{73}$$

Moreover, we define the space $\overline{\mathcal{S}}_{\text{cont}} \subset \overline{\mathcal{S}}$ of all X such that $P=\emptyset$ in the first condition and for all $s,s'\in[0,t]$,

$$\left\|\overline{C}_\theta(t,s)-\overline{C}_\theta(t,s')\right\| \le \sqrt{\Phi_{C_\theta}(T)M_{\overline{\mathcal{S}}}\cdot|s-s'|}\,. \tag{74}$$

The following lemma shows the stochastic processes θ^t, r^t and the functional derivatives $\partial\theta^t/\partial u^s, \partial\ell_t(r^t;z)/\partial w^s$ are well-defined whenever $(C_\ell,R_\ell,\Gamma)\in\mathcal{S}$ and $(C_\theta,R_\theta)\in\overline{\mathcal{S}}$. Its proof can be found in Appendix B.2.

Lemma 5.4. *For any fixed $T>0$, initialization θ^0 and $(C_\ell,R_\ell,\Gamma)\in\mathcal{S}$, $(C_\theta,R_\theta)\in\overline{\mathcal{S}}$, the functions $\theta^t, r^t, \partial\theta^t/\partial u^s, \partial\ell_t(r^t;z)/\partial w^s$ are uniquely defined by Eqs. (12a), (12b), (13a) and (13b).*

We endow $\mathcal{S}$ and $\overline{\mathcal{S}}$ with distances. To begin with, we define the (λ,T)-distance for two Gaussian processes u_1 and u_2 on $\mathbb{R}^k$ by the following formula

$$\mathsf{dist}_{\lambda,T}(u_1,u_2) := \inf_{(u_1,u_2)\sim\gamma\in\Gamma(u_1,u_2)} \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}\left[\|u_1^t-u_2^t\|_2^2\right]}\,, \tag{75}$$

where $\lambda>0$ and $\Gamma(u_1,u_2)$ is the collection of all couplings between the two Gaussian processes u_1,u_2. The for any pair of positive semi-definite kernel functions $C^1,C^2:[0,T]^2\to\mathbb{R}^{k\times k}$, we define their (λ,T)-distance by

$$\mathsf{dist}_{\lambda,T}\left(C^1,C^2\right) := \mathsf{dist}_{\lambda,T}(g_1,g_2)\,, \tag{76}$$

where g_1 and g_2 are two centered Gaussian processes with covariance kernels C^1 and C^2. By Minkowski inequality, for a third kernel C_3 the triangle inequality holds

$$\mathsf{dist}_{\lambda,T}\left(C^1,C^3\right) \le \mathsf{dist}_{\lambda,T}\left(C^1,C^2\right)+\mathsf{dist}_{\lambda,T}\left(C^2,C^3\right)\,.$$

For function pairs $R_\ell^i,\Gamma_i,R_\theta^i$, $i=1,2$, we overload the notation $\mathsf{dist}_{\lambda,T}(\cdot,\cdot)$ to define distances

$$\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right) := \sup_{0\le s<t\le T} e^{-\lambda t}\left\|R_\ell^1(t,s)-R_\ell^2(t,s)\right\|\,, \tag{77a}$$

$$\mathsf{dist}_{\lambda,T}(\Gamma_1,\Gamma_2) := \sup_{0\le t\le T} e^{-\lambda t}\left\|\Gamma_1^t-\Gamma_2^t\right\|\,, \tag{77b}$$

$$\mathsf{dist}_{\lambda,T}\left(R_\theta^1,R_\theta^2\right) := \sup_{0\le s\le t\le T} e^{-\lambda t}\left\|R_\theta^1(t,s)-R_\theta^2(t,s)\right\|\,. \tag{77c}$$

It can be seen the triangle inequality still holds for the function spaces on which R_ℓ,Γ and R_θ are defined. Finally, for any $X_i=(C_\ell^i,R_\ell^i,\Gamma_i)\in\mathcal{S}$ and $Y_i=(C_\theta^i,R_\theta^i)\in\overline{\mathcal{S}}$ and $i=1,2$, we define the distances

$$\mathsf{dist}_{\lambda,T}(X_1,X_2) := \mathsf{dist}_{\lambda,T}\left(C_\ell^1,C_\ell^2\right)+\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right)+\mathsf{dist}_{\lambda,T}(\Gamma_1,\Gamma_2)\,, \tag{78a}$$

$$\mathsf{dist}_{\lambda,T}(Y_1,Y_2) := \mathsf{dist}_{\lambda,T}\left(C_\theta^1,C_\theta^2\right)+\mathsf{dist}_{\lambda,T}\left(R_\theta^1,R_\theta^2\right)\,. \tag{78b}$$

Part III: The contraction mapping $\mathcal{T}$. In this part, we will define the mapping $\mathcal{T}:\mathcal{S}\to\mathcal{S}$ such that any solution of $\mathfrak{S}$ must be a fixed-point of $\mathcal{T}$ a nd show the mapping is a contraction. We will define $\mathcal{T}$ by

$$\mathcal{T}(X) := \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}\circ\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X)\,, \tag{79}$$

where $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}:(C_\ell,R_\ell,\Gamma)\mapsto(\overline{C}_\theta,\overline{R}_\theta)$ and $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}:(\overline{C}_\theta,\overline{R}_\theta)\mapsto\overline{X}:=(\overline{C}_\ell,\overline{R}_\ell,\overline{\Gamma})$, so that $\mathcal{T}:X=(C_\ell,R_\ell,\Gamma)\mapsto\overline{X}=(\overline{C}_\ell,\overline{R}_\ell,\overline{\Gamma})$. Specifically, the mapping $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}$ is defined by first solving Eqs. (12a) and (13a)

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^t = -(\Lambda^t+\Gamma^t)\theta^t-\int_0^t R_\ell(t,s)\theta^s\mathrm{d}s+u^t\,, \qquad u^t\sim\mathsf{GP}(0,C_\ell/\delta)\,, \tag{80}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\theta^t}{\partial u^s} = -(\Lambda^t+\Gamma^t)\frac{\partial\theta^t}{\partial u^s} - \int_s^t R_\ell(t,s')\frac{\partial\theta^{s'}}{\partial u^s}\mathrm{d}s'\,, \qquad 0\le s\le t\le T\,, \tag{81}$$

with boundary condition $\partial\theta^t/\partial u^t = I$. We know θ^t and $\partial\theta^t/\partial u^s$ are uniquely defined by Lemma 5.4. This allows us to define the mapping $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(C_\ell, R_\ell, \Gamma) = (\overline{C}_\theta, \overline{R}_\theta)$ via

$$\overline{C}_\theta(t,s) = \mathbb{E}\left[\theta^t\theta^{s\mathsf{T}}\right]\,, \qquad 0\le s\le t\le T\,, \tag{82}$$

$$\overline{R}_\theta(t,s) = \mathbb{E}\left[\frac{\partial\theta^t}{\partial u^s}\right]\,, \qquad 0\le s\le t\le T\,, \tag{83}$$

with the convention $\overline{C}_\theta(t,s) = \overline{R}_\theta(t,s) = 0$ for all $t<s$. Similarly, we define $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(\overline{C}_\theta,\overline{R}_\theta) = (\overline{C}_\ell,\overline{R}_\ell,\overline{\Gamma})$ through Eqs. (12b) and (13b), namely

$$r^t = -\frac{1}{\delta}\int_0^t \overline{R}_\theta(t,s)\ell_s(r^s;z)\mathrm{d}s + w^t\,, \qquad w^t\sim\mathsf{GP}(0,\overline{C}_\theta)\,, \tag{84}$$

$$\frac{\partial\ell_t(r^t;z)}{\partial w^s} = \nabla_r\ell_t(r^t;z)\cdot\left(-\frac{1}{\delta}\int_s^t\overline{R}_\theta(t,s')\frac{\partial\ell_{s'}(r^{s'};z)}{\partial w^s}\mathrm{d}s' - \frac{1}{\delta}\overline{R}_\theta(t,s)\nabla_r\ell_s(r^s;z)\right)\,, \quad 0\le s<t\le T\,. \tag{85}$$

The random functions $r^t, \partial\ell_t(r^t;z)/\partial w^s$ are uniquely defined by Lemma 5.4. The mapping then determined by setting

$$\overline{C}_\ell(t,s) = \mathbb{E}\left[\ell_t(r^t;z)\ell_s(r^s;z)^\mathsf{T}\right]\,, \qquad 0\le s\le t<T\,, \tag{86}$$

$$\overline{R}_\ell(t,s) = \mathbb{E}\left[\frac{\partial\ell_t(r^t;z)}{\partial w^s}\right]\,, \qquad 0\le s<t<T\,, \tag{87}$$

$$\overline{\Gamma}^t = \mathbb{E}\left[\nabla_r\ell_t(r^t;z)\right]\,, \qquad 0\le t\le T\,, \tag{88}$$

and on regions in $\mathbb{R}_{\ge 0}\times\mathbb{R}_{\ge 0}$ that are not defined for $\overline{C}_\ell, \overline{R}_\ell$, we set their values to be zero. The next lemma shows we can choose $\Phi_{R_\theta}(0)$ and $\Phi_{C_\theta}(0)$ large enough such that $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}:\mathcal{S}\to\overline{\mathcal{S}}$ and $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}:\overline{\mathcal{S}}\to\mathcal{S}$. Its proof can be found in Appendix B.3.

Lemma 5.5. *Under the same assumptions of Theorem 1, suppose $\Phi_{C_\theta}(0) > M_{\theta^0,z}$ and $\Phi_{R_\theta}(0) > 1$. Then $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}$ maps $\mathcal{S}$ into $\overline{\mathcal{S}}_{\rm cont}\subset\overline{\mathcal{S}}$ and $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}$ maps $\overline{\mathcal{S}}$ into $\mathcal{S}_{\rm cont}\subset\mathcal{S}$. In particular, this implies $\mathcal{T} = \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}\circ\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}$ maps $\mathcal{S}$ into $\mathcal{S}_{\rm cont}\subset\mathcal{S}$.*

Next, we want to show $\mathcal{T}$ is a contraction mapping under the $\mathsf{dist}_{\lambda,T}(\cdot,\cdot)$ metric defined in Eq. (78a). To this end, we need the following lemmas for the transformation $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}$.

Lemma 5.6. *Suppose $X_1 = (C_\ell^1, R_\ell^1, \Gamma_1), X_2 = (C_\ell^2, R_\ell^2, \Gamma_2)\in\mathcal{S}$, and further $R_\ell^1 = R_\ell^2$ on $[0,T]^2$ and $\Gamma_1=\Gamma_2$ on $[0,T]$. Let $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_i) = (\overline{C}_\theta^i, \overline{R}_\theta^i)$ for $i=1,2$, then we have for any $\epsilon>0$,*

$$\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right) \le \epsilon\cdot\mathsf{dist}_{\lambda,T}\left(C_\ell^1,C_\ell^2\right)\,, \tag{89a}$$

$$\overline{R}_\theta^1 = \overline{R}_\theta^2\,, \tag{89b}$$

for all $\lambda\ge\overline{\lambda}_1 := \overline{\lambda}_1(\varepsilon,\mathcal{S})$.

We defer its proof to Appendix B.4.1.

Lemma 5.7. *Suppose $X_1 = (C_\ell^1, R_\ell^1, \Gamma_1), X_2 = (C_\ell^2, R_\ell^2, \Gamma_2)\in\mathcal{S}$ and $C_\ell^1 = C_\ell^2$ on $[0,T]^2$. Let $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_i) = (\overline{C}_\theta^i, \overline{R}_\theta^i)$ for $i=1,2$, then we have for any $\epsilon>0$,*

$$\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right) \le \epsilon\cdot\left(\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right)+\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)\right)\,, \tag{90a}$$

$$\mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right) \le \epsilon\cdot\left(\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right)+\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)\right)\,. \tag{90b}$$

for all $\lambda\ge\overline{\lambda}_2 := \overline{\lambda}_2(\varepsilon,\mathcal{S})$.

We defer the proof to Appendix B.4.2. We next derive the lemmas for the transformation $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}$.

Lemma 5.8. *Suppose $Y_1 = (\overline{C}_\theta^1, \overline{R}_\theta^1), Y_2 = (\overline{C}_\theta^2, \overline{R}_\theta^2) \in \overline{\mathcal{S}}$ and $\overline{R}_\theta^1 = \overline{R}_\theta^2$ on $[0,T]^2$. Let $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_i) = (\overline{C}_\ell^i, \overline{R}_\ell^i, \overline{\Gamma}_i)$ for $i = 1, 2$, then there exists a constant $M := M(\mathcal{S})$ such that*

$$\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\overline{C}_\ell^1, \overline{C}_\ell^2\right) &\le M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right), && (91)\\
\mathsf{dist}_{\lambda,T}\left(\overline{R}_\ell^1, \overline{R}_\ell^2\right) &\le M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right), && (92)\\
\mathsf{dist}_{\lambda,T}\left(\overline{\Gamma}^1, \overline{\Gamma}^2\right) &\le M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right), && (93)
\end{aligned}$$

for all $\lambda \ge \overline{\lambda}_3 := \overline{\lambda}_3(\mathcal{S})$.

We defer the proof to Appendix B.4.3.

Lemma 5.9. *Suppose $Y_1 = (\overline{C}_\theta^1, \overline{R}_\theta^1), Y_2 = (\overline{C}_\theta^2, \overline{R}_\theta^2) \in \overline{\mathcal{S}}$ and $\overline{C}_\theta^1 = \overline{C}_\theta^2$ on $[0,T]^2$. Let $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_i) = (\overline{C}_\ell^i, \overline{R}_\ell^i, \overline{\Gamma}_i)$ for $i = 1, 2$, then there exists a constant $M := M(\mathcal{S})$ such that*

$$\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\overline{C}_\ell^1, \overline{C}_\ell^2\right) &\le M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right), && (94)\\
\mathsf{dist}_{\lambda,T}\left(\overline{R}_\ell^1, \overline{R}_\ell^2\right) &\le M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right), && (95)\\
\mathsf{dist}_{\lambda,T}\left(\overline{\Gamma}^1, \overline{\Gamma}^2\right) &\le M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right), && (96)
\end{aligned}$$

for all $\lambda \ge \overline{\lambda}_4 := \overline{\lambda}_4(\mathcal{S})$.

We defer the proof to Appendix B.4.4. Now, we are ready to show that $\mathcal{T}$ is a contraction. We take the constant M to be the maximum one among that of Lemma 5.8 and 5.9. Then we take $\epsilon = (12M)^{-1}$, and any $\lambda \ge \max\left\{\overline{\lambda}_1, \overline{\lambda}_2, \overline{\lambda}_3, \overline{\lambda}_4\right\}$, where $\overline{\lambda}_i$ are defined in above lemmas. For any $X_1 = (C_\ell^1, R_\ell^1, \Gamma_1), X_2 = (C_\ell^2, R_\ell^2, \Gamma_2) \in \mathcal{S}$, we set $Y_i = \mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_i) = (\overline{C}_\theta^i, \overline{R}_\theta^i)$ for $i = 1, 2$. We also define $Y_3 = (\overline{C}_\theta^1, \overline{R}_\theta^2)$, and thus

$$\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\mathcal{T}(X_1), \mathcal{T}(X_2)\right) &= \mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_1), \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_2)\right)\\
&\le \mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_1), \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_3)\right) + \mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_2), \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_3)\right). && (97)
\end{aligned}$$

We can then control $\mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_1), \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_3)\right)$ by Lemma 5.9 and $\mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_2), \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_3)\right)$ by Lemma 5.8 which further gives

$$\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\mathcal{T}(X_1), \mathcal{T}(X_2)\right) &\le \mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_1), \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_3)\right) + \mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_2), \mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}(Y_3)\right)\\
&\le 3M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right) + 3M \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right)\\
&= 3M\mathsf{dist}_{\lambda,T}\left(Y_1, Y_2\right)\\
&= 3M\mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_1), \mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_2)\right). && (98)
\end{aligned}$$

Then, we take $X_3 = (C_\ell^1, R_\ell^2, \Gamma_2)$ and apply Lemma 5.6 and 5.7,

$$\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_1), \mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_2)\right) &\le \mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_1), \mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_3)\right) + \mathsf{dist}_{\lambda,T}\left(\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_2), \mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X_3)\right)\\
&\le 2\epsilon \cdot \left(\mathsf{dist}_{\lambda,T}\left(R_\ell^1, R_\ell^2\right) + \mathsf{dist}_{\lambda,T}\left(\Gamma_1, \Gamma_2\right)\right) + \epsilon \cdot \mathsf{dist}_{\lambda,T}\left(C_\ell^1, C_\ell^2\right)\\
&\le 2\epsilon \cdot \left(\mathsf{dist}_{\lambda,T}\left(C_\ell^1, C_\ell^2\right) + \mathsf{dist}_{\lambda,T}\left(R_\ell^1, R_\ell^2\right) + \mathsf{dist}_{\lambda,T}\left(\Gamma_1, \Gamma_2\right)\right)\\
&\le 2\epsilon \cdot \mathsf{dist}_{\lambda,T}\left(X_1, X_2\right). && (99)
\end{aligned}$$

Substitute into Eq. (98),

$$\mathsf{dist}_{\lambda,T}\left(\mathcal{T}(X_1), \mathcal{T}(X_2)\right) \le 3M \cdot 2\epsilon\mathsf{dist}_{\lambda,T}\left(X_1, X_2\right) = (6M\epsilon)\mathsf{dist}_{\lambda,T}\left(X_1, X_2\right) = \frac{1}{2}\mathsf{dist}_{\lambda,T}\left(X_1, X_2\right), \quad (100)$$

provided that $\epsilon = (12M)^{-1}$.

Since T is finite, it follows from Eqs. (77a) to (77c) that the distances $\mathsf{dist}_{\lambda,T}(\cdot,\cdot)$ for R_ℓ, R_θ and Γ are equivalent to L^∞ distance on $L^\infty([0,T]^2)$ and $L^\infty([0,T])$. By Banach fixed-point theorem, we obtain the uniqueness and existence for the fixed-point R_ℓ, R_θ and Γ in $\mathcal{S}$ and $\overline{\mathcal{S}}$.

It requires a slightly longer argument to prove existence and uniqueness of C_ℓ, C_θ. Consider a pair of covariance kernels $C^1, C^2 : [0,T]^2 \to \mathbb{R}^{k\times k}$ and centered Gaussian processes g_1, g_2 with covariances C_1, C_2. If C^1, C^2 satisfy the conditions of $\mathcal{S}$ (cf. Definition 5.2),

$$
\begin{aligned}
\left\|C^1(t,s) - C^2(t,s)\right\| &= \left\|\mathbb{E}\left[g_1^t {g_1^s}^{\mathsf{T}}\right] - \mathbb{E}\left[g_2^t {g_2^s}^{\mathsf{T}}\right]\right\| \\
&\le \left\|\mathbb{E}\left[g_1^t \left(g_1^s - g_2^s\right)^{\mathsf{T}}\right]\right\| + \left\|\mathbb{E}\left[\left(g_1^t - g_2^t\right) {g_2^s}^{\mathsf{T}}\right]\right\| \\
&\le \sqrt{\mathbb{E}\left[\|g_1^t\|_2^2\right]\cdot \mathbb{E}\left[\|g_1^s - g_2^s\|_2^2\right]} + \sqrt{\mathbb{E}\left[\|g_2^s\|_2^2\right]\cdot \mathbb{E}\left[\|g_1^t - g_2^t\|_2^2\right]} \\
&= \sqrt{\mathrm{Tr}\left(\mathbb{E}\left[g_1^t {g_1^t}^{\mathsf{T}}\right]\right)\cdot \mathbb{E}\left[\|g_1^s - g_2^s\|_2^2\right]} + \sqrt{\mathrm{Tr}\left(\mathbb{E}\left[g_1^s {g_1^s}^{\mathsf{T}}\right]\right)\cdot \mathbb{E}\left[\|g_1^t - g_2^t\|_2^2\right]} \\
&\le \sqrt{k\Phi_{C_\theta}(T)e^{\lambda T}}\cdot e^{-\lambda T}\left(\mathbb{E}\left[\|g_1^s - g_2^s\|_2^2\right] + \mathbb{E}\left[\left\|g_1^t - g_2^t\right\|_2^2\right]\right). \qquad (101)
\end{aligned}
$$

Taking infimum over all couplings of g_1 and g_2 we have

$$
\left\|C^1(t,s) - C^2(t,s)\right\| \le 2\sqrt{k\Phi_{C_\theta}(T)}e^{\lambda T}\mathsf{dist}_{\lambda,T}(C_1, C_2). \qquad (102)
$$

This implies any Cauchy sequence under the metric $\mathsf{dist}_{\lambda,T}(\cdot,\cdot)$ is also a Cauchy sequence in $L^\infty([0,T]^2)$. As we have shown in Lemma 5.5, although we allow the input covariance kernel C_θ to be only piecewise continuous, the output $\overline{C}_\theta$ by mapping $\mathcal{T}$ is always continuous. Moreover, from Eq. (74) we actually have $\overline{C}_\theta$ is Lipschitz continuous with a uniform Lipschitz constant.

By completeness of Lipschitz continuous functions under the L^∞ metric, we can apply Banach fixed-point theorem once more and conclude that there exists a unique $X \in \mathcal{S}$ such that

$$
\mathcal{T}(X) = X. \qquad (103)
$$

To see the uniqueness when C_θ, R_θ are bounded functions in any compact set, we can simply take $\Phi_{C_\theta}(0) \to \infty, \Phi_{R_\theta}(0) \to \infty$.

6 Proof of Theorem 2

We first present a proof roadmap.

I. Discretization. For the general flow system $\mathfrak{F}$ we construct a discrete approximation by Euler's method with step size $\eta > 0$, yielding a system $\mathfrak{F}^\eta$. We show that $\mathfrak{F}^\eta$ is an approximation of $\mathfrak{F}$ which is uniformly good in the high dimensional asymptotics $n, d \to \infty$, $n/d \to \delta$.

II. We introduce a discrete time approximation $\mathfrak{S}^\eta$ for the DMFT system $\mathfrak{S}$. We prove that $\mathfrak{S}^\eta$ exactly characterizes the asymptotics of $\mathfrak{F}^\eta$. This is done by showing that the discretized system is equivalent an AMP algorithm plus post-processing, and leveraging existing analysis of AMP.

III. We prove that the unique solution of $\mathfrak{S}^\eta$ converges to the unique solution of $\mathfrak{S}$ as $\eta \to 0$. The latter therefore characterizes the general flow system $\mathfrak{F}$.

Throughout this section, we denote by d_{W} any distance that metrizes weak convergence of probability distributions in $\mathbb{R}^m$ and with an abuse of notation, weak convergence in $C([0,T], \mathbb{R}^k)$. For instance, we can take $d_{\mathrm{W}} = d_{\mathrm{BL}}$ the bounded Lipschitz distance

$$
d_{\mathrm{BL}}(\mu,\nu) := \sup\left\{\int f\mathrm{d}\mu - \int f\mathrm{d}\nu : \|f\|_\infty \le 1,\ \|f\|_{\mathrm{Lip}} \le 1\right\}. \qquad (104)
$$

We further denote by $\mathrm{W}_2(\mu,\nu)$ the Wasserstein-2 distance between μ and ν. Also, we will focus on proving Eq. (19), since (20) follows by repeating the same argument.

Part I: Discrete time approximation of the flow. For the general flow $\mathfrak{F}$ (whose definition we copy from Eq. (11))

$$\frac{\mathrm{d}\boldsymbol{\theta}^t}{\mathrm{d}t} = -\boldsymbol{\theta}^t\Lambda^t - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t;\boldsymbol{z})\,,$$

we consider a discrete time approximation with step size $\eta > 0$. For all $t_i = i\eta$ and $i \in \mathbb{Z}_{\geq 0}$, we set $\boldsymbol{\theta}^0_\eta = \boldsymbol{\theta}^0$ and

$$\boldsymbol{\theta}^{t_{i+1}}_\eta = \boldsymbol{\theta}^{t_i}_\eta + \eta\cdot\left\{-\boldsymbol{\theta}^{t_i}_\eta\Lambda^{t_i} - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_{t_i}(\boldsymbol{X}\boldsymbol{\theta}^{t_i}_\eta;\boldsymbol{z})\right\}. \tag{105}$$

This defines $\boldsymbol{\theta}^t_\eta$ on all $t_i = i\eta$. We extend it to $t \to \mathbb{R}_{\geq 0}$ as a piecewise linear function. Specifically, we define

$$\lfloor t\rfloor := \max\left\{i\eta \mid i\eta \leq t, i \in \mathbb{Z}_{\geq 0}\right\}\,, \tag{106}$$

and the flow $\mathfrak{F}^\eta$ is given by

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\theta}^t_\eta = -\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor} - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{X}\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta;\boldsymbol{z})\,. \tag{107}$$

Consider the empirical distributions of the rows of $\boldsymbol{\theta}^{\tau_1},\cdots,\boldsymbol{\theta}^{\tau_m}$ and $\boldsymbol{\theta}^{\tau_1}_\eta,\cdots,\boldsymbol{\theta}^{\tau_m}_\eta$ for any $\tau_1,\cdots,\tau_m \in \mathbb{R}_{\geq 0}$, denoted by

$$\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}} := \frac{1}{d}\sum_{j=1}^d \delta_{\left(\boldsymbol{\theta}^{\tau_1}_j,\cdots,\boldsymbol{\theta}^{\tau_m}_j\right)}\,, \tag{108}$$

$$\widehat{\mu}_{\theta^{\tau_1}_\eta,\cdots,\theta^{\tau_m}_\eta} := \frac{1}{d}\sum_{j=1}^d \delta_{\left((\boldsymbol{\theta}^{\tau_1}_\eta)_j,\cdots,(\boldsymbol{\theta}^{\tau_m}_\eta)_j\right)}\,, \tag{109}$$

where $\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}$ and $\widehat{\mu}_{\theta^{\tau_1}_\eta,\cdots,\theta^{\tau_m}_\eta}$ are probability distributions in $\mathbb{R}^{km}$. The following lemma controls the distance between the two distributions uniformly with respect to $n, d \to \infty$. We defer its proof to Appendix C.1.

Lemma 6.1. *Under the same assumptions of Theorem 2, consider the gradient flow $\boldsymbol{\theta}^t$ and its piecewise linear approximation $\boldsymbol{\theta}^t_\eta$ by forward Euler with step size $\eta > 0$ and the same initialization $\boldsymbol{\theta}^0$. Then, almost surely, for any $t \geq 0$:*

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{1}{\sqrt{d}}\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}^s_\eta\right\| = 0\,. \tag{110}$$

As a consequence, for any $\tau_1,\cdots,\tau_m \in [0,T]$, we have almost surely that

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\mathrm{W}_2\left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \widehat{\mu}_{\theta^{\tau_1}_\eta,\cdots,\theta^{\tau_m}_\eta}\right) = 0\,. \tag{111}$$

In particular, since convergence in Wasserstein-2 distance implies weak convergence, we also have almost surely that

$$\lim_{\eta\to 0}\limsup_{n\to\infty} d_{\mathrm{W}}\left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \widehat{\mu}_{\theta^{\tau_1}_\eta,\cdots,\theta^{\tau_m}_\eta}\right) = 0\,. \tag{112}$$

Part II: Characterizing discrete flow $\mathfrak{F}^\eta$ by reduction to AMP. In [CMW20b], the authors show that the exact a general first order method of the type $\mathfrak{F}^\eta$ can be reduced to an AMP algorithm followed by a post-processing operation that operates row-wise on $\boldsymbol{\theta}^t_\eta$ and $\boldsymbol{r}^t_\eta$ (and across multiple times). This allows us to leverage existing high-dimensional characterizations of AMP that go under the name of 'state evolution' [BM11, BLM15, JM13, CL21].

We will use the notation $\lceil t\rceil := \lfloor t\rfloor + \eta$. We introduce the following DMFT system $\mathfrak{S}^\eta := \mathfrak{S}^\eta(\theta^0, \theta^*, z, \delta, \lambda, \ell)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^t_\eta = -(\Lambda^{\lfloor t\rfloor} + \Gamma^{\lfloor t\rfloor}_\eta)\theta^{\lfloor t\rfloor}_\eta - \int_0^{\lfloor t\rfloor} R^\eta_\ell(\lfloor t\rfloor, \lfloor s\rfloor)\theta^{\lfloor s\rfloor}_\eta \mathrm{d}s + u^t_\eta\,, \qquad u^t_\eta \sim \mathsf{GP}(0, C^\eta_\ell/\delta)\,, \tag{113a}$$

$$r_\eta^t = -\frac{1}{\delta}\int_0^{\lfloor t\rfloor} R_\theta^\eta(\lfloor t\rfloor, \lceil s\rceil)\ell_{\lfloor s\rfloor}(r_\eta^s; z)\mathrm{d}s + w_\eta^t\,, \qquad w_\eta^t \sim \mathsf{GP}(0, C_\theta^\eta)\,, \tag{113b}$$

$$R_\theta^\eta(t,s) = \mathbb{E}\left[\frac{\partial \theta_\eta^t}{\partial u_\eta^s}\right]\,, \qquad 0\le s\le t<\infty\,, \tag{113c}$$

$$R_\ell^\eta(t,s) = \mathbb{E}\left[\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t; z)}{\partial w_\eta^s}\right]\,, \qquad 0\le s< t<\infty\,, \tag{113d}$$

$$\Gamma_\eta^t = \mathbb{E}\left[\nabla_r \ell_{\lfloor t\rfloor}(r_\eta^t; z)\right]\,, \tag{113e}$$

$$C_\theta^\eta(t,s) = \mathbb{E}\left[\theta_\eta^{\lfloor t\rfloor}{\theta_\eta^{\lfloor s\rfloor}}^{\mathsf{T}}\right]\,, \qquad 0\le s\le t<\infty\,, \tag{113f}$$

$$C_\ell^\eta(t,s) = \mathbb{E}\left[\ell_{\lfloor t\rfloor}(r_\eta^{\lfloor t\rfloor}; z)\ell_{\lfloor s\rfloor}(r_\eta^{\lfloor s\rfloor}; z)^{\mathsf{T}}\right]\,, \qquad 0\le s\le t<\infty\,, \tag{113g}$$

where the functional derivatives are determined by

$$\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial \theta_\eta^t}{\partial u_\eta^s} = -(\Lambda^{\lfloor t\rfloor} + \Gamma_\eta^{\lfloor t\rfloor})\frac{\partial \theta_\eta^{\lfloor t\rfloor}}{\partial u_\eta^s} - \int_s^{\lfloor t\rfloor} R_\ell^\eta(\lfloor t\rfloor, \lfloor s'\rfloor)\frac{\partial \theta_\eta^{\lfloor s'\rfloor}}{\partial u_\eta^s}\mathrm{d}s'\,, \tag{114a}$$

$$\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t; z)}{\partial w_\eta^s} = \nabla_r \ell_{\lfloor t\rfloor}(r_\eta^t; z)\cdot\left(-\frac{1}{\delta}\int_{\lceil s\rceil}^{\lfloor t\rfloor} R_\theta^\eta(\lfloor t\rfloor, \lceil s'\rceil)\frac{\partial \ell_{\lfloor s'\rfloor}(r_\eta^{s'}; z)}{\partial w_\eta^s}\mathrm{d}s' - \frac{1}{\delta}R_\theta^\eta(\lfloor t\rfloor, \lceil s\rceil)\nabla_r \ell_{\lfloor s\rfloor}(r_\eta^s; z)\right)\,, \tag{114b}$$

where Eq. (114b) is defined for $\lceil s\rceil \le \lfloor t\rfloor$. For $\lceil s\rceil > \lfloor t\rfloor$ we set

$$\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t; z)}{\partial w_\eta^s} = -\frac{1}{\delta}\nabla_r \ell_{\lfloor t\rfloor}(r_\eta^t; z)\nabla_r \ell_{\lfloor t\rfloor}(r_\eta^s; z)\,. \tag{115}$$

The boundary conditions are $\theta_\eta^0 := \theta^0$ and $\partial\theta_\eta^s/\partial u_\eta^s = I$. The system $\mathfrak{S}^\eta$ can be viewed as a discrete approximation of $\mathfrak{S}$. The next lemma shows that the unique solution of $\mathfrak{S}^\eta$ characterizes the asymptotic behavior of $\mathfrak{F}^\eta$.

Lemma 6.2. *Under the assumptions of Theorem 2, suppose $\Phi_{C_\theta}(0) > M_{\theta^0,z}$ and $\Phi_{R_\theta}(0) > 1$, the system $\mathfrak{S}^\eta$ has a unique solution. In particular, the function triplet $(C_\ell^\eta, R_\ell^\eta, \Gamma_\eta)$ that solves $\mathfrak{S}^\eta$ is in the space $\mathcal{S}$ (cf. Definition 5.2). For any $\tau_1,\cdots,\tau_m\in[0,T]$, denote by $\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}$ the joint distribution of $(\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m})$, we have*

$$\operatorname*{p-lim}_{n\to\infty} d_{\mathrm{W}}\left(\widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}, \mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}\right) = 0\,, \tag{116}$$

where $\widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}$ is the empirical distribution of the discretized flow $\boldsymbol{\theta}_\eta^t$, defined in Eq. (108).

We notice that the solution of $\mathfrak{S}^\eta$ is uniquely by induction over time. We postpone the proof to Appendix C.2.

Remark 6.1. In case we are interested in discrete-time flows, e.g. gradient descent with stepsize η, Lemma 6.2 provides the relevant characterization.

Remark 6.2. In the proof of Lemma 6.2 we use the results of [CL21] which establishes universality over the class of matrices satisfying the assumptions of Theorem 2.

If $\boldsymbol{X}$ is a Gaussian matrix, we can use the results of [JM13] which imply convergence in Wasserstein-2 distance. Hence in this case, Theorem 2 will hold with convergence of finite-dimensional distributions in Wasserstein-2 distance, as stated in Remark 3.2.

Part III: Approximating $\mathfrak{S}$ by $\mathfrak{S}^\eta$. We approximate the unique solution of the DMFT system $\mathfrak{S}$ by the unique solution of the discretized system $\mathfrak{S}^\eta$. In particular, we have the following lemma, whose proof is postponed to Appendix C.3.

Lemma 6.3. *Under the assumptions of Theorem 2, suppose $\Phi_{C_\theta}(0) > M_{\theta^0,z}$ and $\Phi_{R_\theta}(0) > 1$, the systems $\mathfrak{S}$ and $\mathfrak{S}^\eta$ both have unique solutions in $\mathcal{S}$. For any $\tau_1, \cdots, \tau_m \in [0,T]$, denote by $\mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}$ the distribution of $(\theta^{\tau_1}, \cdots, \theta^{\tau_m})$ and $\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}$ the distribution of $(\theta_\eta^{\tau_1}, \cdots, \theta_\eta^{\tau_m})$, we have*

$$\lim_{\eta\to\infty} \mathrm{W}_2 \left(\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right) = 0\,. \tag{117}$$

In particular, since convergence in Wasserstein-2 distance implies weak convergence, we also have

$$\lim_{\eta\to\infty} d_{\mathrm{W}} \left(\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right) = 0\,. \tag{118}$$

Using Lemma 6.2,

$$\begin{aligned}
&\operatorname*{p-lim\,sup}_{n\to\infty} d_{\mathrm{W}} \left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right) \\
&\le \operatorname*{p-lim\,sup}_{n\to\infty} \left(d_{\mathrm{W}} \left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}\right) + d_{\mathrm{W}} \left(\widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}, \mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}\right) + d_{\mathrm{W}} \left(\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right)\right) \\
&= \operatorname*{p-lim\,sup}_{n\to\infty} d_{\mathrm{W}} \left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}\right) + d_{\mathrm{W}} \left(\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right)\,.
\end{aligned} \tag{119}$$

Finally, take $\eta \to 0$ and combine Lemma 6.1 and Lemma 6.3, we obtain

$$\begin{aligned}
&\operatorname*{p-lim\,sup}_{n\to\infty} d_{\mathrm{W}} \left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right) \\
&\le \lim_{\eta\to 0} d_{\mathrm{W}} \left(\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right) + \lim_{\eta\to 0} \operatorname*{p-lim\,sup}_{n\to\infty} d_{\mathrm{W}} \left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}\right) \\
&= 0.
\end{aligned} \tag{120}$$

Now, let μ be the law of the DMFT process $(\theta^t)_0^T$, that is a probability law on $C([0,T], \mathbb{R}^k)$. Indeed, by condition (71) and Kolmogorov-Chentsov theorem, $t \mapsto \theta^t$ is almost surely α-Hölder for any $\alpha \in (0,1)$. Also, let $\widehat{\mu}^{(n)} := d^{-1} \sum_{i=1}^d \delta_{(\boldsymbol{\theta}_i)_0^T}$. Denoting by $\mu_{\tau_1,\ldots,\tau_m}$ and $\widehat{\mu}^{(n)}_{\tau_1,\ldots,\tau_n}$, we proved that

$$\operatorname*{p-lim}_{n\to\infty} d_{\mathrm{W}} \left(\widehat{\mu}^{(n)}_{\tau_1,\ldots,\tau_m}, \mu_{\tau_1,\ldots,\tau_m}\right) = 0\,. \tag{121}$$

We are left with the task of proving $d_{\mathrm{W}} \left(\widehat{\mu}^{(n)}, \mu\right) \to 0$ in probability. Recall the following basic fact.

Lemma 6.4. *For a sequence of random variable $(X_n)_{n\ge 1}$, we have $X_n \overset{p}{\to} 0$ if and only if for each diverging subsequence (n_ℓ) there exists a refinement $(n'_\ell) \subseteq (n_\ell)$, such that $X_{n'_\ell} \overset{a.s.}{\to} 0$.*

Let (n_ℓ) be a diverging sequence. Then for any m, and any $\tau_1, \ldots, \tau_m \in [0,T] \cap \mathbb{Q}$, we can construct a subsequence along which $d_{\mathrm{W}} \left(\widehat{\mu}^{(n'_\ell)}_{\tau_1,\ldots,\tau_n}, \mu_{\tau_1,\ldots,\tau_m}\right) \overset{a.s.}{\to} 0$. By successive refinements and a diagonal argument, we can assume that the subsequence is such that

$$\mathbb{P}\Big(\widehat{\mu}^{(n'_\ell)}_{\tau_1,\ldots,\tau_n} \Rightarrow \mu_{\tau_1,\ldots,\tau_m} \quad \forall m,\ \ \forall \tau_1, \ldots, \tau_m \in [0,T] \cap \mathbb{Q}\Big) = 1\,. \tag{122}$$

We finally need a tightness result, whose proof is presented in Appendix C.9.

Lemma 6.5. *Under the assumptions of Theorem 2, there exists $\alpha \in (0,1)$ and, for any $\varepsilon > 0$ there exists $M(\varepsilon) < \infty$ such that*

$$\mathrm{P}\Big(\widehat{\mu}^{(n)}\big(\{\|\theta^0\|_2 > M(\varepsilon)\} \cup \{\|(\theta)_0^T\|_{C^{0,\alpha}} > M(\varepsilon)\}\big) \ge \varepsilon \textit{ for infinitely many } n\Big) = 0\,. \tag{123}$$

(Here $\|f\|_{C^{0,\alpha}}$ denotes the α-Hölder seminorm of function f.)

By Eq. (122) and since finite-dimensional distributions on the rationals uniquely identify the limit on $C([0,T], \mathbb{R}^k)$ [Bil13], we proved $d_{\mathrm{W}} \left(\widehat{\mu}^{(n'_\ell)}, \mu\right) \overset{a.s.}{\to} 0$, and therefore using Lemma 6.4 we obtain the desired claim.

Acknowledgements

This work was supported by NSF through award DMS-2031883 and from the Simons Foundation through Award 814639 for the Collaboration on the Theoretical Foundations of Deep Learning. C. Cheng is supported by the William R. Hewlett Stanford graduate fellowship. M. Celentano is supported by the Miller Institute for Basic Research in Science, University of California Berkeley. We also acknowledge NSF grant CCF-2006489 and the ONR grant N00014-18-1-2729.

References

[ABC20] Ada Altieri, Giulio Biroli, and Chiara Cammarota, *Dynamical mean-field theory and aging dynamics*, Journal of Physics A: Mathematical and Theoretical **53** (2020), no. 37, 375006.

[ABUZ18] Elisabeth Agoritsas, Giulio Biroli, Pierfrancesco Urbani, and Francesco Zamponi, *Out-of-equilibrium dynamical mean-field equations for the perceptron model*, Journal of Physics A: Mathematical and Theoretical **51** (2018), no. 8, 085002.

[ADG01] Gerard Ben Arous, Amir Dembo, and Alice Guionnet, *Aging of spherical spin glasses*, Probability Theory and Related Fields **120** (2001), no. 1, 1–67.

[ADG06] ______, *Cugliandolo-Kurchan equations for dynamics of spin-glasses*, Probability Theory and Related Fields **136** (2006), no. 4, 619–660.

[AG95] Gerard Ben Arous and Alice Guionnet, *Large deviations for langevin spin glass dynamics*, Probability Theory and Related Fields **102** (1995), no. 4, 455–509.

[AG97] ______, *Symmetric langevin spin glass dynamics*, The Annals of Probability **25** (1997), no. 3, 1367–1422.

[Bil13] Patrick Billingsley, *Convergence of probability measures*, John Wiley & Sons, 2013.

[BLM15] Mohsen Bayati, Marc Lelarge, and Andrea Montanari, *Universality in polytope phase transitions and message passing algorithms*, The Annals of Applied Probability **25** (2015), no. 2, 753–822.

[BM11] Mohsen Bayati and Andrea Montanari, *The dynamics of message passing on dense graphs, with applications to compressed sensing*, IEEE Transactions on Information Theory **57** (2011), no. 2, 764–785.

[BMN20] Raphael Berthier, Andrea Montanari, and Phan-Minh Nguyen, *State evolution for approximate message passing with non-separable functions*, Information and Inference: A Journal of the IMA **9** (2020), no. 1, 33–79.

[Bol14] Erwin Bolthausen, *An iterative construction of solutions of the TAP equations for the Sherrington–Kirkpatrick model*, Communications in Mathematical Physics **325** (2014), no. 1, 333–366.

[BS10] Zhidong Bai and Jack Silverstein, *Spectral Analysis of Large Dimensional Random Matrices (2^{nd} edition)*, Springer, 2010.

[CC17] Yuxin Chen and Emmanuel J Candès, *Solving random quadratic systems of equations is nearly as easy as solving linear systems*, Communications on pure and applied mathematics **70** (2017), no. 5, 822–883.

[CHS93] Andrea Crisanti, Heinz Horner, and H-J Sommers, *The spherical p-spin interaction spin-glass model*, Zeitschrift für Physik B Condensed Matter **92** (1993), no. 2, 257–271.

[CK93] Leticia F Cugliandolo and Jorge Kurchan, *Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model*, Physical Review Letters **71** (1993), no. 1, 173.

[CK08] ———, *The out-of-equilibrium dynamics of the Sherrington–Kirkpatrick model*, Journal of Physics A: Mathematical and Theoretical **41** (2008), no. 32, 324018.

[CL21] Wei-Kuo Chen and Wai-Kit Lam, *Universality of approximate message passing algorithms*, Electronic Journal of Probability **26** (2021), 1–44.

[CMW20a] Michael Celentano, Andrea Montanari, and Yuting Wei, *The lasso with general Gaussian designs with applications to hypothesis testing*, arXiv:2007.13716 (2020).

[CMW20b] Michael Celentano, Andrea Montanari, and Yuchen Wu, *The estimation error of general first order methods*, Conference on Learning Theory, PMLR, 2020, pp. 1078–1141.

[DG21] Amir Dembo and Reza Gheissari, *Diffusions interacting through a random matrix: universality via stochastic Taylor expansion*, Probability Theory and Related Fields (2021), 1–41.

[DLZ19] Amir Dembo, Eyal Lubetzky, and Ofer Zeitouni, *Universality for langevin-like spin glass dynamics*, arXiv:1911.08001 (2019).

[DM16] David Donoho and Andrea Montanari, *High dimensional robust M-estimation: asymptotic variance via approximate message passing*, Probability Theory and Related Fields **166** (2016), no. 3, 935–969.

[Fie82] James R Fienup, *Phase retrieval algorithms: a comparison*, Applied optics **21** (1982), no. 15, 2758–2769.

[Gru96] Malte Grunwald, *Sanov results for glauber spin-glass dynamics*, Probability Theory and Related Fields **106** (1996), no. 2, 187–232.

[Gui97] Alice Guionnet, *Averaged and quenched propagation of chaos for spin glass dynamics*, Probability Theory and Related Fields **109** (1997), no. 2, 183–215.

[JM13] Adel Javanmard and Andrea Montanari, *State evolution for general approximate message passing algorithms, with applications to spatial coupling*, Information and Inference: A Journal of the IMA **2** (2013), no. 2, 115–144.

[Kar13] Noureddine El Karoui, *Asymptotic behavior of unregularized and ridge-regularized high-dimensional robust regression estimators: rigorous results.*

[MBC+20] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová, *Marvels and pitfalls of the Langevin algorithm in noisy high-dimensional inference*, Physical Review X **10** (2020), no. 1, 011057.

[MKUZ20] Francesca Mignacco, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová, *Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification*, arXiv:2006.06098 (2020).

[Mon21] Andrea Montanari, *Optimization of the Sherrington–Kirkpatrick hamiltonian*, SIAM Journal on Computing (2021), no. 0, FOCS19–1.

[MRSY19] Andrea Montanari, Feng Ruan, Youngtak Sohn, and Jun Yan, *The generalization error of max-margin linear classifiers: High-dimensional asymptotics in the overparametrized regime*, arXiv:1911.01544 (2019).

[RSZ88] H Rieger, Michael Schreckenberg, and J Zittartz, *Glauber dynamics of the Little-Hopfield model*, Zeitschrift für Physik B Condensed Matter **72** (1988), no. 4, 523–533.

[SC19] Pragya Sur and Emmanuel J. Candès, *A modern maximum-likelihood theory for high-dimensional logistic regression*, Proceedings of the National Academy of Sciences **116** (2019), no. 29, 14516–14525.

[SGB94] K Skouras, C Goutis, and MJ Bramson, *Estimation in linear models using gradient descent with early stopping*, Statistics and Computing **4** (1994), no. 4, 271–278.

[SHN$^+$18] Daniel Soudry, Elad Hoffer, Mor Shpigel Nacson, Suriya Gunasekar, and Nathan Srebro, *The implicit bias of gradient descent on separable data*, The Journal of Machine Learning Research **19** (2018), no. 1, 2822–2878.

[Sto13] Mihailo Stojnic, *A framework to characterize performance of lasso algorithms*, arXiv:1303.7291 (2013).

[SV97] Daniel W Stroock and SR Srinivasa Varadhan, *Multidimensional diffusion processes*, vol. 233, Springer Science & Business Media, 1997.

[SZ81] Haim Sompolinsky and Annette Zippelius, *Dynamic theory of the spin-glass phase*, Physical Review Letters **47** (1981), no. 5, 359.

[SZ82] ______, *Relaxational dynamics of the edwards-anderson model and the mean-field theory of spin-glasses*, Physical Review B **25** (1982), no. 11, 6860.

[TAH18] Christos Thrampoulidis, Ehsan Abbasi, and Babak Hassibi, *Precise error analysis of regularized M-estimators in high dimensions*, IEEE Transactions on Information Theory **64** (2018), no. 8, 5592–5628.

[TOH15] Christos Thrampoulidis, Samet Oymak, and Babak Hassibi, *Regularized linear regression: A precise analysis of the estimation error*, Proceedings of The 28th Conference on Learning Theory (Paris, France) (Peter Grünwald, Elad Hazan, and Satyen Kale, eds.), Proceedings of Machine Learning Research, vol. 40, PMLR, 03–06 Jul 2015, pp. 1683–1709.

A Summary of notations

We use $\|u\|_2$ to denote the ℓ_2 norm of a vector u. We also use $\|M\|$ and $\|M\|_F$ to denote the operator norm and Frobenius norm of a matrix M. For two vectors u, v of the same dimension, we write $u \geq v$ or $u \leq v$ to represent entrywise inequality. For random variables ξ and $\xi_1, \xi_2, \cdots, \xi_n, \cdots$ defined on the same probability space, we denote convergence almost surely, in probability and weakly by $\xi_n \overset{a.s.}{\to} \xi$, $\xi_n \overset{p}{\to} \xi$ and $\xi_n \Rightarrow \xi$, respectively. We will also use convergence in Wasserstein-2 distance, denoted by $\mu_n \overset{\mathrm{W}_2}{\to} \mu$. In particular, for two distributions μ, ν on $\mathbb{R}^k$, the Wasserstein-2 distance of is defined by

$$\mathrm{W}_2\left(\mu, \nu\right) := \inf_{\gamma \in \Gamma(\mu,\nu)} \sqrt{\int \|\xi - \eta\|_2^2 \, \mathrm{d}\gamma(\xi, \eta)}\,,$$

where $\Gamma(\mu, \nu)$ denotes the collection of all couplings of μ and ν.

Throughout the paper we adhere to the convention that a symbol is boldfaced if and only if it is a high dimensional object, namely whose dimension depends on n or d. For instance $\boldsymbol{X} \in \mathbb{R}^{n\times d}, \boldsymbol{\theta} \in \mathbb{R}^{d\times k}$ and $\theta \in \mathbb{R}^k$.

B Auxiliary lemmas for the proof of Theorem 1

B.1 Proofs for the auxiliary real-valued system

B.1.1 Proof of Lemma 5.1

Existence of $f_1(t)$ and $f_2(t)$. We first look at Eqs. (56a) and (56b) and prove the existence of f_1 and f_2. Recall the equations

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t} f_1(t) &= \alpha_1 f_1(t) + \alpha_2 f_2(t)\,, \\ f_2(t) &= \alpha_3 f_1(t) + \alpha_4 \int_0^t f_1(t-s) f_2(s) \mathrm{d}s\,,\end{aligned}$$

and we consider the measurable function pair space

$$\mathcal{S}_{\overline{\mathfrak{S}},1}(\lambda, \varepsilon_1, \varepsilon_2) := \Big\{(f_1, f_2) \mid f_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, \|\!|f_i|\!\|_{\lambda,\infty} \leq \varepsilon_i, i = 1,2; f_1(0) = \beta_1\Big\} \tag{124}$$

with the following metric

$$\mathsf{dist}_\lambda((f_1, f_2), (g_1, g_2)) := 4\alpha_3 \|\!|f_1 - g_1|\!\|_{\lambda,\infty} + \|\!|f_2 - g_2|\!\|_{\lambda,\infty}. \tag{125}$$

Let μ_λ be the measure on $[0, \infty)$ with Radon-Nikodym derivative $e^{-\lambda x}$ with respect to the Lebesgue measure on $[0, \infty)$. By the completeness of the space $L^1(\mu_\lambda)$, we can see that $\mathcal{S}_{\overline{\mathfrak{S}},1}(\lambda, \varepsilon_1, \varepsilon_2)$ is complete under the metric dist_λ. Next, we consider the mapping $\mathcal{T}_{\overline{\mathfrak{S}},1}(f_1, f_2) := (\overline{f}_1, \overline{f}_2)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t} \overline{f}_1(t) = \alpha_1 f_1(t) + \alpha_2 f_2(t)\,, \tag{126}$$

$$\overline{f}_2(t) = \alpha_3 f_1(t) + \alpha_4 \int_0^t f_1(t-s) f_2(s) \mathrm{d}s\,, \tag{127}$$

with $\overline{f}_1(0) = f_1(0)$. We have the following lemma.

Lemma B.1. *There exists constants $\varepsilon_i := \varepsilon_i(\alpha_1, \cdots, \alpha_4, \beta_1) > 0$ for $i = 1, 2$ such that for any $\lambda \geq \overline{\lambda} := \overline{\lambda}(\alpha_1, \cdots, \alpha_4, \beta_1)$, $\mathcal{T}_{\overline{\mathfrak{S}},1}$ is an operator that maps $\mathcal{S}_{\overline{\mathfrak{S}},1}(\lambda, \varepsilon_1, \varepsilon_2)$ into itself, and for any $(f_1, f_2), (g_1, g_2) \in \mathcal{S}_{\overline{\mathfrak{S}},1}$, the transformation $\mathcal{T}_{\overline{\mathfrak{S}},1}$ is a contraction*

$$\mathsf{dist}_\lambda\big(\mathcal{T}_{\overline{\mathfrak{S}},1}(f_1, f_2)\,, \mathcal{T}_{\overline{\mathfrak{S}},1}(g_1, g_2)\big) \leq \frac{1}{2} \mathsf{dist}_\lambda\big((f_1, f_2), (g_1, g_2)\big)\,. \tag{128}$$

We defer its proof to Appendix B.1.2. By completeness and the lemma above, we can conclude by Banach fixed-point theorem that there exists a unique $(f_1, f_2) \in \mathcal{S}_{\overline{\mathfrak{S}},1}(\lambda, \varepsilon_1, \varepsilon_2)$ such that

$$\mathcal{T}_{\overline{\mathfrak{S}},1}(f_1, f_2) = (f_1, f_2)\,, \tag{129}$$

i.e. solving Eqs. (56a) and (56b). By nonnegativity of f_1 and f_2, one can then see that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t} f_1(t) = \alpha_1 f_1(t) + \alpha_2 f_2(t) \geq 0\,,\\
&\frac{\mathrm{d}}{\mathrm{d}t} f_2(t) = \alpha_3 \frac{\mathrm{d}}{\mathrm{d}t} f_1(t) + \alpha_4 \beta_1 f_2(t) + \alpha_4 \int_0^t \frac{\mathrm{d}}{\mathrm{d}t} f_1(t-s) f_2(s) \mathrm{d}s \geq 0\,,
\end{aligned}$$

implying $f_1(t)$ and $f_2(t)$ are nondecreasing. Then, we can upper bound their the other norm

$$\|f_1\|_{\lambda,\infty} \leq \sup_{0\leq t<\infty} e^{-\lambda t} f_1(t) \leq \sup_{0\leq t<\infty} e \cdot \int_t^{t+1} e^{-\lambda s} f_1(s) \mathrm{d}s \leq e \|\!| f_1 |\!\|_{\lambda,\infty} \leq e\epsilon_1\,, \tag{130}$$

and $\|f_2\|_{\lambda,\infty} \leq e\epsilon_2$.

Existence of $f_3(t)$ and $f_4(t)$. Take $f_1(t)$ and $f_2(t)$ that satisfy the first two equations (56a) and (56b). We then seek measurable functions $f_3, f_4 : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}$ solving Eqs. (56c) and (56d), i.e.

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{f_3(t)} &= \sqrt{\alpha_5 f_3(t) + \alpha_6 f_4(t) + \alpha_7 \int_0^t (t-s+1)^2 f_2(t-s)^2 f_3(s) \mathrm{d}s}\,,\\
f_4(t) &= \alpha_8 + \alpha_9 f_3(t) + \alpha_{10} \int_0^t (t-s+1)^2 f_1(t-s)^2 f_4(s) \mathrm{d}s\,.
\end{aligned}$$

Consider the space

$$\mathcal{S}_{\overline{\mathfrak{S}},2}(\lambda, \varepsilon_3, \varepsilon_4) := \left\{ (f_3, f_4) \mid f_i : \mathbb{R}_{\geq 0} \to \mathbb{R}_{\geq 0}, \left\| \sqrt{f_i} \right\|_{\lambda,\infty} \leq \varepsilon_i, i = 3,4; f_3(0) = \beta_2 \right\} \tag{131}$$

with the metric

$$\mathsf{dist}_\lambda((f_3, f_4), (g_3, g_4)) := 4\sqrt{\alpha_9} \left\| \sqrt{f_3} - \sqrt{g_3} \right\|_{\lambda,\infty} + \left\| \sqrt{f_4} - \sqrt{g_4} \right\|_{\lambda,\infty}. \tag{132}$$

The space $\mathcal{S}_{\overline{\mathfrak{S}},2}(\lambda, \varepsilon_3, \varepsilon_4)$ is complete under the metric dist_λ. Then we consider the transformation $\mathcal{T}_{\overline{\mathfrak{S}},2}(f_3, f_4) := (\overline{f}_3, \overline{f}_4)$ such that

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\overline{f}_3(t)} = \sqrt{\alpha_5 f_3(t) + \alpha_6 f_4(t) + \alpha_7 \int_0^t (t-s+1)^2 f_2(t-s)^2 f_3(s) \mathrm{d}s}\,, \tag{133}$$

$$\overline{f}_4(t) = \alpha_8 + \alpha_9 f_3(t) + \alpha_{10} \int_0^t (t-s+1)^2 f_1(t-s)^2 f_4(s) \mathrm{d}s\,, \tag{134}$$

with $\overline{f}_3(0) = f_3(0)$. Similarly, in the following lemma we show that for properly chosen $\varepsilon_3, \varepsilon_4$ and large enough λ, $\mathcal{T}_{\overline{\mathfrak{S}},2}$ is a contraction mapping. We postpone its proof to Appendix B.1.3.

Lemma B.2. *There exists constants $\varepsilon_i := \varepsilon_i(\alpha_1, \cdots, \alpha_8, \beta_1, \beta_2) > 0$ for $i = 3,4$ such that for any $\lambda \geq \overline{\lambda} := \overline{\lambda}(\alpha_1, \cdots, \alpha_8, \beta_1, \beta_2)$, $\mathcal{T}_{\overline{\mathfrak{S}},2}$ is an operator that maps $\mathcal{S}_{\overline{\mathfrak{S}},2}(\lambda, \varepsilon_3, \varepsilon_4)$ into itself, and for any $(f_3, f_4), (g_3, g_4) \in \mathcal{S}_{\overline{\mathfrak{S}},2}$, the transformation $\mathcal{T}_{\overline{\mathfrak{S}},2}$ is a contraction*

$$\mathsf{dist}_\lambda\big(\mathcal{T}_{\overline{\mathfrak{S}},2}(f_3, f_4)\,, \mathcal{T}_{\overline{\mathfrak{S}},2}(g_3, g_4)\big) \leq \frac{1}{2} \mathsf{dist}_\lambda\big((f_3, f_4), (g_3, g_4)\big)\,. \tag{135}$$

The proof of existence is then concluded by the applying Banach fixed-point theorem and Lemma B.2. We see uniqueness in $L^1_{\mathrm{loc}}(\mathbb{R}_{\geq 0})$ by taking $\lambda \to \infty$ as any functions $f_i|_{[0,T]}$ will eventually fall in $\mathcal{S}_{\overline{\mathfrak{S}},1}, \mathcal{S}_{\overline{\mathfrak{S}},2}$ by taking $\lambda \to \infty$. This shows uniqueness up to any finite time T and thus concludes the proof.

B.1.2 Proof of Lemma B.1

We first show that for some properly chosen $\varepsilon_1, \varepsilon_2$ and large enough λ, we have $\mathcal{T}_{\overline{\mathfrak{S}},1} : \mathcal{S}_{\overline{\mathfrak{S}},1}(\lambda, \varepsilon_1, \varepsilon_2) \to \mathcal{S}_{\overline{\mathfrak{S}},1}(\lambda, \varepsilon_1, \varepsilon_2)$. For any $\lambda > 0$, by definition in Eq. (126) we have

$$\overline{f}_1(t) = \beta_1 + \alpha_1 \int_0^t f_1(s)\mathrm{d}s + \alpha_2 \int_0^t f_2(s)\mathrm{d}s\,,$$

and thus by nonnegativity of f_1, f_2 we further have

$$\begin{aligned}
\left|\left|\left|\overline{f}_1\right|\right|\right|_{\lambda,\infty} &= \beta_1 \int_0^\infty e^{-\lambda t}\mathrm{d}t + \alpha_1 \int_0^\infty \mathrm{d}t \int_0^t e^{-\lambda(t-s)} \cdot e^{-\lambda s} f_1(s)\mathrm{d}s + \alpha_2 \int_0^\infty \mathrm{d}t \int_0^t e^{-\lambda(t-s)} \cdot e^{-\lambda s} f_2(s)\mathrm{d}s \\
&= \int_0^\infty e^{-\lambda t}\mathrm{d}t \cdot \Big(\beta_1 + \alpha_1 |||f_1|||_{\lambda,\infty} + \alpha_2 |||f_2|||_{\lambda,\infty}\Big) \\
&= \frac{1}{\lambda} \cdot \Big(\beta_1 + \alpha_1 |||f_1|||_{\lambda,\infty} + \alpha_2 |||f_2|||_{\lambda,\infty}\Big) \le \frac{1}{\lambda} \cdot (\beta_1 + \alpha_1\varepsilon_1 + \alpha_2\varepsilon_2)\,.
\end{aligned} \tag{136}$$

Similarly by Eq. (127) we have

$$\begin{aligned}
\left|\left|\left|\overline{f}_2\right|\right|\right|_{\lambda,\infty} &= \alpha_3 \int_0^\infty e^{-\lambda t} f_1(t)\mathrm{d}t + \alpha_4 \int_0^\infty \mathrm{d}t \int_0^t e^{-\lambda(t-s)} f_1(t-s) \cdot e^{-\lambda s} f_2(s)\mathrm{d}s \\
&= \alpha_3 |||f_1|||_{\lambda,\infty} + \alpha_4 |||f_1|||_{\lambda,\infty} |||f_2|||_{\lambda,\infty} \\
&\le \alpha_3\varepsilon_1 + \alpha_4\varepsilon_1\varepsilon_2\,.
\end{aligned} \tag{137}$$

Take

$$\begin{aligned}
&\varepsilon_1 \le \frac{\varepsilon_2}{\alpha_3 + \alpha_4}\,, \\
&\varepsilon_2 \le 1\,, \\
&\lambda \ge \alpha_1 + \frac{\alpha_2\varepsilon_2 + \beta_1}{\varepsilon_1}\,,
\end{aligned}$$

we immediately have

$$\begin{aligned}
\left|\left|\left|\overline{f}_1\right|\right|\right|_{\lambda,\infty} &\le \frac{1}{\lambda} \cdot (\beta_1 + \alpha_1\varepsilon_1 + \alpha_2\varepsilon_2) \le \varepsilon_1\,, \\
\left|\left|\left|\overline{f}_2\right|\right|\right|_{\lambda,\infty} &\le \alpha_3\varepsilon_1 + \alpha_4\varepsilon_1\varepsilon_2 \le (\alpha_3 + \alpha_4)\varepsilon_1 \le \varepsilon_2\,.
\end{aligned}$$

This shows $\mathcal{T}_{\overline{\mathfrak{S}},1}$ maps $\mathcal{S}_{\overline{\mathfrak{S}},1}(\lambda, \varepsilon_1, \varepsilon_2)$ into itself. Next we show the contraction. Suppose $\mathcal{T}_{\overline{\mathfrak{S}},1}(g_1, g_2) = (\overline{g}_1, \overline{g}_2)$, and from Eq. (126) it follows that

$$\begin{aligned}
\left|\overline{f}_1(t) - \overline{g}_1(t)\right| &= \left|\beta_1 + \alpha_1 \int_0^t f_1(s)\mathrm{d}s + \alpha_2 \int_0^t f_2(s)\mathrm{d}s - \beta_1 - \alpha_1 \int_0^t g_1(s)\mathrm{d}s - \alpha_2 \int_0^t g_2(s)\mathrm{d}s\right| \\
&\le \alpha_1 \int_0^t |f_1(s) - g_1(s)|\,\mathrm{d}s + \alpha_2 \int_0^t |f_2(s) - g_2(s)|\,\mathrm{d}s\,,
\end{aligned} \tag{138}$$

which further implies

$$\begin{aligned}
\left|\left|\left|\overline{f}_1 - \overline{g}_1\right|\right|\right|_{\lambda,\infty} &\le \alpha_1 \int_0^\infty \mathrm{d}t \int_0^t e^{-\lambda(t-s)} \cdot e^{-\lambda s} |f_1(s) - g_1(s)|\,\mathrm{d}s \\
&\qquad + \alpha_2 \int_0^\infty \mathrm{d}t \int_0^t e^{-\lambda(t-s)} \cdot e^{-\lambda s} |f_2(s) - g_2(s)|\,\mathrm{d}s \\
&\le \int_0^\infty e^{-\lambda}\mathrm{d}t \cdot \left(\alpha_1 \int_0^\infty e^{-\lambda s} |f_1(s) - g_1(s)|\,\mathrm{d}s + \alpha_2 \int_0^\infty e^{-\lambda s} |f_2(s) - g_2(s)|\,\mathrm{d}s\right) \\
&\le \frac{1}{\lambda} \cdot \Big(\alpha_1 |||f_1 - g_1|||_{\lambda,\infty} + \alpha_2 |||f_2 - g_2|||_{\lambda,\infty}\Big)\,.
\end{aligned} \tag{139}$$

Substituting into Eq. (127) gives us

$$\begin{aligned}
&\left|\overline{f}_2(t) - \overline{g}_2(t)\right| \\
&= \left|\alpha_3 f_1(t) + \alpha_4 \int_0^t f_1(t-s) f_2(s)\mathrm{d}s - \alpha_3 g_1(t) - \alpha_4 \int_0^t g_1(t-s) g_2(s)\mathrm{d}s\right| \\
&\le \alpha_3 \left|f_1(t) - g_1(t)\right| + \alpha_4 \left(\int_0^t \left|f_1(t-s) - g_1(t-s)\right| f_2(s)\mathrm{d}s + \int_0^t g_1(t-s)|f_2(s) - g_2(s)|\mathrm{d}s\right), \qquad (140)
\end{aligned}$$

and further

$$\begin{aligned}
\left|\left|\left|\overline{f}_2 - \overline{g}_2\right|\right|\right|_{\lambda,\infty} \le \alpha_3 \int_0^t e^{-\lambda s} \left|f_1(s) - g_1(s)\right| \mathrm{d}s + \alpha_4 \int_0^\infty \mathrm{d}t \int_0^t e^{-\lambda(t-s)} \left|f_1(t-s) - g_1(t-s)\right| \cdot e^{-\lambda s} f_2(s)\mathrm{d}s \\
+ \alpha_4 \int_0^\infty \mathrm{d}t \int_0^t e^{-\lambda(t-s)} g_1(t-s) \cdot e^{-\lambda s} |f_2(s) - g_2(s)|\mathrm{d}s \\
= \alpha_3 |||f_1 - g_1|||_{\lambda,\infty} + \alpha_4 |||f_1 - g_1|||_{\lambda,\infty} |||f_2|||_{\lambda,\infty} + \alpha_4 |||g_1|||_{\lambda,\infty} |||f_2 - g_2|||_{\lambda,\infty} \\
\le (\alpha_3 + \alpha_4 \varepsilon_2) \cdot |||f_1 - g_1|||_{\lambda,\infty} + \alpha_4 \varepsilon_1 \cdot |||f_2 - g_2|||_{\lambda,\infty}. \qquad (141)
\end{aligned}$$

Suppose

$$\begin{aligned}
\varepsilon_1 &\le \frac{1}{4\alpha_4}, \\
\varepsilon_2 &\le \frac{\alpha_3}{2\alpha_4}, \\
\lambda &\ge 8\alpha_1 + 8\alpha_2\alpha_3,
\end{aligned}$$

and taking together Eqs. (140) and (141) yields

$$\begin{aligned}
\mathsf{dist}_\lambda\left(\mathcal{T}_{\overline{\mathfrak{S}},1}(f_1, f_2), \mathcal{T}_{\overline{\mathfrak{S}},1}(g_1, g_2)\right) &= \mathsf{dist}_\lambda\left((\overline{f}_1, \overline{f}_2), (\overline{g}_1, \overline{g}_2)\right) \\
&= 4\alpha_3 \left|\left|\left|\overline{f}_1 - \overline{g}_1\right|\right|\right|_{\lambda,\infty} + \left|\left|\left|\overline{f}_2 - \overline{g}_2\right|\right|\right|_{\lambda,\infty} \\
&\le \left(\frac{4\alpha_1\alpha_3}{\lambda} + \alpha_3 + \alpha_4\varepsilon_2\right) \cdot |||f_1 - g_1|||_{\lambda,\infty} + \left(\frac{4\alpha_2\alpha_3}{\lambda} + \alpha_4\varepsilon_1\right) \cdot |||f_2 - g_2|||_{\lambda,\infty} \\
&\le 2\alpha_3 |||f_1 - g_1|||_{\lambda,\infty} + \frac{1}{2} |||f_2 - g_2|||_{\lambda,\infty} \\
&= \frac{1}{2} \mathsf{dist}_\lambda\left((f_1, f_2), (g_1, g_2)\right). \qquad (142)
\end{aligned}$$

The proof is completed.

B.1.3 Proof of Lemma B.2

In the first step, we show that for some properly chosen $\varepsilon_3, \varepsilon_4$ and large enough λ, $\mathcal{T}_{\overline{\mathfrak{S}},2}$ maps $\mathcal{S}_{\overline{\mathfrak{S}},2}(\lambda, \varepsilon_3, \varepsilon_4)$ into itself. We set $F_i(\lambda) := \int_0^\infty e^{-\lambda s}(s+1)^2 f_i(s)^2 \mathrm{d}s$ for $i = 1, 2$. For any $\lambda > 0$, by Eq. (133)

$$\sqrt{\overline{f}_3(t)} = \sqrt{\beta_2} + \int_0^t \mathrm{d}s \sqrt{\alpha_5 f_3(s) + \alpha_6 f_4(s) + \alpha_7 \int_0^s (s - s' + 1)^2 f_2(s - s')^2 f_3(s')\mathrm{d}s'}.$$

Since

$$\begin{aligned}
&e^{-\lambda t} \int_0^t \mathrm{d}s \sqrt{\alpha_5 f_3(s) + \alpha_6 f_4(s) + \alpha_7 \int_0^s (s - s' + 1)^2 f_2(s - s')^2 f_3(s')\mathrm{d}s'} \\
&= \int_0^t e^{-\lambda(t-s)} \cdot e^{-\lambda s} \sqrt{\alpha_5 f_3(s) + \alpha_6 f_4(s) + \alpha_7 \int_0^s (s - s' + 1)^2 f_2(s - s')^2 f_3(s')\mathrm{d}s'}\mathrm{d}s
\end{aligned}$$

$$\leq \left(\int_0^t e^{-\lambda(t-s)}\mathrm{d}s\right)\cdot \sup_{0\leq s\leq t} e^{-\lambda s}\sqrt{\alpha_5 f_3(s)+\alpha_6 f_4(s)+\alpha_7\int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'}$$

$$\leq \frac{1}{\lambda}\cdot \sup_{0\leq s\leq t} e^{-\lambda s}\sqrt{\alpha_5 f_3(s)+\alpha_6 f_4(s)+\alpha_7\int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'}\,, \tag{143}$$

By Lemma B.1 we have $F_i(\lambda)$ is well-defined for any λ large enough and $F_i(\lambda)\to 0$ as $\lambda\to\infty$. In addition by the elementary inequality that $\sqrt{a+b}\leq\sqrt{a}+\sqrt{b}$ for any $a,b>0$,

$$\begin{aligned}
&\sup_{0\leq s\leq t} e^{-\lambda s}\sqrt{\alpha_5 f_3(s)+\alpha_6 f_4(s)+\alpha_7\int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'}\\
&\leq \sup_{0\leq s\leq t} e^{-\lambda s}\cdot\left(\sqrt{\alpha_5 f_3(s)}+\sqrt{\alpha_6 f_4(s)}+\sqrt{\alpha_7\int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'}\right)\\
&\leq \sqrt{\alpha_5}\left\|\sqrt{f_3}\right\|_{\lambda,\infty}+\sqrt{\alpha_6}\left\|\sqrt{f_4}\right\|_{\lambda,\infty}\\
&\qquad+\sqrt{\alpha_7}\sup_{0\leq s\leq t}\sqrt{\int_0^s e^{-2\lambda(s-s')}(s-s'+1)^2 f_2(s-s')^2\cdot e^{-2\lambda s'}f_3(s')\mathrm{d}s'}\\
&\leq \sqrt{\alpha_5}\left\|\sqrt{f_3}\right\|_{\lambda,\infty}+\sqrt{\alpha_6}\left\|\sqrt{f_4}\right\|_{\lambda,\infty}+\sqrt{\alpha_7 F_2(2\lambda)}\left\|\sqrt{f_3}\right\|_{\lambda,\infty}\,.
\end{aligned} \tag{144}$$

Therefore

$$\begin{aligned}
\left\|\sqrt{\overline{f}_3}\right\|_{\lambda,\infty} &= \sup_{0\leq t<\infty} e^{-\lambda t}\left(\sqrt{\beta_2}+\int_0^t \mathrm{d}s\sqrt{\alpha_5 f_3(s)+\alpha_6 f_4(s)+\alpha_7\int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'}\right)\\
&\leq \sqrt{\beta_2}+\frac{1}{\lambda}\cdot\left(\sqrt{\alpha_5}\varepsilon_3+\sqrt{\alpha_6}\varepsilon_4+\sqrt{\alpha_7 F_2(2\lambda)}\varepsilon_3\right)\,.
\end{aligned} \tag{145}$$

Analogously through the same calculations, we can have from Eq. (134) that

$$\begin{aligned}
e^{-\lambda t}\sqrt{\overline{f}_4(t)} &= e^{-\lambda t}\sqrt{\alpha_8+\alpha_9 f_3(t)+\alpha_{10}\int_0^t (t-s+1)^2 f_1(t-s)^2\cdot f_4(s)\mathrm{d}s}\,,\\
&\leq \sqrt{\alpha_8}+\sqrt{\alpha_9}\left\|\sqrt{f_3}\right\|_{\lambda,\infty}+\sqrt{\alpha_{10}F_1(2\lambda)}\left\|\sqrt{f_4}\right\|_{\lambda,\infty}
\end{aligned}$$

and further

$$\left\|\sqrt{\overline{f}_4}\right\|_{\lambda,\infty}\leq \alpha_8+\alpha_9\varepsilon_3+\sqrt{\alpha_{10}F_1(2\lambda)}\varepsilon_4\,. \tag{146}$$

By taking

$$\begin{aligned}
\varepsilon_3 &\geq 2\sqrt{\beta_2}\,,\\
\varepsilon_4 &\geq 2\alpha_9\varepsilon_3+2\alpha_8\,,
\end{aligned}$$

and large enough λ such that

$$\begin{aligned}
\frac{1}{\lambda}\cdot\left(\sqrt{\alpha_5}\varepsilon_3+\sqrt{\alpha_6}\varepsilon_4+\sqrt{\alpha_7 F_2(2\lambda)}\varepsilon_3\right) &\leq \sqrt{\beta_2}\,,\\
\alpha_{10}F_1(\lambda) &\leq \frac{1}{4}\,,
\end{aligned}$$

we can then get from Eqs. (145) and (146)

$$\|\overline{f}_3\|_{\lambda,\infty}\leq\sqrt{\beta_2}+\sqrt{\beta_2}=2\sqrt{\beta_2}\leq\varepsilon_3\,,$$

$$\|\overline{f}_4\|_{\lambda,\infty} \leq \frac{\varepsilon_4}{2} + \frac{\varepsilon_4}{2} \leq \varepsilon_4 \,.$$

Next, we show $\mathcal{T}_{\overline{\mathfrak{S}},2}$ is a contraction mapping. Let $\mathcal{T}_{\overline{\mathfrak{S}},2}(g_3, g_4) = (\overline{g}_3, \overline{g}_4)$, we can get from Eq. (133)

$$\begin{aligned}
&\left|\sqrt{\overline{f}_3(t)} - \sqrt{\overline{g}_3(t)}\right| \\
&= \Bigg|\sqrt{\beta_2} + \int_0^t \mathrm{d}s \sqrt{\alpha_5 f_3(s) + \alpha_6 f_4(s) + \alpha_7 \int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'} \\
&\quad - \sqrt{\beta_2} - \int_0^t \mathrm{d}s \sqrt{\alpha_5 g_3(s) + \alpha_6 g_4(s) + \alpha_7 \int_0^s (s-s'+1)^2 f_2(s-s')^2 g_3(s')\mathrm{d}s'}\Bigg| \\
&\leq \int_0^t \Bigg|\sqrt{\alpha_5 f_3(s) + \alpha_6 f_4(s) + \alpha_7 \int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'} \\
&\quad - \sqrt{\alpha_5 g_3(s) + \alpha_6 g_4(s) + \alpha_7 \int_0^s (s-s'+1)^2 f_2(s-s')^2 g_3(s')\mathrm{d}s'}\Bigg|\mathrm{d}s \\
&\leq \int_0^t \Bigg(\sqrt{\alpha_5} \cdot \left|\sqrt{f_3(s)} - \sqrt{g_3(s)}\right| + \sqrt{\alpha_6} \cdot \left|\sqrt{f_4(s)} - \sqrt{g_4(s)}\right| \\
&\quad + \sqrt{\alpha_7} \cdot \left|\sqrt{\int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'} - \sqrt{\int_0^s (s-s'+1)^2 f_2(s-s')^2 g_3(s')\mathrm{d}s'}\right|\Bigg)\mathrm{d}s \,, \qquad (147)
\end{aligned}$$

where in the last inequality we use the elementary fact $|\sqrt{a+b} - \sqrt{c+d}| \leq \sqrt{|a-c|} + \sqrt{|b-d|}$ for any $a, b, c, d \in \mathbb{R}_{\geq 0}$. By Minkowski inequality we also have

$$\begin{aligned}
&\left|\sqrt{\int_0^s (s-s'+1)^2 f_2(s-s')^2 f_3(s')\mathrm{d}s'} - \sqrt{\int_0^s (s-s'+1)^2 f_2(s-s')^2 g_3(s')\mathrm{d}s'}\right| \\
&\leq \sqrt{\int_0^s (s-s'+1)^2 f_2(s-s')^2 \left(\sqrt{f_3(s')} - \sqrt{g_3(s')}\right)^2 \mathrm{d}s'} \,. \qquad (148)
\end{aligned}$$

Therefore, taken collectively we have

$$\begin{aligned}
&e^{-\lambda t}\left|\sqrt{\overline{f}_3(t)} - \sqrt{\overline{g}_3(t)}\right| \\
&\leq \sqrt{\alpha_5} \cdot \int_0^t e^{-\lambda(t-s)} \cdot e^{-\lambda s}\left|\sqrt{f_3(s)} - \sqrt{g_3(s)}\right|\mathrm{d}s \\
&\quad + \sqrt{\alpha_6} \cdot \int_0^t e^{-\lambda(t-s)} \cdot e^{-\lambda s}\left|\sqrt{f_4(s)} - \sqrt{g_4(s)}\right|\mathrm{d}s \\
&\quad + \sqrt{\alpha_7} \cdot \int_0^t e^{-\lambda(t-s)} \cdot \sqrt{\int_0^s e^{-2\lambda(s-s')}(s-s'+1)^2 f_2(s-s')^2 e^{-2\lambda s'}\left(\sqrt{f_3(s')} - \sqrt{g_3(s')}\right)^2 \mathrm{d}s'\mathrm{d}s} \\
&\leq \frac{1}{\lambda}\left(\sqrt{\alpha_5}\left\|\sqrt{f_3} - \sqrt{g_3}\right\|_{\lambda,t} + \sqrt{\alpha_6}\left\|\sqrt{f_4} - \sqrt{g_4}\right\|_{\lambda,t} + \sqrt{\alpha_7 F_2(2\lambda)}\left\|\sqrt{f_3} - \sqrt{g_3}\right\|_{\lambda,t}\right), \qquad (149)
\end{aligned}$$

which further implies by taking supremum over $t \in [0, \infty)$

$$\left\|\sqrt{\overline{f}_3} - \sqrt{\overline{g}_3}\right\|_{\lambda,\infty} \leq \frac{1}{\lambda}\left(\left(\sqrt{\alpha_5} + \sqrt{\alpha_7 F_2(2\lambda)}\right)\left\|\sqrt{f_3} - \sqrt{g_3}\right\|_{\lambda,\infty} + \sqrt{\alpha_6}\left\|\sqrt{f_4} - \sqrt{g_4}\right\|_{\lambda,\infty}\right). \qquad (150)$$

Next from Eq. (134) it follows similarly that

$$\left|\sqrt{\overline{f}_4(t)} - \sqrt{\overline{g}_4(t)}\right|$$

$$\le \sqrt{\alpha_9} \cdot \left| \sqrt{f_3(t)} - \sqrt{g_3(t)} \right| + \sqrt{\alpha_{10}} \cdot \sqrt{\int_0^t (t-s+1)^2 f_1(t-s)^2 \left(\sqrt{f_4(s)} - \sqrt{g_4(s)} \right)^2 \mathrm{d}s}\,, \tag{151}$$

and

$$\left\| \sqrt{f_4} - \sqrt{g_4} \right\|_{\lambda,\infty} \le \sqrt{\alpha_9} \left\| \sqrt{f_3} - \sqrt{g_3} \right\|_{\lambda,\infty} + \sqrt{\alpha_{10} F_1(2\lambda)} \left\| \sqrt{f_4} - \sqrt{g_4} \right\|_{\lambda,\infty} . \tag{152}$$

We take λ large enough such that

$$\begin{aligned} \frac{\sqrt{\alpha_5} + \sqrt{\alpha_7 F_2(2\lambda)}}{\lambda} &\le \frac{1}{4}\,, \\ \frac{4\sqrt{\alpha_6 \alpha_9}}{\lambda} &\le \frac{1}{4}\,, \\ \sqrt{\alpha_{10} F_1(2\lambda)} &\le \frac{1}{4}\,, \end{aligned}$$

and further with Eqs. (151) and (152),

$$\begin{aligned} \mathsf{dist}_\lambda\big(\mathcal{T}_{\overline{\mathfrak{S}},2}(f_3, f_4)\,, \mathcal{T}_{\overline{\mathfrak{S}},2}(g_3, g_4)\big) &= \mathsf{dist}_\lambda\big((\overline{f}_3, \overline{f}_4), (\overline{g}_3, \overline{g}_4)\big) \\ &= 4\sqrt{\alpha_9} \left\| \sqrt{\overline{f}_3} - \sqrt{\overline{g}_3} \right\|_{\lambda,\infty} + \left\| \sqrt{\overline{f}_4} - \sqrt{\overline{g}_4} \right\|_{\lambda,\infty} \\ &\le 2 \cdot \left(\sqrt{\alpha_9} \left\| \sqrt{f_3} - \sqrt{g_3} \right\|_{\lambda,\infty} + \frac{1}{4} \left\| \sqrt{f_4} - \sqrt{g_4} \right\|_{\lambda,\infty} \right) \\ &= \frac{1}{2} \mathsf{dist}_\lambda\big((f_3, f_4), (g_3, g_4)\big). \end{aligned} \tag{153}$$

B.2 Proof of Lemma 5.4

We first show θ^t is uniquely defined. Note that u^t has covariance kernel C_ℓ/δ, which implies for any $0 \le s \le t \le T$,

$$\begin{aligned} \mathbb{E}\left[\left\| u^t - u^s \right\|_2^2 \right] &= \mathbb{E}\left[\mathrm{Tr}\left((u^t - u^s)(u^t - u^s)^\intercal \right) \right] \\ &= \mathrm{Tr}\left(\mathbb{E}\left[(u^t - u^s)(u^t - u^s)^\intercal \right] \right) \\ &= \frac{1}{\delta} \cdot \mathrm{Tr}\left(C_\theta(t,t) - 2C_\theta(t,s) + C_\theta(s,s) \right) \\ &\le \frac{k}{\delta} \left\| C_\theta(t,t) - 2C_\theta(t,s) + C_\theta(s,s) \right\| . \end{aligned} \tag{154}$$

By the definition of $\overline{\mathcal{S}}$ and in particular Eq. (62), we can invoke Kolmogorov continuity theorem (cf. [SV97, Cor. 2.1.4]) and conclude that the sample path u^t is continuous. In fact, u^t will be locally α-Hölder continuous for any $\alpha \in (0,1)$. Since the sample path u^t is continuous almost surely, u^t will be integrable on $[0,T]$ with probability one. For any fixed sample path u^t, we show there's a unique path $\theta^t \in L^1([0,T])$ satisfying Eq. (12a), namely

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^t = -(\Lambda^t + \Gamma^t)\theta^t - \int_0^t R_\ell(t,s)\theta^s \mathrm{d}s + u^t\,. \tag{155}$$

This can be done by finding a contraction mapping and using Banach fixed-point theorem. Consider the mapping $\mathcal{T}_\theta$ such that $\mathcal{T}_\theta : \theta^t \mapsto \overline{\theta}^t$ such that $\overline{\theta}^0 = \theta^0$ and

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{\theta}^t = -(\Lambda^t + \Gamma^t)\theta^t - \int_0^t R_\ell(t,s)\theta^s \mathrm{d}s + u^t\,. \tag{156}$$

For any θ_1, θ_2 such that $\theta_1^0 = \theta_2^0$, we have

$$\left\|\overline{\theta}_1^t - \overline{\theta}_2^t\right\| \le \int_0^t (M_\Lambda + M_\ell)\left\|\theta_1^s - \theta_2^s\right\|\mathrm{d}s + \int_0^t \mathrm{d}s \int_0^s \left\|R_\ell(s,s')\right\|\left\|\theta_1^{s'} - \theta_2^{s'}\right\|\mathrm{d}s'\,, \tag{157}$$

and further

$$\begin{aligned}\left\|\overline{\theta}_1 - \overline{\theta}_2\right\|_{\lambda,T} \le \sup_{0\le t\le T} \Bigg\{ &\int_0^t e^{-\lambda(t-s)}(M_\Lambda + M_\ell)\cdot e^{-\lambda s}\left\|\theta_1^s - \theta_2^s\right\|\mathrm{d}s \\ &+ \int_0^t e^{-\lambda(t-s)}\mathrm{d}s \int_0^s e^{-\lambda(s-s')}\left\|R_\ell(s,s')\right\|\cdot e^{-\lambda s'}\left\|\theta_1^{s'} - \theta_2^{s'}\right\|\mathrm{d}s'\Bigg\} \\ \le \frac{1}{\lambda}\Bigg\{&(M_\Lambda + M_\ell)\left\|\theta_1 - \theta_2\right\|_{\lambda,T} + \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\ell}(t)\mathrm{d}t\right)\left\|\theta_1 - \theta_2\right\|_{\lambda,T}\Bigg\}\,. \end{aligned} \tag{158}$$

By Lemma 5.1 for Φ_{R_ℓ}, we see the integral $\int_0^\infty e^{-\lambda t}\Phi_{R_\ell}(t)\mathrm{d}t$ will be finite for sufficiently large λ, and thus we can take λ large enough to ensure $\mathcal{T}_\theta$ is a contraction mapping with respect to the metric $\left\|\overline{\theta}_1 - \overline{\theta}_2\right\|_{\lambda,T}$. The uniqueness in $L^\infty([0,T])$ is immediate, while the uniqueness in $L^1([0,T])$ follows from the fact that $\mathcal{T}_\theta$ maps any element in $L^1([0,T])$ into $L^\infty([0,T])$. We can apply the exact same argument to $\partial\theta^t/\partial u^s$ as Eq. (13a) takes the same form of Eq. (12a).

Next we show r^t is uniquely defined. Again by Eq. (71) and Kolmogorov continuity theorem, the sample path w^t is α-Hölder continuous for any $\alpha \in (0,1)$ with probability 1. Consider the mapping $\mathcal{T}_r$ such that $\mathcal{T}_r : r^t \mapsto \overline{r}^t$ and

$$\overline{r}^t = -\frac{1}{\delta}\int_0^t R_\theta(t,s)\ell_s(r^s;z)\mathrm{d}s + w^t\,. \tag{159}$$

Similarly, we get for any r_1, r_2 the mapping is a contraction

$$\begin{aligned}\left\|\overline{r}_1 - \overline{r}_2\right\|_{\lambda,T} &\le \frac{1}{\delta}\left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right)\sup_{0\le t\le T} e^{-\lambda t}\left\|\ell_t(r_1^t;z) - \ell_t(r_2^t;z)\right\|_2 \\ &\overset{\text{(i)}}{\le} \frac{M_\ell}{\delta}\left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right)\sup_{0\le t\le T} e^{-\lambda t}\left\|r_1^t - r_2^t\right\|_2 \\ &\le \frac{M_\ell}{\delta}\left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right)\left\|r_1 - r_2\right\|_{\lambda,T}\,, \end{aligned} \tag{160}$$

where in (i) we use the Lipschitz property of the function ℓ. Again by Lemma 5.1, by taking $\lambda \to \infty$, the integral $\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t$ can be arbitrarily small, implying $\mathcal{T}_r$ is a contraction mapping. The proof of uniqueness and existence for the functional derivative $\partial\ell_t(r^t;z)/\partial w^s$ is the same, provided that now the path r^t is uniquely defined.

B.3 Proof of Lemma 5.5

$\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}$ maps $\mathcal{S}$ into $\overline{\mathcal{S}}_{\mathrm{cont}}$. Directly from Eq. (80), we can get

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\theta^t\right\|_2 &\le \left(\left\|\Lambda^t\right\| + \left\|\Gamma^t\right\|\right)\left\|\theta^t\right\|_2 + \int_0^t \left\|R_\ell(t,s)\right\|\left\|\theta^s\right\|_2\mathrm{d}s + \left\|u^t\right\|_2 \\ &\le (M_\Lambda + M_\ell)\left\|\theta^t\right\|_2 + \int_0^t \Phi_{R_\ell}(t-s)\left\|\theta^s\right\|_2\mathrm{d}s + \left\|u^t\right\|_2\,, \end{aligned} \tag{161}$$

where the last line follows from the assumptions that $\|\Lambda^t\| \le M_\Lambda$, $\|\Gamma^t\| \le M_\ell$ and $\|R_\ell(t,s)\| \le \Phi_{R_\ell}(t-s)$, which further gives us

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\mathbb{E}\left\|\theta^t\right\|_2^2} = \frac{\mathbb{E}\left[\|\theta^t\|_2 \frac{\mathrm{d}}{\mathrm{d}t}\|\theta^t\|_2\right]}{\sqrt{\mathbb{E}\left\|\theta^t\right\|_2^2}} \le \sqrt{\mathbb{E}\left[\left(\frac{\mathrm{d}}{\mathrm{d}t}\left\|\theta^t\right\|_2\right)^2\right]}$$

$$\le \sqrt{\mathbb{E}\left[\left((M_\Lambda + M_\ell)\left\|\theta^t\right\|_2 + \int_0^t \Phi_{R_\ell}(t-s)\left\|\theta^s\right\|_2 \mathrm{d}s + \left\|u^t\right\|_2\right)^2\right]}$$

$$\overset{\text{(i)}}{\le} \sqrt{\mathbb{E}\left[\left((M_\Lambda + M_\ell)^2\left\|\theta^t\right\|_2^2 + \int_0^t (t-s+1)^2\Phi_{R_\ell}(t-s)^2\left\|\theta^s\right\|_2^2 \mathrm{d}s + \left\|u^t\right\|_2^2\right)\cdot\left(1+\int_0^t (t-s+1)^{-2}\mathrm{d}s + 1\right)\right]}$$

$$\le \sqrt{3\cdot\mathbb{E}\left[(M_\Lambda + M_\ell)^2\left\|\theta^t\right\|_2^2 + \int_0^t (t-s+1)^2\Phi_{R_\ell}(t-s)^2\left\|\theta^s\right\|_2^2 \mathrm{d}s + \left\|u^t\right\|_2^2\right]}$$

$$\le \sqrt{3\cdot\left\{(M_\Lambda + M_\ell)^2\,\mathbb{E}\left[\left\|\theta^t\right\|_2^2\right] + \int_0^t (t-s+1)^2\Phi_{R_\ell}(t-s)^2\mathbb{E}\left[\left\|\theta^s\right\|_2^2\right] \mathrm{d}s + \mathbb{E}\left[\left\|u^t\right\|_2^2\right]\right\}}$$

$$\le \sqrt{3\cdot\left\{(M_\Lambda + M_\ell)^2\,\mathbb{E}\left[\left\|\theta^t\right\|_2^2\right] + \int_0^t (t-s+1)^2\Phi_{R_\ell}(t-s)^2\mathbb{E}\left[\left\|\theta^s\right\|_2^2\right] \mathrm{d}s + \frac{k}{\delta}\Phi_{C_\ell}(t)\right\}}\,, \tag{162}$$

where in (i) we use Cauchy-Schwarz inequality and in the last line it is used that

$$\mathbb{E}\left[\left\|u^t\right\|_2^2\right] = \mathrm{Tr}\left(\mathbb{E}\left[u^t {u^t}^\mathsf{T}\right]\right) \le k\left\|\mathbb{E}\left[u^t {u^t}^\mathsf{T}\right]\right\| = \frac{k}{\delta}\left\|C_\ell(t,t)\right\| \le \frac{k}{\delta}\Phi_{C_\ell}(t)\,. \tag{163}$$

While since $\Phi_{C_\theta}(0) > M_{\theta^0,z} \ge \mathbb{E}\left[\left\|\theta^0\right\|_2^2\right]$ and recall Eq. (58c)

$$\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\Phi_{C_\theta}(t)} = \sqrt{3\cdot\left\{(M_\Lambda + M_\ell)^2\Phi_{C_\theta}(t) + \frac{k}{\delta}\Phi_{C_\ell}(t) + \int_0^t (t-s+1)^2\Phi_{R_\ell}(t-s)^2\Phi_{C_\theta}(s)\mathrm{d}s\right\}}\,,$$

we obtain that $\mathbb{E}\left[\left\|\theta^t\right\|_2^2\right] < \Phi_{C_\theta}(t)$ for all $t \in [0,T]$. We thus have

$$\left\|\overline{C}_\theta(t,t)\right\| = \left\|\mathbb{E}\left[\theta^t {\theta^t}^\mathsf{T}\right]\right\| \le \mathbb{E}\left[\left\|\theta^t\right\|_2^2\right] \le \Phi_{C_\theta}(t)\,. \tag{164}$$

Next we look at the definition for the formal partial derivative $\partial\theta^t/\partial u^s$, as it is not a random function we have from Eq. (81) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{R}_\theta(t,s) = -\left(\Lambda^t + \Gamma^t\right)\overline{R}_\theta(t,s) - \int_s^t R_\ell(t,s')\overline{R}_\theta(s',s)\mathrm{d}s'\,, \tag{165}$$

for $0 \le s \le t \le T$ and with $\overline{R}_\theta(s,s) = I$. Substituting in assumptions of $\mathcal{S}$ in Definition 5.2, it holds that

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left\|\overline{R}_\theta(t,s)\right\| &\le \left\|\frac{\mathrm{d}}{\mathrm{d}t}\overline{R}_\theta(t,s)\right\| \\
&\le \left(\left\|\Lambda^t\right\| + \left\|\Gamma^t\right\|\right)\left\|\overline{R}_\theta(t,s)\right\| + \int_s^t \left\|R_\ell(t,s')\right\|\left\|\overline{R}_\theta(s',s)\right\|\mathrm{d}s' \\
&\le (M_\Lambda + M_\ell)\left\|\overline{R}_\theta(t,s)\right\| + \int_s^t \Phi_{R_\ell}(t-s')\left\|\overline{R}_\theta(s',s)\right\|\mathrm{d}s'\,,
\end{aligned} \tag{166}$$

with $\left\|\overline{R}_\theta(s,s)\right\| = 1$. Since $\Phi_{R_\theta}(0) > 1$ and by Eq. (58a)

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{R_\theta}(t) = (M_\Lambda + M_\ell)\Phi_{R_\theta}(t) + \int_0^t \Phi_{R_\ell}(t-s)\Phi_{R_\theta}(s)\mathrm{d}s\,,$$

we can obtain that

$$\left\|\overline{R}_\theta(t,s)\right\| \le \Phi_{R_\theta}(t-s). \tag{167}$$

Finally, we note that for any $0 \le s \le t \le T$,

$$\left\|\overline{C}_\theta(t,t) - 2\overline{C}_\theta(t,s) + \overline{C}_\theta(s,s)\right\| = \left\|\mathbb{E}\left[(\theta^t - \theta^s)(\theta^t - \theta^s)^\mathsf{T}\right]\right\| \le \mathbb{E}\left[\left\|(\theta^t - \theta^s)(\theta^t - \theta^s)^\mathsf{T}\right\|\right] \le \mathbb{E}\left[\left\|\theta^t - \theta^s\right\|_2^2\right],$$

and thus further

$$\begin{aligned}
&\left\|\overline{C}_\theta(t,t) - 2\overline{C}_\theta(t,s) + \overline{C}_\theta(s,s)\right\| \\
&\le \mathbb{E}\left[\left\|\int_s^t \left\{-(\Lambda^{t'} + \Gamma^{t'})\theta^{t'} - \int_0^{t'} R_\ell(t',s')\theta^{s'}\mathrm{d}s' + u^{t'}\right\}\mathrm{d}t'\right\|_2^2\right] \\
&\le (t-s)^2 \sup_{0\le t\le T} \mathbb{E}\left[\left((M_\Lambda + M_\ell)\left\|\theta^t\right\|_2 + \int_0^t \Phi_{R_\ell}(t-s)\left\|\theta^s\right\|_2\mathrm{d}s + \left\|u^t\right\|_2\right)^2\right] \\
&\le (t-s)^2 \cdot \sup_{0\le t\le T} 3\mathbb{E}\left[(M_\Lambda + M_\ell)^2\left\|\theta^t\right\|_2^2 + \int_0^t (t-s+1)^2\Phi_{R_\ell}(t-s)^2\left\|\theta^s\right\|_2^2\mathrm{d}s + \left\|u^t\right\|_2^2\right] \\
&\le (t-s)^2 \cdot \sup_{0\le t\le T} \cdot 3\left\{(M_\Lambda + M_\ell)^2\Phi_{C_\theta}(t) + \int_0^t (t-s+1)^2\Phi_{R_\ell}(t-s)^2\Phi_{C_\theta}(s)\mathrm{d}s + \frac{k}{\delta}\Phi_{C_\ell}(t)\right\}. \qquad (168)
\end{aligned}$$

Since $\Phi_{R_\ell}, \Phi_{R_\theta}$ and Φ_{C_ℓ} are nondecreasing, we get

$$\begin{aligned}
\left\|\overline{C}_\theta(t,t) - 2\overline{C}_\theta(t,s) + \overline{C}_\theta(s,s)\right\| &\le 3\left\{\left[(M_\Lambda + M_\ell)^2 + T(T+1)^2\Phi_{R_\ell}(T)^2\right]\Phi_{C_\theta}(T) + \frac{k}{\delta}\Phi_{C_\ell}(T)\right\}(t-s)^2 \\
&= M_{\overline{\mathcal{S}}}(t-s)^2. \qquad (169)
\end{aligned}$$

Similarly, we can see the continuity of $\overline{C}_\theta$ by Cauchy-Schwarz (as an even stronger result, we show $\overline{C}_\theta$ is Lipschitz continuous)

$$\left\|\overline{C}_\theta(t,s) - \overline{C}_\theta(t,s')\right\| \le \sqrt{\mathbb{E}\left[\left\|\theta^t\right\|_2^2\right] \cdot \mathbb{E}\left[\left\|\theta^s - \theta^{s'}\right\|_2^2\right]} \le \sqrt{\Phi_{C_\theta}(T)M_{\overline{\mathcal{S}}}} \cdot |s - s'|. \qquad (170)$$

This shows that $(\overline{C}_\theta, \overline{R}_\theta) \in \overline{\mathcal{S}}_{\mathrm{cont}}$ and concludes the first part.

$\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}$ maps $\overline{\mathcal{S}}$ into $\mathcal{S}_{\mathrm{cont}}$. Next we will show $\left\|\overline{C}_\ell(t,t)\right\| \le \Phi_{C_\ell}(t)$ assuming that $(\overline{C}_\theta, \overline{R}_\theta) \in \overline{\mathcal{S}}$. By Definition 5.3 and Eq. (84), it follows that

$$\begin{aligned}
\left\|\ell_t\left(r^t;z\right)\right\|_2 &\le \left\|\ell_t(0;z)\right\|_2 + M_\ell\left\|r^t\right\|_2 \\
&\le \left\|\ell_t(0;z)\right\|_2 + \frac{M_\ell}{\delta}\int_0^t \left\|\overline{R}_\theta(t,s)\right\|\left\|\ell_s\left(r^s;z\right)\right\|_2\mathrm{d}s + M_\ell\left\|w^t\right\|_2 \\
&\le \left\|\ell_t(0;z)\right\|_2 + \frac{M_\ell}{\delta}\int_0^t \Phi_{R_\theta}(t-s)\left\|\ell_s\left(r^s;z\right)\right\|_2\mathrm{d}s + M_\ell\left\|w^t\right\|_2. \qquad (171)
\end{aligned}$$

Hence

$$\begin{aligned}
&\mathbb{E}\left[\left\|\ell_t\left(r^t;z\right)\right\|_2^2\right] \\
&\le \mathbb{E}\left[\left(\left\|\ell_t(0;z)\right\|_2 + \frac{M_\ell}{\delta}\int_0^t \Phi_{R_\theta}(t-s)\left\|\ell_s\left(r^s;z\right)\right\|_2\mathrm{d}s + M_\ell\left\|w^t\right\|_2\right)^2\right] \\
&\overset{\text{(i)}}{\le} \mathbb{E}\left[\left(\left\|\ell_t(0;z)\right\|_2^2 + \frac{M_\ell^2}{\delta^2}\int_0^t (t-s+1)^2\Phi_{R_\theta}(t-s)^2\left\|\ell_s\left(r^s;z\right)\right\|_2^2\mathrm{d}s + M_\ell^2\left\|w^t\right\|_2^2\right) \cdot \left(1 + \int_0^t (t-s+1)^{-2}\mathrm{d}s + 1\right)\right] \\
&\le 3\left\{\mathbb{E}\left[\left\|\ell_t(0;z)\right\|_2^2\right] + \frac{M_\ell^2}{\delta^2}\int_0^t (t-s+1)^2\Phi_{R_\theta}(t-s)^2\mathbb{E}\left[\left\|\ell_s\left(r^s;z\right)\right\|_2^2\right]\mathrm{d}s + M_\ell^2\mathbb{E}\left[\left\|w^t\right\|_2^2\right]\right\} \\
&\le 3\left\{M_{\theta^0,z} + \frac{M_\ell^2}{\delta^2}\int_0^t (t-s+1)^2\Phi_{R_\theta}(t-s)^2\mathbb{E}\left[\left\|\ell_s\left(r_s;z\right)\right\|_2^2\right]\mathrm{d}s + kM_\ell^2\Phi_{C_\theta}(t)\right\}, \qquad (172)
\end{aligned}$$

where in (i) we use Cauchy-Schwarz inequality and in the last line it is used that $\mathbb{E}\left[\|w^t\|_2^2\right] \le k\left\|\overline{C}_\theta(t,t)\right\| \le k\Phi_{C_\theta}(t)$. By Eq. (58d) we note that

$$\Phi_{C_\ell}(t) = 3\cdot\left\{M_{\theta^0,z} + kM_\ell^2\Phi_{C_\theta}(t) + \frac{M_\ell^2}{\delta^2}\int_0^t (t-s+1)^2\Phi_{R_\theta}(t-s)^2\Phi_{C_\ell}(s)\mathrm{d}s\right\},$$

and along with the fact that $\Phi_{C_\ell}(0) \ge 3M_{\theta^0,z} + 3kM_\ell^2\Phi_{C_\theta}(0) > M_{\theta^0,z} \ge \mathbb{E}\left[\left\|\ell_0\left(r^0;z\right)\right\|_2^2\right]$, it must follow that

$$\left\|\overline{C}_\ell(t,t)\right\| = \left\|\mathbb{E}\left[\ell_t(r^t;z)\ell_t(r^t;z)^\mathsf{T}\right]\right\| \le \mathbb{E}\left[\left\|\ell_t\left(r^t;z\right)\right\|_2^2\right] \le \Phi_{C_\ell}(t)\,. \tag{173}$$

Next we show $\overline{R}_\theta(t,s) \le \Phi_{R_\theta}(t-s)$ for all $0 \le s < t \le T$. By Eq. (85),

$$\frac{\partial \ell_t(r^t;z)}{\partial w^s} = \nabla_r\ell_t(r^t;z)\cdot\left(-\frac{1}{\delta}\int_s^t \overline{R}_\theta(t,s')\frac{\partial \ell_{s'}(r^{s'};z)}{\partial w^s}\mathrm{d}s' - \frac{1}{\delta}\overline{R}_\theta(t,s)\nabla_r\ell_s(r^s;z)\right),$$

and the Lipschitz property of the function ℓ, i.e. $\|\nabla_r\ell_t(r^t;z)\| \le M_\ell$, we have

$$\begin{aligned}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r^t;z)}{\partial w^s}\right\|\right] &\le \frac{M_\ell}{\delta}\left(\int_s^t \left\|\overline{R}_\theta(t,s')\right\|\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r^{s'};z)}{\partial w^s}\right\|\right]\mathrm{d}s' + M_\ell\left\|\overline{R}_\theta(t,s)\right\|\right)\\ &\le \frac{M_\ell}{\delta}\left(\int_s^t \Phi_{R_\theta}(t-s')\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r^{s'};z)}{\partial w^s}\right\|\right]\mathrm{d}s' + M_\ell\Phi_{R_\theta}(t-s)\right).\end{aligned} \tag{174}$$

Comparing to the deterministic system in Eq. (58b),

$$\Phi_{R_\ell}(t) = \frac{M_\ell}{\delta}\cdot\left\{M_\ell\Phi_{R_\theta}(t) + \int_0^t \Phi_{R_\theta}(t-s)\Phi_{R_\ell}(s)\mathrm{d}s\right\},$$

we see $\mathbb{E}\left[\left\|\frac{\partial \ell_t(r^t;z)}{\partial w^s}\right\|\right] \le \Phi_{R_\ell}(t-s)$ for all $0 \le s \le t \le T$, and further

$$\left\|\overline{R}_\ell(t,s)\right\| = \left\|\mathbb{E}\left[\frac{\partial \ell_t(r^t;z)}{\partial w^s}\right]\right\| \le \mathbb{E}\left[\left\|\frac{\partial \ell_t(r^t;z)}{\partial w^s}\right\|\right] \le \Phi_{R_\ell}(t-s). \tag{175}$$

Then we conclude from the Lipschitz property and Eq. (88) that

$$\left\|\overline{\Gamma}^t\right\| = \left\|\mathbb{E}\left[\nabla_r\ell_t(r^t;z)\right]\right\| \le \mathbb{E}\left[\left\|\nabla_r\ell_t(r^t;z)\right\|\right] \le M_\ell\,. \tag{176}$$

We note that for any $0 \le s \le t \le T$,

$$\begin{aligned}&\left\|\overline{C}_\ell(t,t) - 2\overline{C}_\ell(t,s) + \overline{C}_\ell(s,s)\right\|\\ &= \left\|\mathbb{E}\left[(\ell_t(r^t;z) - \ell_s(r^s;z))(\ell_t(r^t;z) - \ell_s(r^s;z))^\mathsf{T}\right]\right\|\\ &\le \mathbb{E}\left[\left\|\ell_t(r^t;z) - \ell_s(r^s;z)\right\|_2^2\right]\\ &\le M_\ell\cdot\mathbb{E}\left[\left(\left\|r^t - r^s\right\|_2 + |t-s|\right)^2\right]\\ &\le M_\ell\cdot\mathbb{E}\left[\left(\frac{1}{\delta}\int_s^t \Phi_{R_\theta}(t-s')\left\|\ell_{s'}(r^{s'};z)\right\|_2\mathrm{d}s' + \left\|w^t - w^s\right\|_2 + |t-s|\right)^2\right]\\ &\le M_\ell\cdot 3\mathbb{E}\left[\left(\frac{1}{\delta^2}\int_s^t (t-s'+1)^2\Phi_{R_\theta}(t-s')^2\left\|\ell_{s'}(r^{s'};z)\right\|_2^2\mathrm{d}s' + \left\|w^t - w^s\right\|_2^2 + (t-s)^2\right)\right],\end{aligned}$$

where in the last line we use the Cauchy-Schwarz inequality. Further we take ind

$$
\begin{aligned}
\mathbb{E}\left[\left\|\ell_{s'}(r^{s'};z)\right\|_2^2\right] &\leq \Phi_{C_\ell}(s'),\\
\mathbb{E}\left[\left\|w^t-w^s\right\|_2^2\right] &\leq k\left\|\overline{C}_\theta(t,t)-2\overline{C}_\theta(t,s)+\overline{C}_\theta(s,s)\right\|\\
&\leq 3k\left\{\left[(M_\Lambda+M_\ell)^2+T(T+1)^2\Phi_{R_\ell}(T)^2\right]\Phi_{C_\theta}(T)+\frac{k}{\delta}\Phi_{C_\ell}(T)\right\}(t-s)^2,
\end{aligned}
$$

it then follows that

$$
\begin{aligned}
&\left\|\overline{C}_\ell(t,t)-2\overline{C}_\ell(t,s)+\overline{C}_\ell(s,s)\right\|\\
&\leq 3M_\ell(t-s)^2\\
&\qquad\cdot\left\{\frac{1}{\delta^2}(T+1)^2\Phi_{R_\theta}(T)^2\Phi_{C_\ell}(T)+3k\left\{\left[(M_\Lambda+M_\ell)^2+T(T+1)^2\Phi_{R_\ell}(T)^2\right]\Phi_{C_\theta}(T)+\frac{k}{\delta}\Phi_{C_\ell}(T)\right\}+1\right\}\\
&=M_{\mathcal{S}}(t-s)^2.
\end{aligned}
$$

Similar to Eq. (74) in the previous part, we also have Lipschitz continuity for $\overline{C}_\ell$, namely $\forall s,s'\in[0,t]$,

$$
\left\|\overline{C}_\ell(t,s)-\overline{C}_\ell(t,s')\right\|\leq\sqrt{\mathbb{E}\left[\|\ell_t(r^t;z)\|_2^2\right]\cdot\mathbb{E}\left[\|\ell_s(r^s;z)-\ell_{s'}(r^{s'};z)\|_2^2\right]}\leq\sqrt{\Phi_{C_\ell}(T)M_{\mathcal{S}}}\cdot|s-s'|. \tag{177}
$$

This concludes the proof.

B.4 Proofs for contraction property of the mapping $\mathcal{T}$

B.4.1 Proof of Lemma 5.6

Controlling the distance between $\overline{C}_\theta^1$ and $\overline{C}_\theta^2$. By Eq. (80), the equations that define θ_1 and θ_2 can be put as for all $t\in[0,T]$ and $i=1,2$,

$$
\frac{\mathrm{d}}{\mathrm{d}t}\theta_i^t=-(\Lambda^t+\Gamma^t)\theta_i^t-\int_0^t R_\ell(t,s)\theta_i^s\mathrm{d}s+u_i^t, \tag{178}
$$

where u_i^t are centered Gaussian processes with autocovariances C_ℓ^i/δ and $R_\ell:=R_\ell^1=R_\ell^2$. By definition, we can couple u_1^t and u_2^t such that

$$
\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\|u_1^t-u_2^t\|_2^2\right]}\leq 2\cdot\mathsf{dist}_{\lambda,T}\left(u_1^t,u_2^t\right)=2\cdot\mathsf{dist}_{\lambda,T}\left(C_\ell^1/\delta,C_\ell^2/\delta\right)=\frac{2}{\sqrt{\delta}}\mathsf{dist}_{\lambda,T}\left(C_\ell^1,C_\ell^2\right). \tag{179}
$$

We observe that

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left\|\theta_1^t-\theta_2^t\right\|_2 &\leq\left\|\frac{\mathrm{d}}{\mathrm{d}t}\left(\theta_1^t-\theta_2^t\right)\right\|_2\\
&=\left\|\int_0^t R_\ell(t,s)(\theta_1^s-\theta_2^s)\mathrm{d}s+\left(\Lambda^t+\Gamma^t\right)\left(\theta_1^t-\theta_2^t\right)-\left(u_1^t-u_2^t\right)\right\|_2\\
&\leq\int_0^t\Phi_{R_\ell}(t-s)\left\|\theta_1^s-\theta_2^s\right\|_2\mathrm{d}s+(M_\Lambda+M_\ell)\left\|\theta_1^t-\theta_2^t\right\|_2+\left\|u_1^t-u_2^t\right\|_2.
\end{aligned} \tag{180}
$$

By Lemma 5.1 we can choose a $\bar\lambda$ large enough such that $\int_0^\infty e^{-\bar\lambda s}\Phi_{R_\ell}(s)\mathrm{d}s\leq M_\Lambda+M_\ell$, which implies that

$$
\begin{aligned}
&e^{-\bar\lambda t}\frac{\mathrm{d}}{\mathrm{d}t}\|\theta_1^t-\theta_2^t\|_2\\
&\leq\int_0^t e^{-\bar\lambda(t-s)}\Phi_{R_\ell}(t-s)\cdot e^{-\bar\lambda s}\left\|\theta_1^s-\theta_2^s\right\|_2\mathrm{d}s+e^{-\bar\lambda t}\left(M_\Lambda+L\right)\left\|\theta_1^t-\theta_2^t\right\|_2+e^{-\bar\lambda t}\left\|u_1^t-u_2^t\right\|_2
\end{aligned}
$$

$$\le \left(\int_0^\infty e^{-\overline{\lambda}s}\Phi_{R_\ell}(s)\mathrm{d}s\right)\cdot \sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2 + e^{-\overline{\lambda}t}\left(M_\Lambda+L\right)\left\|\theta_1^t-\theta_2^t\right\|_2 + e^{-\overline{\lambda}t}\left\|u_1^t-u_2^t\right\|_2$$

$$\le \left(M_\Lambda+M_\ell+\int_0^\infty e^{-\overline{\lambda}s}\Phi_{R_\ell}(s)\mathrm{d}s\right)\cdot \sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2 + e^{-\overline{\lambda}t}\left\|u_1^t-u_2^t\right\|_2$$

$$\le 2\left(M_\Lambda+M_\ell\right)\cdot \sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2 + e^{-\overline{\lambda}t}\left\|u_1^t-u_2^t\right\|_2 \,. \tag{181}$$

Using the observation

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2 &\le \max\left\{\frac{\mathrm{d}}{\mathrm{d}t}e^{-\overline{\lambda}t}\left\|\theta_1^t-\theta_2^t\right\|_2,0\right\} \le \max\left\{e^{-\overline{\lambda}t}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\theta_1^t-\theta_2^t\right\|_2,0\right\} \\ &\le 2\left(M_\Lambda+M_\ell\right)\cdot \sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2 + e^{-\overline{\lambda}t}\left\|u_1^t-u_2^t\right\|_2 \,,\end{aligned} \tag{182}$$

we can derive that

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-2(M_\Lambda+M_\ell)t}\sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2\right) \le e^{-2(M_\Lambda+M_\ell)t-\overline{\lambda}t}\left\|u_1^t-u_2^t\right\|_2 \,, \tag{183}$$

and consequently by Cauchy-Schwarz inequality

$$\begin{aligned}e^{-4(M_\Lambda+M_\ell)t-2\overline{\lambda}t}\left\|\theta_1^t-\theta_2^t\right\|_2^2 &\le \left(e^{-2(M_\Lambda+M_\ell)t}\sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2\right)^2 \le \left(\int_0^t e^{-2(M_\Lambda+M_\ell)s-\overline{\lambda}s}\left\|u_1^s-u_2^s\right\|_2\mathrm{d}s\right)^2 \\ &\le \left(\int_0^t \frac{1}{(t-s+1)^2}\mathrm{d}s\right)\left(\int_0^t (t-s+1)^2 e^{-4(M_\Lambda+M_\ell)s-2\overline{\lambda}s}\left\|u_1^s-u_2^s\right\|_2^2\mathrm{d}s\right) \\ &\le \int_0^t (t-s+1)^2 e^{-4(M_\Lambda+M_\ell)s-2\overline{\lambda}s}\left\|u_1^s-u_2^s\right\|_2^2\mathrm{d}s \,.\end{aligned} \tag{184}$$

Taking expectation on both sides, and choose some $\lambda \ge 2(M_\Lambda+M_\ell)+\overline{\lambda}$, we have

$$\begin{aligned}e^{-2\lambda t}\mathbb{E}\left[\left\|\theta_1^t-\theta_2^t\right\|_2^2\right] &\le e^{-2\left(\lambda-2(M_\Lambda+M_\ell)-\overline{\lambda}\right)t}\cdot e^{-4(M_\Lambda+M_\ell)t-2\overline{\lambda}t}\mathbb{E}\left[\left\|\theta_1^t-\theta_2^t\right\|_2^2\right] \\ &\le e^{-2\left(\lambda-2(M_\Lambda+L)-\overline{\lambda}\right)t}\cdot\int_0^t (t-s+1)^2 e^{-4(M_\Lambda+L)s-2\overline{\lambda}s}\mathbb{E}\left[\left\|u_1^s-u_2^s\right\|_2^2\right]\mathrm{d}s \\ &\le \int_0^t e^{-2\left(\lambda-2(M_\Lambda+L)-\overline{\lambda}\right)(t-s)}(t-s+1)^2\cdot e^{-2\lambda s}\mathbb{E}\left[\left\|u_1^s-u_2^s\right\|_2^2\right]\mathrm{d}s \\ &\le \left(\int_0^\infty e^{-2\left(\lambda-2(M_\Lambda+L)-\overline{\lambda}\right)t}(t+1)^2\mathrm{d}t\right)\cdot\sup_{0\le s\le t} e^{-2\lambda s}\mathbb{E}\left[\left\|u_1^s-u_2^s\right\|_2^2\right] \,.\end{aligned} \tag{185}$$

Taking supremum on both sides for $t\in[0,T]$ and choosing a large enough λ yields

$$\begin{aligned}\sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\theta_1^t-\theta_2^t\right\|_2^2\right]} &\le \left(\int_0^\infty e^{-2\left(\lambda-2(M_\Lambda+L)-\overline{\lambda}\right)t}(t+1)^2\mathrm{d}t\right)\cdot\frac{2}{\sqrt{\delta}}\cdot\mathsf{dist}_{\lambda,T}\left(C_\ell^1,C_\ell^2\right) \\ &\le \epsilon\cdot\mathsf{dist}_{\lambda,T}\left(C_\ell^1,C_\ell^2\right) \,,\end{aligned} \tag{186}$$

for any prescribed $\epsilon>0$. Consider a centered Gaussian process $\begin{bmatrix} g_1 \\ g_2\end{bmatrix}\in\mathbb{R}^{2p}$ with covariance $\mathbb{E}\left[\begin{bmatrix}\theta_1^t\\ \theta_2^t\end{bmatrix}\begin{bmatrix}\theta_1^s\\ \theta_2^s\end{bmatrix}^\mathsf{T}\right]$. Clearly $\mathbb{E}\left[\left\|g_1^t-g_2^t\right\|_2^2\right]=\mathbb{E}\left[\left\|\theta_1^t-\theta_2^t\right\|_2^2\right]$ for all $t\in[0,\infty)$. Since g_1 and g_2 have covariance kernels $\overline{C}_\theta^1$ and $\overline{C}_\theta^2$, we have

$$\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right)\le\sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\theta_1^t-\theta_2^t\right\|_2^2\right]}\le\epsilon\cdot\mathsf{dist}_{\lambda,T}\left(C_\ell^1,C_\ell^2\right) \,. \tag{187}$$

Controlling the distance between $\overline{R}_\theta^1$ and $\overline{R}_\theta^2$. Note that both $\overline{R}_\theta^1$ and $\overline{R}_\theta^2$ are defined by the same ODE by Eq. (81) for $i = 1, 2$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{R}_\theta^i(t,s) = -\Lambda^t \overline{R}_\theta^i(t,s) - \int_s^t R_\ell(t,s')\overline{R}_\theta^i(s',s)\mathrm{d}s' \,, \tag{188}$$

and with the same boundary condition $\overline{R}_\theta^i(s,s) = I$. Thus $\overline{R}_\theta^1 = \overline{R}_\theta^2$ on $[0,T]^2$.

B.4.2 Proof of Lemma 5.7

Controlling the distance between $\overline{C}_\theta^1$ and $\overline{C}_\theta^2$. Since $C_\ell^1 = C_\ell^2$ on $[0,T]^2$, we have for all $t \in [0,T]$ and $i = 1, 2$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta_i^t = -\left(\Lambda^t + \Gamma_i^t\right)\theta_i^t - \int_0^t R_\ell^i(t,s)\theta_i^s \mathrm{d}s + u_t \,, \tag{189}$$

where u_t is a centered Gaussian process with the covariance kernel $C_\ell/\delta := C_\ell^1/\delta = C_\ell^2/\delta$. Using

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\theta_1^t - \theta_2^t\right) = -\left(\Lambda^t + \Gamma_1^t\right)\left(\theta_1^t - \theta_2^t\right) - \left(\Gamma_1^t - \Gamma_2^t\right)\theta_2^t - \int_0^t R_\ell^1(t,s)(\theta_1^s - \theta_2^s)\mathrm{d}s - \int_0^t (R_\ell^1(t,s) - R_\ell^2(t,s))\theta_2^s \mathrm{d}s \,, \tag{190}$$

it follows that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\|\theta_1^t - \theta_2^t\|_2 \le \left\|\frac{\mathrm{d}}{\mathrm{d}t}\left(\theta_1^t - \theta_2^t\right)\right\|_2 \\
&= \left\|\left(\Lambda^t + \Gamma_1^t\right)\left(\theta_1^t - \theta_2^t\right) + \int_0^t R_\ell^1(t,s)(\theta_1^s - \theta_2^s)\mathrm{d}s + \left(\Gamma_1^t - \Gamma_2^t\right)\theta_2^t + \int_0^t (R_\ell^1(t,s) - R_\ell^2(t,s))\theta_2^s \mathrm{d}s\right\|_2 \\
&\le (M_\Lambda + M_\ell)\left\|\theta_1^t - \theta_2^t\right\|_2 + \int_0^t \Phi_{R_\ell}(t-s)\left\|\theta_1^s - \theta_2^s\right\|_2 \mathrm{d}s + \left\|\Gamma_1^t - \Gamma_2^t\right\|\left\|\theta_2^t\right\|_2 \\
&\qquad + \int_0^t \left\|R_\ell^1(t,s) - R_\ell^2(t,s)\right\|\left\|\theta_2^s\right\|_2 \mathrm{d}s \,.
\end{aligned} \tag{191}$$

By Lemma 5.1 we can choose a $\bar{\lambda}$ large enough such that $\int_0^\infty e^{-\overline{\lambda}s}\Phi_{R_\ell}(s)\mathrm{d}s \le M_\Lambda + M_\ell$, and therefore

$$\begin{aligned}
&e^{-\overline{\lambda}t}\frac{\mathrm{d}}{\mathrm{d}t}\|\theta_1^t - \theta_2^t\|_2 \\
&\le (M_\Lambda + M_\ell)\, e^{-\overline{\lambda}t}\left\|\theta_1^t - \theta_2^t\right\|_2 + \int_0^t e^{-\overline{\lambda}(t-s)}\Phi_{R_\ell}(t-s)\cdot e^{-\overline{\lambda}s}\left\|\theta_1^s - \theta_2^s\right\|_2 \mathrm{d}s \\
&\qquad + e^{-\overline{\lambda}t}\left\|\Gamma_1^t - \Gamma_2^t\right\|\left\|\theta_2^t\right\|_2 + e^{-\overline{\lambda}t}\int_0^t \left\|R_\ell^1(t,s) - R_\ell^2(t,s)\right\|\left\|\theta_2^s\right\|_2 \mathrm{d}s \\
&\le (M_\Lambda + M_\ell)\, e^{-\overline{\lambda}t}\left\|\theta_1^t - \theta_2^t\right\|_2 + \left(\int_0^\infty e^{-\overline{\lambda}s}\Phi_{R_\ell}(s)\mathrm{d}s\right)\cdot \sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s - \theta_2^s\right\|_2 \\
&\qquad + e^{-\overline{\lambda}t}\left\|\Gamma_1^t - \Gamma_2^t\right\|\left\|\theta_2^t\right\|_2 + e^{-\overline{\lambda}t}\int_0^t \left\|R_\ell^1(t,s) - R_\ell^2(t,s)\right\|\left\|\theta_2^s\right\|_2 \mathrm{d}s \\
&= \left(M_\Lambda + M_\ell + \int_0^\infty e^{-\overline{\lambda}s}\Phi_{R_\ell}(s)\mathrm{d}s\right)\cdot \sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s - \theta_2^s\right\|_2 + e^{-\overline{\lambda}t}\left\|\Gamma_1^t - \Gamma_2^t\right\|\left\|\theta_2^t\right\|_2 \\
&\qquad + e^{-\overline{\lambda}t}\int_0^t \left\|R_\ell^1(t,s) - R_\ell^2(t,s)\right\|\left\|\theta_2^s\right\|_2 \mathrm{d}s \\
&\le 2\,(M_\Lambda + M_\ell)\, e^{-\overline{\lambda}t}\sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s - \theta_2^s\right\|_2 + e^{-\overline{\lambda}t}\left\|\Gamma_1^t - \Gamma_2^t\right\|\left\|\theta_2^t\right\|_2 + e^{-\overline{\lambda}t}\int_0^t \left\|R_\ell^1(t,s) - R_\ell^2(t,s)\right\|\left\|\theta_2^s\right\|_2 \mathrm{d}s \,.
\end{aligned} \tag{192}$$

Similar to the proof in Appendix B.4.1, we obtain

$$
\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-2(M_\Lambda+M_\ell)t}\sup_{0\le s\le t} e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2\right)\\
&\le e^{-2(M_\Lambda+M_\ell)t-\overline{\lambda}t}\left(\left\|\Gamma_1^t-\Gamma_2^t\right\|\left\|\theta_2^t\right\|_2+\int_0^t\left\|R_\ell^1(t,s)-R_\ell^2(t,s)\right\|\left\|\theta_2^s\right\|_2\mathrm{d}s\right), \qquad (193)
\end{aligned}
$$

and consequently

$$
\begin{aligned}
&e^{-4(M_\Lambda+M_\ell)t-2\overline{\lambda}t}\left\|\theta_1^t-\theta_2^t\right\|_2^2\\
&\le\left(e^{-2(M_\Lambda+M_\ell)t}\sup_{0\le s\le t}e^{-\overline{\lambda}s}\left\|\theta_1^s-\theta_2^s\right\|_2\right)^2\\
&\le\left(\int_0^t e^{-2(M_\Lambda+M_\ell)s-\overline{\lambda}s}\left(\left\|\Gamma_1^s-\Gamma_2^s\right\|\left\|\theta_2^s\right\|_2+\int_0^s\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|\left\|\theta_{s'}^2\right\|_2\mathrm{d}s'\right)\mathrm{d}s\right)^2\\
&\overset{\text{(i)}}{\le}\left\{\int_0^t(t-s+1)^{-2}\left(1+\int_0^s(s'+1)^{-2}\mathrm{d}s'\right)\mathrm{d}s\right\}\cdot\left\{\int_0^t e^{-4(M_\Lambda+M_\ell)s-2\overline{\lambda}s}(t-s+1)^2\right.\\
&\qquad\left.\cdot\left(\left\|\Gamma_1^s-\Gamma_2^s\right\|^2\left\|\theta_2^s\right\|_2^2+\int_0^s(s'+1)^2\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|^2\left\|\theta_{s'}^2\right\|_2^2\mathrm{d}s'\right)\mathrm{d}s\right\}\\
&\le 2\int_0^t e^{-4(M_\Lambda+M_\ell)s-2\overline{\lambda}s}(t-s+1)^2\left(\left\|\Gamma_1^s-\Gamma_2^s\right\|^2\left\|\theta_2^s\right\|_2^2+\int_0^s(s'+1)^2\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|^2\left\|\theta_{s'}^2\right\|_2^2\mathrm{d}s'\right)\mathrm{d}s, \qquad (194)
\end{aligned}
$$

where we invoke Cauchy-Schwarz inequality in (i). Take expectation on both sides and use Lemma 5.5 which implies that $\mathbb{E}\left[\left\|\theta_{s'}^2\right\|_2^2\right]\le k\left\|\mathbb{E}\left[\theta_{s'}^2\theta_{s'}^{2,\mathsf{T}}\right]\right\|_2\le k\Phi_{C_\theta}(s')$, we have

$$
\begin{aligned}
&e^{-4(M_\Lambda+M_\ell)t-2\overline{\lambda}t}\mathbb{E}\left[\left\|\theta_1^t-\theta_2^t\right\|_2^2\right]\\
&\le 2\int_0^t\left\{\left(e^{-4(M_\Lambda+M_\ell)s-2\overline{\lambda}s}(t-s+1)^2\right)\right.\\
&\qquad\left.\cdot\left(\left\|\Gamma_1^s-\Gamma_2^s\right\|^2\mathbb{E}\left[\left\|\theta_s^2\right\|_2^2\right]+\int_0^s(s'+1)^2\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|^2\mathbb{E}\left[\left\|\theta_{s'}^2\right\|_2^2\right]\mathrm{d}s'\right)\right\}\mathrm{d}s\\
&\le 2k\int_0^t e^{-4(M_\Lambda+M_\ell)s-2\overline{\lambda}s}(t-s+1)^2\left(\left\|\Gamma_1^s-\Gamma_2^s\right\|^2\Phi_{C_\theta}(s)+\int_0^s(s'+1)^2\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|^2\Phi_{C_\theta}(s')\mathrm{d}s'\right)\mathrm{d}s. \qquad (195)
\end{aligned}
$$

Now we take $\lambda>2(M_\Lambda+M_\ell)+\overline{\lambda}$, and for any $t\in[0,T]$,

$$
\begin{aligned}
&e^{-2\lambda t}\mathbb{E}\left[\left\|\theta_1^t-\theta_2^t\right\|_2^2\right]\\
&\le 2k\int_0^t\left[e^{-2\left(\lambda-2(M_\Lambda+M_\ell)-\overline{\lambda}\right)(t-s)}(t-s+1)^2\right.\\
&\qquad\left.\cdot e^{-2\lambda s}\left(\left\|\Gamma_1^s-\Gamma_2^s\right\|^2\Phi_{C_\theta}(s)+k\int_0^s(s'+1)^2\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|^2\Phi_{C_\theta}(s')\mathrm{d}s'\right)\right]\mathrm{d}s\\
&\le 2k\left(\int_0^\infty e^{-2\left(\lambda-2(M_\Lambda+M_\ell)-\overline{\lambda}\right)t}(t+1)^2\mathrm{d}t\right)\\
&\qquad\cdot\sup_{0\le s\le t}e^{-2\lambda s}\left(\left\|\Gamma_1^s-\Gamma_2^s\right\|^2\Phi_{C_\theta}(s)+k\int_0^s(s'+1)^2\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|^2\Phi_{C_\theta}(s')\mathrm{d}s'\right)\\
&\le 2k\left(\int_0^\infty e^{-2\left(\lambda-2(M_\Lambda+M_\ell)-\overline{\lambda}\right)t}(t+1)^2\mathrm{d}t\right)
\end{aligned}
$$

$$\cdot\left(\Phi_{C_\theta}(t)\cdot\sup_{0\le s\le t}e^{-2\lambda s}\left\|\Gamma_1^s-\Gamma_2^s\right\|^2+\left(k\int_0^s(s'+1)^2\Phi_{C_\theta}(s')\mathrm{d}s'\right)\cdot\sup_{0\le s'\le s}e^{-2\lambda s}\left\|R_\ell^1(s,s')-R_\ell^2(s,s')\right\|^2\right)$$

$$\le 2k\left(\int_0^\infty e^{-2\left(\lambda-2(M_\Lambda+M_\ell)-\overline{\lambda}\right)t}(t+1)^2\mathrm{d}t\right)\cdot\left(\Phi_{C_\theta}(T)\cdot\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)^2+kT(T+1)^2\Phi_{C_\theta}(T)\cdot\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right)^2\right). \tag{196}$$

Therefore, we can always take a large enough λ such that for any $\epsilon>0$

$$e^{-\lambda t}\sqrt{\mathbb{E}\left[\|\theta_1^t-\theta_2^t\|_2^2\right]}\le\epsilon\cdot\sqrt{\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)^2+\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right)^2}\le\epsilon\cdot\left(\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)+\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right)\right). \tag{197}$$

Using the same argument in Appendix B.4.1, we conclude that

$$\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right)\le\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\|\theta_1^t-\theta_2^t\|_2^2\right]}\le\epsilon\cdot\left(\mathsf{dist}_{\lambda,T}\left(R_\ell^1,R_\ell^2\right)+\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)\right). \tag{198}$$

Controlling the distance between $\overline{R}_\theta^1$ and $\overline{R}_\theta^2$. Again from Eq. (81) we get for any $0\le s\le t\le T$ and $i=1,2$,

$$\frac{\mathrm{d}}{\mathrm{d}t}\overline{R}_\theta^i(t,s)=-\left(\Lambda^t+\Gamma_i^t\right)\overline{R}_\theta^i(t,s)-\int_s^t R_\ell^i(t,s')\overline{R}_\theta^i(s',s)\mathrm{d}s', \tag{199}$$

with the same boundary conditions $\overline{R}_\theta^i(s,s)=I$, and thus for any $0\le s\le t\le T$,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right)\\
&=-\left(\Lambda^t+\Gamma_1^t\right)\left(\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right)-\left(\Gamma_1^t-\Gamma_2^t\right)\overline{R}_\theta^2(t,s)\\
&\quad-\int_s^t R_\ell^1(t,s')(\overline{R}_\theta^1(s',s)-\overline{R}_\theta^2(s',s))\mathrm{d}s'-\int_s^t(R_\ell^1(t,s')-R_\ell^2(t,s'))\overline{R}_\theta^2(s',s)\mathrm{d}s',
\end{aligned} \tag{200}$$

and $\overline{R}_\theta^1(s,s)-\overline{R}_\theta^2(s,s)=0$. It then follows that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left\|\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right\|\le\left\|\frac{\mathrm{d}}{\mathrm{d}t}\left(\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right)\right\|\\
&\le\left\|\left(\Lambda^t+\Gamma_1^t\right)\left(\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right)+\int_s^t R_\ell^1(t,s')(\overline{R}_\theta^1(s',s)-\overline{R}_\theta^2(s',s))\mathrm{d}s'\right.\\
&\qquad\left.+\left(\Gamma_1^t-\Gamma_2^t\right)\overline{R}_\theta^2(t,s)+\int_s^t(R_\ell^1(t,s')-R_\ell^2(t,s'))\overline{R}_\theta^2(s',s)\mathrm{d}s'\right\|\\
&\le(M_\Lambda+M_\ell)\left\|\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right\|+\int_s^t\Phi_{R_\ell}(t-s')\left\|\overline{R}_\theta^1(s',s)-\overline{R}_\theta^2(s',s)\right\|\mathrm{d}s'\\
&\qquad+\left\|\Gamma_1^t-\Gamma_2^t\right\|\cdot\Phi_{R_\theta}(t-s)+\int_s^t\left\|R_\ell^1(t,s')-R_\ell^2(t,s')\right\|\cdot\Phi_{R_\theta}(s'-s)\mathrm{d}s',
\end{aligned} \tag{201}$$

where in the last line we use $\|\Gamma_1^t\|\le M_\ell$ and $\left\|\overline{R}_\theta^2(t,s)\right\|\le\Phi_{R_\theta}(t-s)$ by invoking Lemma 5.5. We now proceed almost identically to the proof in the previous part. We find some large enough $\overline{\lambda}$ such that $\int_0^\infty e^{-\overline{\lambda}t}\Phi_{R_\ell}(t)\mathrm{d}t\le M_\Lambda+M_\ell$ and on which $\overline{\lambda}$ it holds that for any $0\le s\le t\le T$,

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left(e^{-2(M_\Lambda+M_\ell)t}\sup_{s\le s'\le t}e^{-\overline{\lambda}s'}\left\|\overline{R}_\theta^1(s',s)-\overline{R}_\theta^2(s',s)\right\|\right)\\
&\le e^{-2(M_\Lambda+M_\ell)t-\overline{\lambda}t}\left(\left\|\Gamma_1^t-\Gamma_2^t\right\|\Phi_{R_\theta}(t-s)+\int_s^t\left\|R_\ell^1(t,s')-R_\ell^2(t,s')\right\|\Phi_{R_\theta}(s'-s)\mathrm{d}s'\right)
\end{aligned} \tag{202}$$

and then we have

$$
\begin{aligned}
&e^{-2(M_\Lambda+M_\ell)t-\overline{\lambda}t}\left\|\overline{R}^1_\theta(t,s)-\overline{R}^2_\theta(t,s)\right\| \\
&\le e^{-2(M_\Lambda+M_\ell)t}\sup_{s\le s'\le t}e^{-\overline{\lambda}s'}\left\|\overline{R}^1_\theta(s',s)-\overline{R}^2_\theta(s',s)\right\| \\
&\le \int_s^t e^{-2(M_\Lambda+M_\ell)s'-\overline{\lambda}s'}\left(\left\|\Gamma^1_{s'}-\Gamma^2_{s'}\right\|\Phi_{R_\theta}(s'-s)+\int_s^{s'}\left\|R^1_\ell(s',s'')-R^2_\ell(s',s'')\right\|\Phi_{R_\theta}(s''-s)\mathrm{d}s''\right)\mathrm{d}s'\,.
\end{aligned}
\tag{203}
$$

For any $\lambda>2(M_\Lambda+M_\ell)+\overline{\lambda}$ and $0\le s\le t\le T$,

$$
\begin{aligned}
&e^{-\lambda t}\left\|\overline{R}^1_\theta(t,s)-\overline{R}^2_\theta(t,s)\right\| \\
&\le \int_s^t e^{-\left(\lambda-2(M_\Lambda+M_\ell)-\overline{\lambda}\right)(t-s')} \\
&\qquad\cdot e^{-\lambda s'}\left(\left\|\Gamma^1_{s'}-\Gamma^2_{s'}\right\|\Phi_{R_\theta}(s'-s)+\int_s^{s'}\left\|R^1_\ell(s',s'')-R^2_\ell(s',s'')\right\|\Phi_{R_\theta}(s''-s)\mathrm{d}s''\right)\mathrm{d}s' \\
&\le \frac{1}{\lambda-2\left(M_\Lambda+M_\ell\right)-\overline{\lambda}} \\
&\qquad\cdot\left(\sup_{s\le s'\le t}e^{-\lambda s'}\left\|\Gamma^1_{s'}-\Gamma^2_{s'}\right\|\Phi_{R_\theta}(s'-s)+\sup_{s\le s'\le t}\int_s^{s'}e^{-\lambda s'}\left\|R^1_\ell(s',s'')-R^2_\ell(s',s'')\right\|\Phi_{R_\theta}(s''-s)\mathrm{d}s''\right) \\
&\le \frac{1}{\lambda-2\left(M_\Lambda+M_\ell\right)-\overline{\lambda}} \\
&\qquad\cdot\left(\Phi_{R_\theta}(t)\cdot\sup_{0\le s\le t}e^{-\lambda s}\left\|\Gamma^1_s-\Gamma^2_s\right\|_2+\left(\int_0^t\Phi_{R_\theta}(s)\mathrm{d}s\right)\cdot\sup_{0\le s'\le s\le t}e^{-\lambda s}\left\|R^1_\ell(s,s')-R^2_\ell(s,s')\right\|\right) \\
&\le \frac{T\Phi_{R_\theta}(T)}{\lambda-2\left(M_\Lambda+M_\ell\right)-\overline{\lambda}}\cdot\left(\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)+\mathsf{dist}_{\lambda,T}\left(R^1_\ell,R^2_\ell\right)\right)\,.
\end{aligned}
\tag{204}
$$

For any $\epsilon>0$, we can take a large enough λ such that

$$
\mathsf{dist}_{\lambda,T}\left(\overline{R}^1_\theta,\overline{R}^2_\theta\right)=\sup_{0\le s\le t\le T}e^{-\lambda t}\left\|\overline{R}^1_\theta(t,s)-\overline{R}^2_\theta(t,s)\right\|\le\epsilon\cdot\left(\mathsf{dist}_{\lambda,T}\left(R^1_\ell,R^2_\ell\right)+\mathsf{dist}_{\lambda,T}\left(\Gamma_1,\Gamma_2\right)\right)\,. \tag{205}
$$

B.4.3 Proof of Lemma 5.8

Controlling the distance between $\overline{C}^1_\ell$ and $\overline{C}^2_\ell$. Given that $\overline{R}_\theta:=\overline{R}^1_\theta=\overline{R}^2_\theta$ on $[0,T]^2$, we can write the equations that define r_1 and r_2 as

$$
r^t_i=-\frac{1}{\delta}\int_0^t\overline{R}_\theta(t,s)\ell_s(r^s_i;z)\mathrm{d}s+w^t_i\,, \tag{206}
$$

for $i=1,2$, where w^t_i are centered Gaussian processes with covariance kernels $\overline{C}^i_\theta$. We couple w^t_1 and w^t_2 such that they achieve small (λ,T)-distance, namely

$$
\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|w^t_1-w^t_2\right\|_2^2\right]}\le 2\cdot\mathsf{dist}_{\lambda,T}\left(w^t_1,w^t_2\right)=2\cdot\mathsf{dist}_{\lambda,T}\left(\overline{C}^1_\theta,\overline{C}^2_\theta\right)\,. \tag{207}
$$

For any $t\le T$, we have

$$
\begin{aligned}
&e^{-\lambda t}\left\|r^t_1-r^t_2\right\|_2 \\
&\le e^{-\lambda t}\left(\frac{1}{\delta}\int_0^t\left\|\overline{R}_\theta(t,s)\right\|\left\|\ell_s(r^s_1;z)-\ell_s(r^s_2;z)\right\|_2\mathrm{d}s+\left\|w^t_1-w^t_2\right\|_2\right)
\end{aligned}
$$

$$\leq \frac{1}{\delta}\int_0^t e^{-\lambda(t-s)}\left\|\overline{R}_\theta(t,s)\right\|\cdot e^{-\lambda s}\left\|\ell_s(r_1^s;z)-\ell_s(r_2^s;z)\right\|_2 \mathrm{d}s + e^{-\lambda t}\left\|w_1^t-w_2^t\right\|_2$$
$$\leq \frac{M_\ell}{\delta}\int_0^t e^{-\lambda(t-s)}\Phi_{R_\theta}(t-s)\cdot e^{-\lambda s}\left\|r_1^s-r_2^s\right\|_2 \mathrm{d}s + e^{-\lambda t}\left\|w_1^t-w_2^t\right\|_2 . \tag{208}$$

Therefore square both sides and taking expectations, we have

$$\begin{aligned}
&e^{-2\lambda t}\mathbb{E}\left[\left\|r_1^t-r_2^t\right\|_2^2\right]\\
&\leq \mathbb{E}\left[\left(\frac{M_\ell}{\delta}\int_0^t e^{-\lambda(t-s)}\Phi_{R_\theta}(t-s)\cdot e^{-\lambda s}\left\|r_1^s-r_2^s\right\|_2 \mathrm{d}s + e^{-\lambda t}\left\|w_1^t-w_2^t\right\|_2\right)^2\right]\\
&\leq \left(\int_0^t (t-s+1)^{-2}\mathrm{d}s+1\right)\\
&\quad\cdot \mathbb{E}\left[\frac{M_\ell^2}{\delta^2}\int_0^t (t-s+1)^2 e^{-2\lambda(t-s)}\Phi_{R_\theta}(t-s)^2\cdot e^{-2\lambda s}\left\|r_1^s-r_2^s\right\|_2^2 \mathrm{d}s + e^{-2\lambda t}\left\|w_1^t-w_2^t\right\|_2^2\right]\\
&\leq \frac{2M_\ell^2}{\delta^2}\int_0^t (t-s+1)^2 e^{-2\lambda(t-s)}\Phi_{R_\theta}(t-s)^2\cdot e^{-2\lambda s}\mathbb{E}\left[\left\|r_1^s-r_2^s\right\|_2^2\right] + 2e^{-2\lambda t}\mathbb{E}\left[\left\|w_1^t-w_2^t\right\|_2^2\right]\\
&\leq \frac{2M_\ell^2}{\delta^2}\cdot\left(\int_0^t e^{-2\lambda t}(t+1)^2\Phi_{R_\theta}(t)^2\mathrm{d}t\right)\cdot \sup_{0\leq s\leq t} e^{-2\lambda s}\mathbb{E}\left[\left\|r_1^s-r_2^s\right\|_2^2\right] + 2\cdot\left(2\cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right)\right)^2 .
\end{aligned}\tag{209}$$

Note that the right hand side is increasing in t. By taking λ to be large enough such that

$$\frac{2M_\ell^2}{\delta^2}\cdot\int_0^t e^{-2\lambda t}(t+1)^2\Phi_{R_\theta}(t)^2\mathrm{d}t \leq \frac{1}{2}, \tag{210}$$

we have

$$\sup_{0\leq s\leq t} e^{-2\lambda s}\mathbb{E}\left[\left\|r_1^s-r_2^s\right\|_2^2\right] \leq 16\cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right)^2 . \tag{211}$$

It then can be established following the same argument in Appendix B.4.1 that

$$\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\overline{C}_\ell^1,\overline{C}_\ell^2\right) &\leq \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\ell_t(r_1^t;z)-\ell_t(r_1^t;z)\right\|_2^2\right]} \leq M_\ell\cdot\sqrt{\sup_{0\leq s\leq t} e^{-2\lambda s}\left[\mathbb{E}\left\|r_1^s-r_2^s\right\|_2^2\right]}\\
&\leq 4M_\ell\cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right) .
\end{aligned}\tag{212}$$

Controlling the distances between $\overline{R}_\ell^1$ and $\overline{R}_\ell^2$, $\overline{\Gamma}_1$ and $\overline{\Gamma}_2$. From Eq. (85) we can obtain for any $0\leq s\leq t\leq T$ and $i=1,2$,

$$\frac{\partial \ell_t(r_i^t;z)}{\partial w^s} = \nabla_r \ell_t(r^t;z)\cdot\frac{\partial r_i^t}{\partial w^s}, \tag{213}$$
$$\frac{\partial r_i^t}{\partial w^s} := -\frac{1}{\delta}\int_s^t \overline{R}_\theta(t,s')\frac{\partial \ell_{s'}(r_i^{s'};z)}{\partial w^{s'}}\mathrm{d}s' - \frac{1}{\delta}\overline{R}_\theta(t,s)\nabla_r\ell_s(r_i^s;z). \tag{214}$$

and by Eq. (88),

$$\overline{\Gamma}_i^t = \mathbb{E}\left[\nabla_r\ell_t(r_i^t,z)\right] . \tag{215}$$

Therefore, for any λ satisfying Eq. (210) we have

$$e^{-\lambda t}\left\|\overline{\Gamma}_1^t-\overline{\Gamma}_2^t\right\| \leq e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_2^t;z)\right\|\right] \leq e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_2^t;z)\right\|^2\right]}, \tag{216}$$

and then we use Eq. (211) and obtain

$$e^{-\lambda t}\left\|\overline{\Gamma}_1^t-\overline{\Gamma}_2^t\right\|\le e^{-\lambda t}\sqrt{\mathbb{E}\left[\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_2^t;z)\|^2\right]}$$
$$\le M_\ell\cdot e^{-\lambda t}\sqrt{\mathbb{E}\left[\|r_1^t-r_2^t\|_2^2\right]}=4M_\ell\cdot\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right). \tag{217}$$

Thus

$$\mathsf{dist}_{\lambda,T}\left(\overline{\Gamma}_1,\overline{\Gamma}_2\right)=\sup_{t\in[0,T]}e^{-\lambda t}\left\|\overline{\Gamma}_1^t-\overline{\Gamma}_2^t\right\|\le 4M_\ell\cdot\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right). \tag{218}$$

Next we focus on the λ-distance between $\overline{R}_\ell^1$ and $\overline{R}_\ell^2$. For any $0\le s<t\le T$, we have

$$\overline{R}_\ell^i(t,s)=\mathbb{E}\left[\frac{\partial\ell_t(r_i^t;z)}{\partial w^s}\right]=\mathbb{E}\left[\nabla_r\ell_t(r^t;z)\cdot\frac{\partial r_i^t}{\partial w^s}\right], \tag{219}$$

which implies

$$\begin{aligned}&\left\|\overline{R}_\ell^1(t,s)-\overline{R}_\ell^2(t,s)\right\|\\&=\left\|\mathbb{E}\left[\nabla_r\ell_t(r_1^t;z)\cdot\frac{\partial r_1^t}{\partial w^s}-\nabla_r\ell_t(r_2^t;z)\cdot\frac{\partial r_2^t}{\partial w^s}\right]\right\|\\&\le\mathbb{E}\left[\left\|\nabla_r\ell_t(r_1^t;z)\cdot\frac{\partial r_1^t}{\partial w^s}-\nabla_r\ell_t(r_2^t;z)\cdot\frac{\partial r_2^t}{\partial w^s}\right\|\right]\\&\le\mathbb{E}\left[\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_2^t;z)\|.\left\|\frac{\partial r_1^t}{\partial w^s}\right\|+\|\nabla_r\ell_t(r_2^t;z)\|\cdot\left\|\frac{\partial r_1^t}{\partial w^s}-\frac{\partial r_2^t}{\partial w^s}\right\|\right]\\&\le\sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]}\cdot\sqrt{\mathbb{E}\left[\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_2^t;z)\|^2\right]}+M_\ell\cdot\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}-\frac{\partial r_2^t}{\partial w^s}\right\|\right],\end{aligned} \tag{220}$$

where in the last line we use the fact that $\|\nabla_r\ell_t(r_2^t;z)\|\le M_\ell$. Taking in Eq. (217), we have

$$e^{-\lambda t}\left\|\overline{R}_\ell^1(t,s)-\overline{R}_\ell^2(t,s)\right\|\le e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial\ell_t(r_1^t;z)}{\partial w^s}-\frac{\partial\ell_t(r_2^t;z)}{\partial w^s}\right\|\right]$$
$$\le\sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]}\cdot 4M_\ell\cdot\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1,\overline{C}_\theta^2\right)+M_\ell\cdot e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}-\frac{\partial r_2^t}{\partial w^s}\right\|\right]. \tag{221}$$

It only remains to bound the quantities $\sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]}$ and $e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}-\frac{\partial r_2^t}{\partial w^s}\right\|\right]$. Substituting in the definition of $\frac{\partial r_i^t}{\partial w^s}$ gives us

$$\begin{aligned}\left\|\frac{\partial r_i^t}{\partial w^s}\right\|&\le\frac{1}{\delta}\int_s^t\|\overline{R}_\theta(t,s')\|\left\|\frac{\partial\ell_{s'}(r_i^{s'},z)}{\partial w^s}\right\|\mathrm{d}s'+\frac{1}{\delta}\|\overline{R}_\theta(t,s)\|\cdot\|\nabla_r\ell_s(r_i^s,z)\|\\&\le\frac{1}{\delta}\int_s^t\Phi_{R_\theta}(t-s')\cdot\left\|\nabla_r\ell_{s'}(r_i^{s'},z)\right\|\cdot\left\|\frac{\partial r_i^{s'}}{\partial w^s}\right\|\mathrm{d}s'+\frac{M_\ell}{\delta}\Phi_{R_\theta}(t-s)\\&\le\frac{M_\ell\Phi_{R_\theta}(T)}{\delta}\left(1+\int_s^t\left\|\frac{\partial r_i^{s'}}{\partial w^s}\right\|\mathrm{d}s'\right).\end{aligned} \tag{222}$$

Invoking Gronwall's inequality gives the upper-bound

$$\left\|\frac{\partial r_i^t}{\partial w^s}\right\|\le\frac{M_\ell\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell\Phi_{R_\theta}(T)}{\delta}(t-s)\right)\le\frac{M_\ell\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right), \tag{223}$$

and thus

$$\sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]} \leq \frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right). \tag{224}$$

On the other hand we have using Eq. (214),

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right] \\
&= \frac{1}{\delta} \cdot e^{-\lambda t}\mathbb{E}\left[\left\|\int_s^t \overline{R}_\theta(t,s')\left(\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right)\mathrm{d}s'\right\|\right] \\
&\quad + \frac{1}{\delta} \cdot e^{-\lambda t}\mathbb{E}\left[\left\|\overline{R}_\theta(t,s)\left(\nabla_r \ell_s(r_1^s;z) - \nabla_r \ell_s(r_2^s;z)\right)\right\|\right].
\end{aligned} \tag{225}$$

We bound the two parts separately. First we have

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\int_s^t \overline{R}_\theta(t,s')\left(\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right)\mathrm{d}s'\right\|\right] \\
&\leq \int_s^t e^{-\lambda(t-s')}\left\|\overline{R}_\theta(t,s')\right\| \cdot e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]\mathrm{d}s' \\
&\leq \int_s^t e^{-\lambda(t-s')}\Phi_{R_\theta}(t-s') \cdot e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]\mathrm{d}s' \\
&\leq \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{s\leq s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right].
\end{aligned} \tag{226}$$

Next we get

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\overline{R}_\theta(t,s)\left(\nabla_r \ell_s(r_1^s;z) - \nabla_r \ell_s(r_2^s;z)\right)\right\|\right] \\
&\leq e^{-\lambda(t-s)}\Phi_{R_\theta}(t-s) \cdot e^{-\lambda s}\mathbb{E}\left[\left\|\nabla_r \ell_s(r_1^s;z) - \nabla_r \ell_s(r_2^s;z)\right\|\right] \\
&\leq \Phi_{R_\theta}(T) \cdot 4M_\ell \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right),
\end{aligned} \tag{227}$$

where we invoke Eq. (217) again in the last step. Take Eq. (226) and (227) into (225) and we get

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right] \\
&\leq \frac{1}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{s\leq s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right] + \frac{4M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right).
\end{aligned} \tag{228}$$

Further we substitute Eq. (224) into Eq. (221), we get

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t;z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t;z)}{\partial w^s}\right\|\right] \\
&\leq \sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]} \cdot 4M_\ell \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right) + M_\ell \cdot e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right] \\
&\leq \frac{4M_\ell^2 \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right) \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right) \\
&\quad + \frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{s\leq s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]
\end{aligned}$$

$$+ \frac{4M_\ell^2 \Phi_{R_\theta}(T)}{\delta} \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right)$$

$$\leq \frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{s\leq s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]$$

$$+ \left(\frac{4M_\ell^2\Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right) + \frac{4M_\ell^2\Phi_{R_\theta}(T)}{\delta}\right) \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right). \quad (229)$$

Additionally we can also choose λ such that $\frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \leq \frac{1}{2}$ and using the fact that the right hand side of the inequality is increasing in t, we can get

$$e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t;z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t;z)}{\partial w^s}\right\|\right]$$

$$\leq \sup_{s\leq s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]$$

$$\leq \left(\frac{8M_\ell^2\Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right) + \frac{8M_\ell^2\Phi_{R_\theta}(T)}{\delta}\right) \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right). \quad (230)$$

From Eq. (221), it follows that

$$\mathsf{dist}_{\lambda,T}\left(\overline{R}_\ell^1, \overline{R}_\ell^2\right) \leq \sup_{0\leq s<t\leq T} e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t;z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t;z)}{\partial w^s}\right\|_2\right]$$

$$\leq \left(\frac{8M_\ell^2\Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right) + \frac{8M_\ell^2\Phi_{R_\theta}(T)}{\delta}\right) \cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^1, \overline{C}_\theta^2\right). \quad (231)$$

The proof is completed by taking

$$M := \max\left\{4M_\ell, \left(\frac{8M_\ell^2\Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right) + \frac{8M_\ell^2\Phi_{R_\theta}(T)}{\delta}\right)\right\}. \quad (232)$$

B.4.4 Proof of Lemma 5.9

Controlling the distance between $\overline{C}_\ell^1$ and $\overline{C}_\ell^2$. Since $\overline{C}_\theta^1 = \overline{C}_\theta^2$ on $[0,T]^2$, we can write the equations that define r^1 and r^2 as

$$r_i^t = -\frac{1}{\delta}\int_0^t \overline{R}_\theta^i(t,s)\ell_s(r_i^s;z)\mathrm{d}s + w^t\,, \quad (233)$$

for $i = 1, 2$, where w^t is a centered Gaussian process with autocovariance $\overline{C}_\theta := \overline{C}_\theta^1 = \overline{C}_\theta^2$. For any $t \in [0,T]$, we have

$$e^{-\lambda t}\left\|r_1^t - r_2^t\right\|_2$$

$$\leq e^{-\lambda t}\left(\frac{1}{\delta}\int_0^t \left\|\overline{R}_\theta^1(t,s)\right\| \|\ell_s(r_1^s;z) - \ell_s(r_2^s;z)\|_2\,\mathrm{d}s + \int_0^t \left\|\overline{R}_\theta^1(t,s) - \overline{R}_\theta^2(t,s)\right\| \|\ell_s(r_2^s;z)\|_2\,\mathrm{d}s\right)$$

$$\leq \frac{1}{\delta}\int_0^t e^{-\lambda(t-s)}\left\|\overline{R}_\theta^1(t,s)\right\| \cdot e^{-\lambda s}\|\ell_s(r_1^s;z) - \ell_s(r_2^s;z)\|_2\,\mathrm{d}s + \int_0^t e^{-\lambda t}\left\|\overline{R}_\theta^1(t,s) - \overline{R}_\theta^2(t,s)\right\| \|\ell_s(r_2^s;z)\|_2\,\mathrm{d}s$$

$$\leq \frac{M_\ell}{\delta}\int_0^t e^{-\lambda(t-s)}\Phi_{R_\theta}(t-s) \cdot e^{-\lambda s}\|r_1^s - r_2^s\|_2\,\mathrm{d}s + \left(\int_0^t \|\ell_s(r_2^s;z)\|_2\,\mathrm{d}s\right) \cdot \sup_{0\leq s\leq t\leq T} e^{-\lambda t}\left\|\overline{R}_\theta^1(t,s) - \overline{R}_\theta^2(t,s)\right\|$$

$$\leq \frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{0\leq s\leq t} e^{-\lambda s}\|r_1^s - r_2^s\|_2 + \left(\int_0^t \|\ell_s(r_2^s;z)\|_2\,\mathrm{d}s\right) \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right). \quad (234)$$

The right hand side is increasing in t and therefore

$$\sup_{0\leq s\leq t} e^{-\lambda s}\|r_1^s - r_2^s\|_2 \leq \frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{0\leq s\leq t} e^{-\lambda s}\|r_1^s - r_2^s\|_2$$

$$+\left(\int_0^t \|\ell_s(r_2^s;z)\|_2\,\mathrm{d}s\right)\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\,, \tag{235}$$

which by choosing λ large enough such that $\frac{M_\ell}{\delta}\cdot\left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right)\le\frac{1}{2}$ yields

$$e^{-\lambda t}\left\|r_1^t-r_2^t\right\|_2\le\sup_{0\le s\le t}e^{-\lambda s}\left\|r_1^s-r_2^s\right\|_2\le 2\left(\int_0^t \|\ell_s(r_2^s;z)\|_2\,\mathrm{d}s\right)\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right). \tag{236}$$

Therefore

$$\begin{aligned}
e^{-\lambda t}\sqrt{\mathbb{E}\left[\|r_1^t-r_2^t\|_2^2\right]} &\le 2\sqrt{\mathbb{E}\left[\left(\int_0^t \|\ell_s(r_2^s;z)\|_2\,\mathrm{d}s\right)^2\right]}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\\
&\le 2\sqrt{\mathbb{E}\left[t\cdot\int_0^t \|\ell_s(r_2^s;z)\|_2^2\,\mathrm{d}s\right]}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\\
&= 2\sqrt{t\cdot\int_0^t \mathbb{E}\left[\mathrm{Tr}\left(\ell_s(r_2^s;z)\ell_s(r_2^s;z)^\top\right)\right]\mathrm{d}s}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\\
&\le 2\sqrt{kt\cdot\int_0^t \|\mathbb{E}\left[\ell_s(r_2^s;z)\ell_s(r_2^s;z)^\top\right]\|\,\mathrm{d}s}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\\
&\le 2\sqrt{kt\cdot\int_0^t \Phi_{C_\ell}(s)\mathrm{d}s}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\\
&\le 2t\sqrt{k\Phi_{C_\ell}(t)}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\,,
\end{aligned} \tag{237}$$

and consequently by Lipschitz continuity

$$e^{-\lambda t}\sqrt{\mathbb{E}\left[\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_1^t;z)\|_2^2\right]}\le M_\ell\cdot e^{-\lambda t}\sqrt{\mathbb{E}\left[\|r_1^t-r_2^t\|_2^2\right]}\le 2M_\ell t\sqrt{k\Phi_{C_\ell}(t)}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right). \tag{238}$$

We then use the same argument in Appendix B.4.1 which gives us

$$\mathsf{dist}_{\lambda,T}\left(\overline{C}_\ell^1,\overline{C}_\ell^2\right)\le\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_1^t;z)\|_2^2\right]}\le 2M_\ell t\sqrt{k\Phi_{C_\ell}(t)}\cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right). \tag{239}$$

Controlling the distances between $\overline{R}_\ell^1$ and $\overline{R}_\ell^2$, $\overline{\Gamma}_1$ and $\overline{\Gamma}_2$. From Eq. (85) we can obtain for any $0\le s\le t\le T$ and $i=1,2$,

$$\frac{\partial\ell_t(r_i^t;z)}{\partial w^s}=\nabla_r\ell_t(r^t;z)\cdot\frac{\partial r_i^t}{\partial w^s}\,, \tag{240}$$

$$\frac{\partial r_i^t}{\partial w^s}:=-\frac{1}{\delta}\int_s^t\overline{R}_\theta^i(t,s')\frac{\partial\ell_{s'}(r_i^{s'};z)}{\partial w^{s'}}\mathrm{d}s'-\frac{1}{\delta}\overline{R}_\theta^i(t,s)\nabla_r\ell_s(r_i^s;z)\,. \tag{241}$$

and by Eq. (88),

$$\overline{\Gamma}_i^t=\mathbb{E}\left[\nabla_r\ell_t(r_i^t,z)\right]\,. \tag{242}$$

First, for any λ satisfying $\frac{M_\ell}{\delta}\cdot\left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right)\le\frac{1}{2}$, we have

$$e^{-\lambda t}\left\|\overline{\Gamma}_1^t-\overline{\Gamma}_2^t\right\|_2\le e^{-\lambda t}\mathbb{E}\left[\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_2^t;z)\|\right]\le e^{-\lambda t}\sqrt{\mathbb{E}\left[\|\nabla_r\ell_t(r_1^t;z)-\nabla_r\ell_t(r_2^t;z)\|^2\right]}$$

$$= 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)} \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right), \tag{243}$$

where in the last line we invoke Eq. (238). Thus

$$\mathsf{dist}_{\lambda,T}\left(\overline{\Gamma}^1, \overline{\Gamma}^2\right) = \sup_{t\in[0,T]} e^{-\lambda t}\left\|\overline{\Gamma}_1^t - \overline{\Gamma}_2^t\right\|_2 \le 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)} \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right). \tag{244}$$

Next we turn to the (λ, T)-distance between $\overline{R}_\ell^1$ and $\overline{R}_\ell^2$. Note that for any $0 \le s < t \le T$, we have

$$\overline{R}_\ell^i(t,s) = \mathbb{E}\left[\frac{\partial \ell_t(r_i^t; z)}{\partial w^s}\right] = \mathbb{E}\left[\nabla_r \ell_t(r_i^t; z) \cdot \frac{\partial r_i^t}{\partial w^s}\right], \tag{245}$$

which gives us that

$$\begin{aligned}
&\left\|\overline{R}_\ell^1(t,s) - \overline{R}_\ell^2(t,s)\right\| \le \mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t; z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t; z)}{\partial w^s}\right\|\right] \\
&= \mathbb{E}\left[\left\|\nabla_r \ell_t(r_1^t; z) \cdot \frac{\partial r_1^t}{\partial w^s} - \nabla_r \ell_t(r_2^t; z) \cdot \frac{\partial r_2^t}{\partial w^s}\right\|\right] \\
&\le \mathbb{E}\left[\left\|\nabla_r \ell_t(r_1^t; z) - \nabla_r \ell_t(r_2^t; z)\right\| \cdot \left\|\frac{\partial r_1^t}{\partial w^s}\right\| + \left\|\nabla_r \ell_t(r_2^t; z)\right\| \cdot \left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right] \\
&\le \sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]} \cdot \sqrt{\mathbb{E}\left[\left\|\nabla_r \ell_t(r_1^t; z) - \nabla_r \ell_t(r_2^t; z)\right\|^2\right]} + M_\ell \cdot \mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right],
\end{aligned} \tag{246}$$

where in the last line we use Cauchy-Schwarz inequality and $\|\nabla_r \ell_t(r_2^t; z)\| \le M_\ell$. Taking in Eq. (243), we can have for all λ that $\frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \le \frac{1}{2}$,

$$\begin{aligned}
&e^{-\lambda t}\left\|\overline{R}_\ell^1(t,s) - \overline{R}_\ell^2(t,s)\right\| \le e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t; z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t; z)}{\partial w^s}\right\|\right] \\
&\le \sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]} \cdot 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)} \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right) + M_\ell \cdot e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right].
\end{aligned} \tag{247}$$

It only remains to bound the quantities $\sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]}$ and $e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right]$. From Eq. (241) we have

$$\begin{aligned}
\left\|\frac{\partial r_i^t}{\partial w^s}\right\| &\le \frac{1}{\delta}\int_s^t \left\|\overline{R}_\theta^i(t,s')\right\| \left\|\frac{\partial \ell_{s'}(r_{s'}^i; z)}{\partial w^s}\right\| \mathrm{d}s' + \frac{1}{\delta}\left\|\overline{R}_\theta^i(t,s)\right\|_2 \cdot \|\nabla_r \ell_s(r_i^s; z)\|_2 \\
&\le \frac{1}{\delta}\int_s^t \Phi_{R_\theta}(t-s') \cdot \left\|\nabla_r \ell_{s'}(r_i^{s'}; z)\right\| \cdot \left\|\frac{\partial r_{s'}^i}{\partial w^s}\right\| \mathrm{d}s' + \frac{M_\ell}{\delta}\Phi_{R_\theta}(t-s) \\
&\le \frac{M_\ell \Phi_{R_\theta}(T)}{\delta}\left(1 + \int_s^t \left\|\frac{\partial r_{s'}^i}{\partial w^s}\right\| \mathrm{d}s'\right).
\end{aligned} \tag{248}$$

This allows us to invoke Gronwall's inequality, giving an non-random upper bound

$$\left\|\frac{\partial r_i^t}{\partial w^s}\right\| \le \frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell \Phi_{R_\theta}(T)}{\delta}(t-s)\right) \le \frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right), \tag{249}$$

and thus

$$\sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]} \le \frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right). \tag{250}$$

On the other hand we have

$$
\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}-\frac{\partial r_2^t}{\partial w^s}\right\|\right]\\
&=\frac{1}{\delta}\cdot e^{-\lambda t}\mathbb{E}\left[\left\|\int_s^t\left(\overline{R}_\theta^1(t,s')\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s}-\overline{R}_\theta^2(t,s')\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right)\mathrm{d}s'\right\|\right]\\
&\quad+\frac{1}{\delta}\cdot e^{-\lambda t}\mathbb{E}\left[\left\|\overline{R}_\theta^1(t,s)\nabla_r\ell_r(r_1^s;z)-\overline{R}_\theta^2(t,s)\nabla_r\ell_r(r_2^s;z)\right\|\right]. \qquad (251)
\end{aligned}
$$

We bound the two parts respectively, first we have

$$
\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\int_s^t\left(\overline{R}_\theta^1(t,s')\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s}-\overline{R}_\theta^2(t,s')\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right)\mathrm{d}s'\right\|\right]\\
&\le e^{-\lambda t}\mathbb{E}\left[\int_s^t\left\|\overline{R}_\theta^1(t,s')\left(\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s}-\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right)\right\|\mathrm{d}s'\right]\\
&\quad+e^{-\lambda t}\mathbb{E}\left[\int_s^t\left\|\left(\overline{R}_\theta^1(t,s')-\overline{R}_\theta^2(t,s')\right)\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\mathrm{d}s'\right]\\
&\le\int_s^t e^{-\lambda(t-s')}\left\|\overline{R}_\theta^1(t,s')\right\|\cdot e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s}-\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]\mathrm{d}s'\\
&\quad+\int_s^t e^{-\lambda t}\left\|\overline{R}_\theta^1(t,s')-\overline{R}_\theta^2(t,s')\right\|\cdot\mathbb{E}\left[\left\|\nabla_r\ell_{s'}(r_2^{s'};z)\cdot\frac{\partial r_{s'}^2}{\partial w^s}\right\|\right]\mathrm{d}s'\\
&\le\int_s^t e^{-\lambda(t-s')}\Phi_{R_\theta}(t-s')\cdot e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s}-\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]\mathrm{d}s'\\
&\quad+M_\ell\int_s^t e^{-\lambda t}\left\|\overline{R}_\theta^1(t,s')-\overline{R}_\theta^2(t,s')\right\|\cdot\mathbb{E}\left[\left\|\frac{\partial r_{s'}^2}{\partial w^s}\right\|\right]\mathrm{d}s'\\
&\le\left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right)\cdot\sup_{s<s'\le t}e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s}-\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]\\
&\quad+M_\ell\left(\int_s^t\mathbb{E}\left[\left\|\frac{\partial r_{s'}^2}{\partial w^s}\right\|\right]\mathrm{d}s'\right)\cdot\mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)\\
&\le\left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right)\cdot\sup_{s<s'\le t}e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s}-\frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right]\\
&\quad+\frac{M_\ell^2T\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)\cdot\mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right), \qquad (252)
\end{aligned}
$$

where in the last line we use the upper bound from Eq. (249). For the second term in Eq. (251) we have

$$
\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\overline{R}_\theta^1(t,s)\nabla_r\ell_r(r_1^s;z)-\overline{R}_\theta^2(t,s)\nabla_r\ell_r(r_2^s;z)\right\|\right]\\
&\le e^{-\lambda t}\mathbb{E}\left[\left\|\left(\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right)\nabla_r\ell_r(r_1^s;z)\right\|\right]+e^{-\lambda t}\mathbb{E}\left[\left\|\overline{R}_\theta^2(t,s)\left(\nabla_r\ell_r(r_1^s;z)-\nabla_r\ell_r(r_2^s;z)\right)\right\|\right]\\
&\le M_\ell\cdot e^{-\lambda t}\left\|\overline{R}_\theta^1(t,s)-\overline{R}_\theta^2(t,s)\right\|+e^{-\lambda(t-s)}\Phi_{R_\theta}(t-s)\cdot e^{-\lambda s}\mathbb{E}\left[\|\nabla_r\ell_r(r_1^s;z)-\nabla_r\ell_r(r_2^s;z)\|\right]\\
&\le M_\ell\cdot\mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right)+\Phi_{R_\theta}(T)\cdot 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)}\cdot\mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1,\overline{R}_\theta^2\right), \qquad (253)
\end{aligned}
$$

where we invoke Eq. (243) in the last line. Define

$$
\overline{M}_1:=\frac{M_\ell^2T\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)+M_\ell+\Phi_{R_\theta}(T)\cdot 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)}, \qquad (254)
$$

$$\overline{M}_2 := \frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right) \cdot 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)} + \frac{M_\ell}{\delta}\overline{M}_1\,, \tag{255}$$

and take Eqs. (252) and (253) into (251) and we get

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right] \\
&\leq \frac{1}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{s<s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right] + \frac{1}{\delta}\overline{M}_1 \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right).
\end{aligned} \tag{256}$$

Further substituting Eq. (250) into Eq. (247), it follows that

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t;z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t;z)}{\partial w^s}\right\|\right] \\
&\leq \sqrt{\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s}\right\|^2\right]} \cdot 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)} \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right) + M_\ell \cdot e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial r_1^t}{\partial w^s} - \frac{\partial r_2^t}{\partial w^s}\right\|\right] \\
&\leq \frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right) \cdot 2M_\ell T\sqrt{k\Phi_{C_\ell}(T)} \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right) \\
&\quad + \frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{s<s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right] \\
&\quad + \frac{M_\ell}{\delta}\overline{M}_1 \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right) \\
&\leq \frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \cdot \sup_{s<s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right] + \overline{M}_2 \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right).
\end{aligned} \tag{257}$$

Recall that we choose λ such that $\frac{M_\ell}{\delta} \cdot \left(\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\right) \leq \frac{1}{2}$ and the right hand side of the inequality is increasing in t, we can get

$$\begin{aligned}
e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t;z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t;z)}{\partial w^s}\right\|_2\right] &\leq \sup_{s<s'\leq t} e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial \ell_{s'}(r_1^{s'};z)}{\partial w^s} - \frac{\partial \ell_{s'}(r_2^{s'};z)}{\partial w^s}\right\|\right] \\
&\leq 2\overline{M}_2 \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right).
\end{aligned} \tag{258}$$

Again by Eq. (247),

$$\mathsf{dist}_{\lambda,T}\left(\overline{R}_\ell^1, \overline{R}_\ell^2\right) \leq \sup_{0\leq s<t\leq T} e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(r_1^t;z)}{\partial w^s} - \frac{\partial \ell_t(r_2^t;z)}{\partial w^s}\right\|_2\right] \leq 2\overline{M}_2 \cdot \mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^1, \overline{R}_\theta^2\right). \tag{259}$$

The proof is completed by taking

$$M := \max\left\{2M_\ell T\sqrt{k\Phi_{C_\ell}(T)}, 2\overline{M}_2\right\}. \tag{260}$$

C Auxiliary lemmas for the proof of Theorem 2

C.1 Proof of Lemma 6.1

Since

$$\mathrm{W}_2\left(\widehat{\mu}_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}, \widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}\right) \leq \sqrt{\frac{1}{d}\sum_{j=1}^d\sum_{l=1}^m \left\|\boldsymbol{\theta}_j^{\tau_l} - (\boldsymbol{\theta}_\eta^{\tau_l})_j\right\|_2^2} = \sqrt{\frac{1}{d}\sum_{l=1}^m \|\boldsymbol{\theta}^{\tau_l} - \boldsymbol{\theta}_\eta^{\tau_l}\|_F^2}\,, \tag{261}$$

the lemma follows immediately once we establish claim (110). To prove this, we first have

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left\|\boldsymbol{\theta}^t\right\| &\le \left\|\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\theta}^t\right\| \\
&= \left\|-\boldsymbol{\theta}^t\Lambda^t - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t;\boldsymbol{z})\right\| \\
&= \left\|-\boldsymbol{\theta}^t\Lambda^t - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\left(\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t;\boldsymbol{z}) - \boldsymbol{\ell}_t(\boldsymbol{0};\boldsymbol{z})\right) - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_t(\boldsymbol{0};\boldsymbol{z})\right\| \\
&\le \left(M_\Lambda + \frac{M_\ell\left\|\boldsymbol{X}\right\|^2}{\delta}\right)\left\|\boldsymbol{\theta}^t\right\| + \frac{\left\|\boldsymbol{X}\right\|\left\|\boldsymbol{\ell}_t(\boldsymbol{0};\boldsymbol{z})\right\|}{\delta}, \qquad (262)
\end{aligned}$$

where in the last line we use Assumption 1. Thus by Gronwall we have

$$\left\|\boldsymbol{\theta}^t\right\| \le e^{\left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\right)t}\cdot\left(\left\|\boldsymbol{\theta}^0\right\| + \frac{t\left\|\boldsymbol{X}\right\|\left\|\boldsymbol{\ell}_t(\boldsymbol{0};\boldsymbol{z})\right\|}{\delta}\right). \qquad (263)$$

Recall the discrete time approximation system $\mathcal{F}^\eta$ in Eq. (107),

$$\frac{\mathrm{d}}{\mathrm{d}t}\boldsymbol{\theta}^t_\eta = -\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor} - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{X}\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta;\boldsymbol{z}),$$

which allows us to write

$$\frac{\mathrm{d}}{\mathrm{d}t}\left(\boldsymbol{\theta}^t - \boldsymbol{\theta}^t_\eta\right) = -\left(\boldsymbol{\theta}^t\Lambda^t - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor}\right) - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\left(\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t;\boldsymbol{z}) - \boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{X}\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta;\boldsymbol{z})\right). \qquad (264)$$

Therefore, we can bound

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^t_\eta\right\| \\
&\le \left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^t_\eta\right\|\left\|\Lambda^t\right\| + \left\|\boldsymbol{\theta}^t_\eta\Lambda^t - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor}\right\| + \frac{M_\ell\left\|\boldsymbol{X}\right\|}{\delta}\cdot\left(\left\|\boldsymbol{X}\right\|\left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\| + |t - \lfloor t\rfloor|\right) \\
&\le M_\Lambda\left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^t_\eta\right\| + \left\|\boldsymbol{\theta}^t_\eta\Lambda^t - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor}\right\| + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\left(\left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^t_\eta\right\| + \left\|\boldsymbol{\theta}^t_\eta - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\|\right) + \frac{M_\ell\eta\left\|\boldsymbol{X}\right\|}{\delta} \\
&\le \left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\right)\left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^t_\eta\right\| + \frac{M_\ell\eta\left\|\boldsymbol{X}\right\|}{\delta} + \left\|\boldsymbol{\theta}^t_\eta\Lambda^t - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor}\right\| + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\left\|\boldsymbol{\theta}^t_\eta - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\|. \qquad (265)
\end{aligned}$$

Substituting in $\boldsymbol{\theta}^t_\eta = \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta - (t - \lfloor t\rfloor)\left(\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor} + \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{X}\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta;\boldsymbol{z})\right)$ gives us

$$\begin{aligned}
&\left\|\boldsymbol{\theta}^t_\eta\Lambda^t - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor}\right\| + \frac{M_\ell\left\|\boldsymbol{X}\right\|^2}{\delta}\left\|\boldsymbol{\theta}^t_\eta - \boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\| \\
&\le \left\|\Lambda^t - \Lambda^{\lfloor t\rfloor}\right\|\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\| + \left\|(t - \lfloor t\rfloor)\left(\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor} + \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{X}\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta;\boldsymbol{z})\right)\Lambda^t\right\| \\
&\qquad + \frac{M_\ell\left\|\boldsymbol{X}\right\|^2}{\delta}\left\|(t - \lfloor t\rfloor)\left(\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\Lambda^{\lfloor t\rfloor} + \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{X}\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta;\boldsymbol{z})\right)\right\| \\
&\le M_\Lambda\eta\cdot\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\|_2 + \max_{s\in[0,t]}\left\|\Lambda^s\right\|\eta\cdot\left\{\left(\max_{s\in[0,t]}\left\|\Lambda^s\right\| + \frac{M_\ell\left\|\boldsymbol{X}\right\|^2}{\delta}\right)\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\| + \frac{\left\|\boldsymbol{X}\right\|\left\|\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{0};\boldsymbol{z})\right\|}{\delta}\right\} \\
&\qquad + \frac{M_\ell\eta\left\|\boldsymbol{X}\right\|^2}{\delta}\left\{\left(\max_{s\in[0,t]}\left\|\Lambda^s\right\| + \frac{M_\ell\left\|\boldsymbol{X}\right\|^2}{\delta}\right)\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\| + \frac{\left\|\boldsymbol{X}\right\|\left\|\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{0};\boldsymbol{z})\right\|}{\delta}\right\} \\
&\overset{\text{(i)}}{\le} M_\Lambda\eta\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\| + \eta\left(M_\Lambda + \frac{M_\ell\left\|\boldsymbol{X}\right\|^2}{\delta}\right)^2\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}_\eta\right\| + \eta\left(M_\Lambda + \frac{M_\ell\left\|\boldsymbol{X}\right\|^2}{\delta}\right)\frac{\left\|\boldsymbol{X}\right\|\left\|\boldsymbol{\ell}_{\lfloor t\rfloor}(\boldsymbol{0};\boldsymbol{z})\right\|}{\delta}
\end{aligned}$$

$$\leq \lambda_1(\eta, t, \|\boldsymbol{X}\|)\left(\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}\right\| + \left\|\boldsymbol{\theta}^{\lfloor t\rfloor} - \boldsymbol{\theta}_\eta^{\lfloor t\rfloor}\right\| + \left\|\boldsymbol{\ell}_{\lfloor t\rfloor}(\mathbf{0};\boldsymbol{z})\right\|\right), \tag{266}$$

where in (i) we use $\max_{s\in[0,t]}\|\Lambda^s\| \leq M_\Lambda$ and we define

$$\lambda_1 := \lambda_1(\eta, t, u) = \eta \cdot \left\{M_\Lambda + \frac{M_\Lambda u^2}{\delta} + \left(M_\Lambda + \frac{M_\Lambda u^2}{\delta}\right)^2\right\}. \tag{267}$$

Making use of Eqs. (265), (266) and the fact that

$$\frac{\mathrm{d}}{\mathrm{d}t}\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\| \leq \frac{\mathrm{d}}{\mathrm{d}t}\left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}_\eta^t\right\|,$$

we can further obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\| \\
&\leq \left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\right)\left\|\boldsymbol{\theta}^t - \boldsymbol{\theta}_\eta^t\right\| + \frac{M_\ell\eta\|\boldsymbol{X}\|}{\delta} + \lambda_1(\eta, t, \|\boldsymbol{X}\|)\left(\left\|\boldsymbol{\theta}^{\lfloor t\rfloor}\right\| + \left\|\boldsymbol{\theta}^{\lfloor t\rfloor} - \boldsymbol{\theta}_\eta^{\lfloor t\rfloor}\right\| + \left\|\boldsymbol{\ell}_{\lfloor t\rfloor}(\mathbf{0};\boldsymbol{z})\right\|\right) \\
&\leq \left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta} + \lambda_1\right)\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\| + \frac{M_\ell\eta\|\boldsymbol{X}\|}{\delta} \\
&\quad + \lambda_1\left(e^{\left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\right)t}\cdot\left(\left\|\boldsymbol{\theta}^0\right\| + \frac{t\|\boldsymbol{X}\|\left\|\boldsymbol{\ell}_t(\mathbf{0};\boldsymbol{z})\right\|}{\delta}\right) + \left\|\boldsymbol{\ell}_{\lfloor t\rfloor}(\mathbf{0};\boldsymbol{z})\right\|\right) \\
&\leq \left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta} + \lambda_1\right)\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\| + \left(\frac{M_\ell\eta\|\boldsymbol{X}\|}{\delta} + \lambda_1 M_\ell\eta\right) + \\
&\quad + \lambda_1\left(e^{\left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\right)t}\cdot\left(\left\|\boldsymbol{\theta}^0\right\| + \frac{t\|\boldsymbol{X}\|\left\|\boldsymbol{\ell}_t(\mathbf{0};\boldsymbol{z})\right\|}{\delta}\right) + \left\|\boldsymbol{\ell}_t(\mathbf{0};\boldsymbol{z})\right\|\right) \\
&\leq \left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta} + \lambda_1\right)\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\| + \lambda_2 + \lambda_3\left(\left\|\boldsymbol{\theta}^0\right\| + \left\|\boldsymbol{\ell}_t(\mathbf{0};\boldsymbol{z})\right\|\right),
\end{aligned} \tag{268}$$

where

$$\lambda_2 := \lambda_2(\eta, t, \|\boldsymbol{X}\|) = \frac{M_\ell\eta\|\boldsymbol{X}\|}{\delta} + \lambda_1 M_\ell\eta, \tag{269}$$

$$\lambda_3 := \lambda_3(\eta, t, \|\boldsymbol{X}\|) = \lambda_1 e^{\left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta}\right)t}\left(1 + \frac{t\|\boldsymbol{X}\|}{\delta}\right). \tag{270}$$

This would then imply

$$\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\| \leq e^{\left(M_\Lambda + \frac{M_\ell\|\boldsymbol{X}\|^2}{\delta} + \lambda_1\right)t}\cdot\left(\lambda_2 t + \lambda_3 t\left\|\boldsymbol{\theta}^0\right\| + \lambda_3\int_0^t\left\|\boldsymbol{\ell}_s(\mathbf{0};\boldsymbol{z})\right\|\mathrm{d}s\right). \tag{271}$$

Recall the Assumption 1 which allows us to invoke Bai-Yin law (cf. [BS10, Thm. 5.8]) that guarantees the almost sure convergence of the largest singular value of $\boldsymbol{X}$, namely $\|\boldsymbol{X}\| \to 1+\sqrt{\delta}$ with probability 1. Hence for any fixed time point $t \in \mathbb{R}_{\geq 0}$,

$$\limsup_{n\to\infty}\frac{\sup_{0\leq s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\|}{t + t\left\|\boldsymbol{\theta}^0\right\| + \int_0^t\left\|\boldsymbol{\ell}_s(\mathbf{0};\boldsymbol{z})\right\|\mathrm{d}s} \leq e^{\left(M_\Lambda + \frac{M_\ell(1+\sqrt{\delta})^2}{\delta} + \lambda_1(\eta,t,1+\sqrt{\delta})\right)t}(\lambda_2(\eta, t, 1+\sqrt{\delta}) + \lambda_3(\eta, t, 1+\sqrt{\delta})). \tag{272}$$

However, since both $\lambda_2(\eta, t, 1+\sqrt{\delta}) = 0$ and $\lambda_3(\eta, t, 1+\sqrt{\delta}) \to 0$ as $\eta \to 0$, it then holds almost surely that

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\frac{\sup_{s\leq t}\left\|\boldsymbol{\theta}^s - \boldsymbol{\theta}_\eta^s\right\|}{t + t\left\|\boldsymbol{\theta}^0\right\| + \int_0^t\left\|\boldsymbol{\ell}_s(\mathbf{0};\boldsymbol{z})\right\|\mathrm{d}s} = 0. \tag{273}$$

Finally, note that

$$\frac{\left\|\boldsymbol{\theta}^0\right\|_F}{\sqrt{d}} \to \sqrt{\mathbb{E}\left[\left\|\theta^0\right\|^2\right]}, \qquad \frac{\left\|\boldsymbol{\ell}_s\left(\mathbf{0};\boldsymbol{z}\right)\right\|_F}{\sqrt{d}} \to \sqrt{\mathbb{E}\left[\left\|\boldsymbol{\ell}_s(0;z)\right\|^2\right]}, \tag{274}$$

by the assumptions. By equivalence of Frobenius norm and spectral norm, we get almost surely that

$$\lim_{\eta\to 0}\limsup_{n\to\infty}\sup_{s\le t}\frac{\left\|\boldsymbol{\theta}^s-\boldsymbol{\theta}_\eta^s\right\|}{\sqrt{d}} = 0\,. \tag{275}$$

C.2 Proof of Lemma 6.2

We introduce the following approximate message passing (AMP) algorithm that admits an asymptotic characterization by state evolution. For sequences of Lipschitz functions $f_i:\mathbb{R}^{k(i+1)+1}\to\mathbb{R}^k$ and $g_i:\mathbb{R}^{k(i+1)}\to\mathbb{R}^k$ with $i=0,1,\cdots$, we consider the following matrix sequences $\left\{\boldsymbol{a}^{i+1},\boldsymbol{b}^i\right\}_{i\ge 0}$ in $\mathbb{R}^{d\times k}$ and $\mathbb{R}^{n\times k}$ respectively, generated by

$$\boldsymbol{a}^{i+1} = -\frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{f}_i(\boldsymbol{b}^0,\cdots,\boldsymbol{b}^i;\boldsymbol{z}) + \sum_{j=0}^{i}\boldsymbol{g}_j(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^j;\boldsymbol{\theta}^0)\xi_{i,j}\,, \tag{276}$$

$$\boldsymbol{b}^i = \boldsymbol{X}\boldsymbol{g}_i(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^i;\boldsymbol{\theta}^0) + \frac{1}{\delta}\sum_{j=0}^{i-1}\boldsymbol{f}_j(\boldsymbol{b}^0,\cdots,\boldsymbol{b}^j;\boldsymbol{z})\zeta_{i,j}\,, \tag{277}$$

where $\boldsymbol{f}_i,\boldsymbol{g}_i$ are functions that apply f_i,g_i row-wise similar to $\boldsymbol{\ell}_t$. $\{\xi_{i,j}\}_{0\le j\le i}$ and $\{\zeta_{i,j}\}_{0\le j\le i-1}$ are sequences of deterministic matrices in $\mathbb{R}^{k\times k}$ that depend on the $\{f_i,g_i\}_{i\ge 0}$ in a specific way that we shall explicitly define later. The algorithm is initialized by $\boldsymbol{g}_0(\boldsymbol{\theta}^0)=\boldsymbol{\theta}^0,\boldsymbol{b}^0=\boldsymbol{X}\boldsymbol{\theta}^0$. To relate this AMP algorithm with the discretized flow system $\mathfrak{F}^\eta$, we consider the specific choice of

$$\boldsymbol{g}_i(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^i;\boldsymbol{\theta}^0) := \boldsymbol{\theta}_\eta^{t_i}\,, \tag{278}$$

$$\boldsymbol{f}_i(\boldsymbol{b}^0,\cdots,\boldsymbol{b}^i;\boldsymbol{z}) := \boldsymbol{\ell}_{t_i}(\boldsymbol{X}\boldsymbol{\theta}_\eta^{t_i};\boldsymbol{z})\,, \tag{279}$$

where $t_i=i\eta$. We next show that $\boldsymbol{\theta}_\eta^{t_i}$ is indeed a function of $\boldsymbol{a}^1,\cdots,\boldsymbol{a}^i,\boldsymbol{\theta}^0$ and $-\boldsymbol{\ell}_{t_i}(\boldsymbol{X}\boldsymbol{\theta}_\eta^{t_i};\boldsymbol{z})$ is indeed a function of $\boldsymbol{b}^0,\cdots,\boldsymbol{b}^i,\boldsymbol{z}$. This can be seen by induction

$$\begin{aligned}
\boldsymbol{\theta}_\eta^{t_i} &= \boldsymbol{\theta}_\eta^{t_{i-1}} + \eta\cdot\left\{-\boldsymbol{\theta}_\eta^{t_{i-1}}\Lambda^{t_{i-1}} - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{\ell}_{t_{i-1}}(\boldsymbol{X}\boldsymbol{\theta}_\eta^{t_{i-1}};\boldsymbol{z})\right\}\\
&= \boldsymbol{\theta}_\eta^{t_{i-1}} + \eta\cdot\left\{-\boldsymbol{\theta}_\eta^{t_{i-1}}\Lambda^{t_{i-1}} - \frac{1}{\delta}\boldsymbol{X}^\mathsf{T}\boldsymbol{f}_{i-1}(\boldsymbol{b}^0,\cdots,\boldsymbol{b}^{i-1};\boldsymbol{z})\right\}\\
&= \boldsymbol{g}_{i-1}(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^{i-1};\boldsymbol{\theta}^0)\left(I-\eta\Lambda^{t_{i-1}}\right) + \eta\left(\boldsymbol{a}^i - \sum_{j=0}^{i-1}\boldsymbol{g}_j(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^j;\boldsymbol{\theta}^0)\xi_{i-1,j}\right)\,,
\end{aligned} \tag{280}$$

$$\begin{aligned}
\boldsymbol{\ell}_{t_i}(\boldsymbol{X}\boldsymbol{\theta}_\eta^{t_i};\boldsymbol{z}) &= \boldsymbol{\ell}_{t_i}(\boldsymbol{X}\boldsymbol{g}_i(\boldsymbol{a}^1,\cdots,\boldsymbol{a}^i;\boldsymbol{\theta}^0);\boldsymbol{z})\\
&= \boldsymbol{\ell}_{t_i}\left(\boldsymbol{b}^i - \frac{1}{\delta}\sum_{j=0}^{i-1}\boldsymbol{f}_j(\boldsymbol{b}^0,\cdots,\boldsymbol{b}^j;\boldsymbol{z})\zeta_{i,j};\boldsymbol{z}\right)\,.
\end{aligned} \tag{281}$$

By Lipschitz property of ℓ_t in Assumption 1, we can see by this inductive definition, g_i and f_i are all Lipschitz continuous. To apply the standard AMP result in [CL21] we only need to specify the matrices $\{\xi_{i,j}\}_{0\le j\le i}$ and $\{\zeta_{i,j}\}_{0\le j\le i-1}$. To this end we iteratively define sequences of centered Gaussian vectors $\left\{\overline{u}_\eta^{t_{i+1}},\overline{w}_\eta^{t_i}\right\}_{i\ge 0}$ in $\mathbb{R}^k$ according to

$$\mathbb{E}\left[\overline{w}_\eta^{t_i}\left(\overline{w}_\eta^{t_j}\right)^\mathsf{T}\right] = \mathbb{E}\left[g_i(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_i};\theta^0)g_j(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_j};\theta^0)^\mathsf{T}\right]\,, \qquad 0\le j\le i<\infty\,, \tag{282a}$$

$$\mathbb{E}\left[\overline{u}_\eta^{t_{i+1}}\left(\overline{u}_\eta^{t_{j+1}}\right)^{\mathsf{T}}\right]=\frac{1}{\delta}\mathbb{E}\left[f_i(\overline{w}_\eta^{t_0},\cdots,\overline{w}_\eta^{t_i};z)f_j(\overline{w}_\eta^{t_0},\cdots,\overline{w}_\eta^{t_j};z)^{\mathsf{T}}\right], \qquad 0\le j\le i<\infty\,, \tag{282b}$$

$$\zeta_{i,j}=\mathbb{E}\left[\frac{\partial}{\partial\overline{u}_\eta^{t_{j+1}}}g_i(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_i};\theta^0)\right], \qquad 0\le j\le i-1\,, \tag{282c}$$

$$\xi_{i,j}=\mathbb{E}\left[\frac{\partial}{\partial\overline{w}_\eta^{t_j}}f_i(\overline{w}_\eta^{t_0},\cdots,\overline{w}_\eta^{t_i};z)\right], \qquad 0\le j\le i\,. \tag{282d}$$

Here the expectation is taking over the Gaussian random vectors $\overline{u}_\eta^{t_i},\overline{w}_\eta^{t_i}$ and also on the independently distributed random variables $(\theta^0,z)\sim\mu_{\theta^0,z}$.

The above equations define inductively the matrices $\{\xi_{i,j}\}_{0\le j\le i}$, $\{\zeta_{i,j}\}_{0\le j\le i-1}$ and also the Gaussian vectors $\left\{\overline{u}_\eta^{t_{i+1}},\overline{w}_\eta^{t_i}\right\}_{i\ge 0}$. The sequence is initialized by $\overline{w}_\eta^{t_0}\sim\mathsf{N}\left(0,\mathbb{E}\left[\theta^0\left(\theta^0\right)^{\mathsf{T}}\right]\right)$ and $\xi_{0,0}=\mathbb{E}\left[\frac{\partial}{\partial\overline{w}_\eta^{t_0}}f_0(\overline{w}_\eta^{t_0};z)\right]$. Suppose for some $r=0,1,\cdots$, we have define $\overline{u}_\eta^{t_i},\overline{w}_\eta^{t_i}$ and the matrices $\zeta_{i,j},\xi_{i,j}$ for $i\le r$. According to Eqs. (280) and (281), the functions $f_0,\cdots,f_{r+1};g_0,\cdots,g_{r+1}$ are all explicitly defined. Substituting into Eq. (282b) we can then determine $\overline{w}_\eta^{t_{r+1}}$ and next by Eq. (282a) we obtain $\overline{u}_\eta^{t_{r+1}}$. Finally, by Eqs. (282c) and (282d) the matrices $\zeta_{i,j},\xi_{i,j}$ for $i=r+1$ are determined.

Under the conditions of Theorem 2, we can invoke [CL21, Theorem 2.4] and [JM13, Theorem 1] to obtain[2] that, for any fixed $t_1=\eta,\cdots,t_m=m\eta$ and any Lipschitz bounded function $\psi:\mathbb{R}^{k(m+1)}\to\mathbb{R}$,

$$\frac{1}{d}\sum_{j=1}^{d}\psi\left(\left(\boldsymbol{a}_\eta^{t_1}\right)_j,\cdots,\left(\boldsymbol{a}_\eta^{t_m}\right)_j;\left(\boldsymbol{\theta}^0\right)_j\right)\xrightarrow{p}\mathbb{E}\left[\psi\left(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_m};\theta^0\right)\right]. \tag{283}$$

For any Lipschitz bounded function $\widetilde{\psi}:(\mathbb{R}^k)^{m+1}\to\mathbb{R}$, define $\psi:(\mathbb{R}^k)^{m+1}\to\mathbb{R}$ via

$$\psi(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_m};\theta^0):=\widetilde{\psi}(\overline{\theta}_\eta^{t_1},\cdots,\overline{\theta}_\eta^{t_m}),\overline{\theta}_\eta^{t_1}:=g_{i_1}(\overline{u}_\eta^1,\cdots,\overline{u}_\eta^{i_1};\theta^0),$$
$$\cdots$$
$$\overline{\theta}_\eta^{t_m}:=g_{i_m}(\overline{u}_\eta^1,\cdots,\overline{u}_\eta^{i_m};\theta^0).$$

By the Lipschitz property of $g_{i_1},\cdots,g_{i_m}$, ψ is also Lipschitz bounded. We thus proved that, for any Lipschitz boiunded function $\widetilde{\psi}$,

$$\frac{1}{d}\sum_{j=1}^{d}\widetilde{\psi}\left(\left(\boldsymbol{\theta}_\eta^{t_1}\right)_j,\cdots,\left(\boldsymbol{\theta}_\eta^{t_m}\right)_j;\left(\boldsymbol{\theta}^0\right)_j\right)\xrightarrow{p}\mathbb{E}\left[\widetilde{\psi}\left(\overline{\theta}_\eta^{t_1},\cdots,\overline{\theta}_\eta^{t_m};\theta^0\right)\right]. \tag{284}$$

The next lemma relates the random variables $\overline{\theta}_\eta^{t_1},\cdots,\overline{\theta}_\eta^{t_m}$ to the DMFT system $\mathfrak{S}^\eta$. We defer the proof to Appendix C.4.

Lemma C.1. *The discrete-time DMFT system $\mathfrak{S}^\eta$ has a unique solution in the space $\mathcal{S}$ and $(\theta_\eta^t)_{t=i\eta,i\le m}\overset{\mathrm{d}}{=}(\theta_\eta^t)_{t=i\eta,i\le m}$. Further $t\mapsto\theta_\eta^t$ is piecewise linear with knots $t_i=i\eta$.*

Fix T and set $m=T/\eta$. By this lemma, and since $t\mapsto\boldsymbol{\theta}_\eta^t$ is also piecewise linear with knots at $t=i\eta$, Eq. (284) implies that, for any ℓ, any $\tau_1,\ldots\tau_\ell\in[0,T]$, and any bounded Lipschitz function $\psi:(\mathbb{R}^k)^\ell\to\mathbb{R}$, we have

$$\frac{1}{d}\sum_{j=1}^{d}\psi\left(\left(\boldsymbol{\theta}_\eta^{\tau_1}\right)_j,\cdots,\left(\boldsymbol{\theta}_\eta^{\tau_\ell}\right)_j\right)\xrightarrow{p}\mathbb{E}\left[\psi\left(\overline{\theta}_\eta^{t_1},\cdots,\overline{\theta}_\eta^{t_m};\theta^0\right)\right]. \tag{285}$$

The proof is completed by applying the following basic fact about weak convergence to the probability measures $\nu_n=\widehat{\mu}_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_\ell}}$, $\nu=\mu_{\theta_\eta^{\tau_1},\cdots,\theta_\eta^{\tau_m}}$ on $\mathbb{R}^d$, $d=\ell k$.

[2]Note that [JM13, Theorem 1] only considers AMP algorithms on which the nonlinearities depends on the last iterate. However by enlarging the dimension k, this also covers the case of nonlinearities depend on any constant number of previous times. This reduction is explained in several earlier papers, e.g. [Mon21, Appendix A].

Lemma C.2. *Let $(\nu_n)_{n\geq 1}$ be a sequence of random probability measures on $\mathbb{R}^d$, and assume that, for any bounded Lipschitz function $\psi : \mathbb{R}^d \to \mathbb{R}$, we have $\int \psi(x)\, \nu_n(\mathrm{d}x) \xrightarrow{p} \int \psi(x)\, \nu(\mathrm{d}x)$.*

Then $d_{\mathrm{W}}(\nu_n, \nu) \xrightarrow{p} 0$.

Proof. By Lemma 6.4, it is sufficient to show that for any subsequence $(n_j)_{j\geq 1}$ we can construct a further subsequence $(n'_j)_{j\geq 1}$ such that $d_{\mathrm{W}}\left(\nu_{n'_j}, \nu\right) \xrightarrow{a.s.} 0$.

Fix such a subsequence (n_j), and let $(\psi_i)_{i\in\mathbb{N}}$ be a countable collection of bounded Lipschitz functions on $\mathbb{R}^d$ which determine weak convergence (i.e. such that $\int \psi_i(x)\, q_n(\mathrm{d}x) \to \int \psi_i(x)\, q(\mathrm{d}x)$ imply $d_{\mathrm{W}}(q_n, q) \to 0$.) One can take for instance all functions of the form $\psi(x) = (1 - d(x, Q)/\varepsilon)_+$ where $Q \subseteq \mathbb{R}^d$ is a rectangle with rational corners, and $\varepsilon > 0$ is rational.

By Borel-Cantelli, we can construct a subsequence $(n^1_j) \subseteq (n_j)$ such that $\int \psi_1(x)\nu_{n^1_j}(\mathrm{d}x) \to \int \psi_1(x)\nu(\mathrm{d}x)$. Refining this sequence, we obtain, for each k a subsequence (n^k_j) such that $\int \psi_a(x)\nu_{n^k_j}(\mathrm{d}x) \to \int \psi_a(x)\nu(\mathrm{d}x)$ for all $a \leq k$. Taking the diagonal $n'_j = n^j_j$ yields a subsequence along which $d_{\mathrm{W}}\left(\nu_{n'_j}, \nu\right)$ as desired. ☐

This concludes the proof of Lemma 6.2.

Remark C.1. By [BMN20], Eq. (283) holds for test functions ψ which are pseudo-Lipschitz of order 2 when the matrix $\boldsymbol{X}$ has Gaussian entries. In this case, using the same argument as above, we may conclude that (116) holds also for the Wassersetin distance.

C.3 Proof of Lemma 6.3

First, we define the transformation $\mathcal{T}^\eta = \mathcal{T}^\eta_{\overline{\mathcal{S}}\to\mathcal{S}} \circ \mathcal{T}^\eta_{\mathcal{S}\to\overline{\mathcal{S}}}$ where we let $\mathcal{T}^\eta_{\mathcal{S}\to\overline{\mathcal{S}}} : (C_\ell, R_\ell, \Gamma) \mapsto (\overline{C}_\theta, \overline{R}_\theta)$ and $\mathcal{T}^\eta_{\overline{\mathcal{S}}\to\mathcal{S}} : (\overline{C}_\theta, \overline{R}_\theta) \mapsto (\overline{C}_\ell, \overline{R}_\ell, \overline{\Gamma})$. We remind the readers that $\mathcal{T}^\eta_{\mathcal{S}\to\overline{\mathcal{S}}}$ does not necessarily map $\mathcal{S}$ into $\overline{\mathcal{S}}$ and nor does $\mathcal{T}^\eta_{\overline{\mathcal{S}}\to\mathcal{S}}$ map $\overline{\mathcal{S}}$ into $\mathcal{S}$. We use this notation here because exactly similar to our previous definitions of $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}$ and $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}$, the transformation $\mathcal{T}^\eta_{\mathcal{S}\to\overline{\mathcal{S}}}$ is defined by taking the input function triplet through Eqs. (113a) and (114a) and then we obtain $(\overline{C}_\theta, \overline{R}_\theta)$ by Eqs. (113f) and (113c); $\mathcal{T}_{\overline{\mathcal{S}}\to\mathcal{S}}$ is defined by taking the input function pair into Eqs. (113b), (114b) and (115) and $(\overline{C}_\ell, \overline{R}_\ell, \overline{\Gamma})$ is obtained by Eqs. (113g), (113d) and (113e).

As we have shown in the proof of Lemma 6.2, the mappings $\mathcal{T}^\eta_{\mathcal{S}\to\overline{\mathcal{S}}}$ and $\mathcal{T}^\eta_{\overline{\mathcal{S}}\to\mathcal{S}}$ are essentially determined recursively on the discrete time knots $t_i = i\eta$, $i = 0, 1, \cdots$, so they are uniquely defined. We express the solution of the system $\mathfrak{S}^\eta$ as the unique fixed-point of $\mathcal{T}^\eta$, namely if we let $X^\eta = (C^\eta_\ell, R^\eta_\ell, \Gamma_\eta)$ be the function triplet that solves $\mathfrak{S}^\eta$, it holds that

$$\mathcal{T}^\eta(X^\eta) = X^\eta \,. \tag{286}$$

Suppose $\mathcal{T}^\eta_{\mathcal{S}\to\overline{\mathcal{S}}}(X^\eta) = (C^\eta_\theta, R^\eta_\theta)$, we have the following lemma characterizing the unique solution of $\mathfrak{S}^\eta$.

Lemma C.3. *Under the same conditions of Lemma 6.3, the unique solution of $\mathfrak{S}^\eta$ satisfies $X^\eta = (C^\eta_\ell, R^\eta_\ell, \Gamma_\eta) \in \mathcal{S}$ and $(C^\eta_\theta, R^\eta_\theta) \in \overline{\mathcal{S}}$.*

Let $X \in \mathcal{S}$ be the unique fixed-point of $\mathcal{T}$, we can then control the distance between X and X^η by

$$\begin{aligned}\mathsf{dist}_{\lambda,T}(X, X^\eta) &= \mathsf{dist}_{\lambda,T}(\mathcal{T}(X), \mathcal{T}^\eta(X^\eta)) \\ &\leq \underbrace{\mathsf{dist}_{\lambda,T}(\mathcal{T}(X), \mathcal{T}(X^\eta))}_{\text{(I)}} + \underbrace{\mathsf{dist}_{\lambda,T}(\mathcal{T}(X^\eta), \mathcal{T}^\eta(X^\eta))}_{\text{(II)}}\,,\end{aligned} \tag{287}$$

where by Eq. (100) we can choose λ large enough such that

$$\text{(I)} \leq \frac{1}{2}\mathsf{dist}_{\lambda,T}(X, X^\eta)\;. \tag{288}$$

The following lemma controls the quantity (ii). We defer its proof to Appendix C.6.

Lemma C.4. *Under the same conditions of Lemma 6.3, it holds for all* $\lambda \geq \overline{\lambda}_5 := \overline{\lambda}_5(\mathcal{S}, \overline{\mathcal{S}})$ *that*

$$\mathsf{dist}_{\lambda,T}\left(\mathcal{T}(X^\eta), \mathcal{T}^\eta(X^\eta)\right) \leq h(\eta) \tag{289}$$

for some nondecreasing continuous function $h(\eta)$ *with* $h(0) = 0$*. Here the function* h *only depends on the spaces* $\mathcal{S}$ *and* $\overline{\mathcal{S}}$*.*

Substituting Lemma C.4 and Eq. (288) into Eq. (287) yields

$$\mathsf{dist}_{\lambda,T}\left(X, X^\eta\right) \leq 2h(\eta) \to 0 \tag{290}$$

as $\eta \to 0$. The following lemma estabilishes if X and X^η are close and the step size η is small, we can couple θ^t and θ^t_η such that their (λ, T)-distance is small. A proof can be found in Appendix C.7.

Lemma C.5. *Under the same conditions of Lemma 6.3, for all* $\lambda \geq \overline{\lambda}_6 := \overline{\lambda}_6(\mathcal{S}, \overline{\mathcal{S}})$ *we can find a coupling for* θ^t *and* θ^t_η *such that*

$$\sup_{0\leq t\leq T} e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\theta^t - \theta^t_\eta\right\|_2^2\right]} \leq H(\eta, \mathsf{dist}_{\lambda,T}\left(X, X^\eta\right)), \tag{291}$$

where H *is a nondecreasing function in each coordinate and* $\lim_{(u,v)\to(0,0)} H(u,v) \to 0$*. Here the function* H *only depends on the spaces* $\mathcal{S}$ *and* $\overline{\mathcal{S}}$*.*

By coupling θ^t and θ^t_η as in Lemma C.5, we can then conclude the proof by invoking Lemma C.4 and Lemma C.5 since

$$\mathrm{W}_2\left(\mu_{\theta^{\tau_1}_\eta,\cdots,\theta^{\tau_m}_\eta}, \mu_{\theta^{\tau_1},\cdots,\theta^{\tau_m}}\right) \leq \sqrt{\frac{1}{m}\sum_{j=1}^m \left\|\theta^{\tau_j}_\eta - \theta^{\tau_j}\right\|_2^2} \leq e^{\lambda T}\cdot H(\eta, 2h(\eta)). \tag{292}$$

The proof is completed by taking $\eta \to 0$.

C.4 Proof of Lemma C.1

First we show any solution of $\mathfrak{S}^\eta$ must be uniquely determined by its values at discrete time knots $t_i = i\eta$ for $i = 0, 1, \cdots$. From Eqs. (113f) and (113g) we have u^t_η and w^t_η must be piecewise constant, namely

$$u^t_\eta = u^{\lfloor t\rfloor}_\eta, \qquad w^t_\eta = w^{\lfloor t\rfloor}_\eta, \tag{293}$$

and therefore we have θ^t_η is piecewise linear with time knots t_i and r^t_η is piecewise constant with time knots t_i. Finally, from Eqs. (114a) and (114b) we have $\partial\theta^t_\eta/\partial u^s_\eta$ is piecewise linear and $\partial\ell_{\lfloor t\rfloor}(r^t_\eta; z)/\partial w^s_\eta$ is piecewise constant with time knots t_i, this then implies R^η_θ is piecewise linear and R^η_ℓ is piecewise constant with knots t_i. By Eq. (113e) we have Γ^t_η must be piecewise constant with knots t_i. We then conclude that $\mathfrak{S}^\eta$ is uniquely determined at $t_i = i\eta$.

We show by induction the unique solution of $\mathfrak{S}^\eta$ at discrete time knots $t_i = i\eta$ must be

$$(\theta^{t_0}_\eta, \cdots, \theta^{t_r}_\eta) \stackrel{d}{=} (\overline{\theta}^{t_0}_\eta, \cdots, \overline{\theta}^{t_r}_\eta), \tag{294a}$$

$$R^\eta_\theta(t_i, t_j) = \zeta_{i,j-1}/\eta, \qquad 0 \leq j \leq i \leq r, \tag{294b}$$

$$R^\eta_\ell(t_i, t_j) = \xi_{i,j}/\eta, \qquad 0 \leq j < i \leq r, \tag{294c}$$

$$\Gamma^{t_i}_\eta = \xi_{i,i}, \qquad 0 \leq i \leq r, \tag{294d}$$

where we define

$$\zeta_{i,-1} = \mathbb{E}\left[\frac{\partial}{\partial\left(\theta^0/\eta\right)} g_i(\overline{u}^{t_1}_\eta, \cdots, \overline{u}^{t_i}_\eta; \theta^0)\right]. \tag{295}$$

For $r = 0$, provided that $\theta^0_\eta \stackrel{d}{=} \overline{\theta}^0_\eta \stackrel{d}{=} \theta^0$, it follows immediately that $w^0_\eta \stackrel{d}{=} \overline{w}^0_\eta \stackrel{d}{=} \mathsf{N}\left(0, \mathbb{E}\left[\theta^0\left(\theta^0\right)^{\mathsf{T}}\right]\right)$ and therefore $\Gamma^0_\eta = \mathbb{E}\left[\nabla_r \ell_0(r^0_\eta; z)\right] = \mathbb{E}\left[\frac{\partial}{\partial \overline{w}^0_\eta} f_0(\overline{w}^0_\eta; z)\right] = \xi_{0,0}$. Suppose the induction hypothesis holds for r, we next show Eqs. (294a) to (294d) hold for $r + 1$.

Induction on Eq. (294a). First, by Eqs. (282b) and (113f) we have

$$\mathbb{E}\left[\overline{w}_\eta^{t_i}\left(\overline{w}_\eta^{t_j}\right)^\top\right] = \mathbb{E}\left[g_i(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_i};\theta^0)g_j(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_j};\theta^0)^\top\right] = \mathbb{E}\left[\overline{\theta}_\eta^{t_i}\left(\overline{\theta}_\eta^{t_j}\right)^\top\right] = \mathbb{E}\left[\theta_\eta^{t_i}\left(\theta_\eta^{t_j}\right)^\top\right] = C_\theta^\eta(t_i,t_j)\,, \tag{296}$$

which implies $(\overline{w}_\eta^{t_0},\cdots,\overline{w}_\eta^{t_r}) \stackrel{d}{=} (w_\eta^{t_0},\cdots,w_\eta^{t_r})$. Similarly it also holds $(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_{r+1}}) \stackrel{d}{=} (u_\eta^{t_0},\cdots,u_\eta^{t_r})$. Thus, substituting into Eq. (113a) gives us for $t\in[t_r,t_{r+1})$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\theta_\eta^t &= -(\Lambda^{\lfloor t\rfloor}+\Gamma_\eta^{\lfloor t\rfloor})\theta_\eta^{\lfloor t\rfloor} - \int_0^{\lfloor t\rfloor} R_\ell^\eta(\lfloor t\rfloor,\lfloor s\rfloor)\theta_\eta^{\lfloor s\rfloor}\mathrm{d}s + u_\eta^t \\
&= -(\Lambda^{t_r}+\Gamma_\eta^{t_r})\theta_\eta^{t_r} - \int_0^{t_r} R_\ell^\eta(\lfloor t\rfloor,\lfloor s\rfloor)\theta_\eta^{\lfloor s\rfloor}\mathrm{d}s + u_\eta^{t_r} \\
&= -(\Lambda^{t_r}+\xi_{r,r})\theta_\eta^{t_r} - \sum_{j=0}^{r-1}\xi_{r,j}\theta_\eta^{t_j} + u_\eta^{t_r}\,,
\end{aligned} \tag{297}$$

and further

$$\theta_\eta^{t_{r+1}} = (I-\eta\Lambda^{t_r})\theta_\eta^{t_r} + \eta\left(u_\eta^{t_r} - \sum_{j=0}^{r}\xi_{r,j}\theta_\eta^{t_j}\right)\,. \tag{298}$$

Comparing to Eq. (280) which asserts

$$\overline{\theta}_\eta^{t_{r+1}} = (I-\eta\Lambda^{t_r})\overline{\theta}_\eta^{t_r} + \eta\left(\overline{u}_\eta^{t_{r+1}} - \sum_{j=0}^{r}\xi_{r,j}\overline{\theta}_\eta^{t_j}\right)\,, \tag{299}$$

which immediately implies Eq. (294a) holds for $r+1$.

Induction on Eq. (294b). With the same calculations applied to Eq. (114a), for an $0\le i\le r$ it follows that

$$\begin{aligned}
\frac{\partial\theta_\eta^{t_{r+1}}}{\partial u_\eta^{t_i}} &= (I-\eta\Lambda^{t_r})\frac{\partial\theta_\eta^{t_r}}{\partial u_\eta^{t_i}} - \eta\sum_{j=i}^{r}\xi_{r,j}\frac{\partial\theta_\eta^{t_j}}{\partial u_\eta^{t_i}} \\
&= (I-\eta\Lambda^{t_r})\frac{\partial\theta_\eta^{t_r}}{\partial u_\eta^{t_i}} - \eta\sum_{j=i+1}^{r}\xi_{r,j}\frac{\partial\theta_\eta^{t_j}}{\partial u_\eta^{t_i}} - \eta\xi_{r,i}\,,
\end{aligned} \tag{300}$$

where we use $\partial\theta_\eta^{t_i}/\partial u_\eta^{t_i} = I$. We slightly abuse the notation here by taking $\overline{u}_\eta^{t_0} := \theta^0/\eta$, and as a direct consequence of Eq. (280), we get the recursion when $0\le i\le r$,

$$\begin{aligned}
&\frac{\partial}{\partial\overline{u}_\eta^{t_i}}g_{r+1}(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_{r+1}};\theta^0) = \frac{\partial\overline{\theta}_\eta^{t_{r+1}}}{\partial\overline{u}_\eta^{t_i}} \\
&= (I-\eta\Lambda^{t_r})\frac{\partial}{\partial\overline{u}_\eta^{t_i}}g_r(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_r};\theta^0) - \eta\sum_{j=i+1}^{r}\xi_{r,j}\frac{\partial}{\partial\overline{u}_\eta^{t_i}}g_j(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_j};\theta^0) \\
&= (I-\eta\Lambda^{t_r})\frac{\partial}{\partial\overline{u}_\eta^{t_i}}g_r(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_r};\theta^0) - \eta\sum_{j=i+1}^{r}\xi_{r,j}\frac{\partial}{\partial\overline{u}_\eta^{t_i}}g_j(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_j};\theta^0) - \eta^2\xi_{r,i}\,,
\end{aligned} \tag{301}$$

where in the last line it is used that $\partial g_i(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_i};\theta^0)/\partial\overline{u}_\eta^{t_i} = \eta I$. We thus have

$$\left(\eta^{-1}\frac{\partial}{\partial\overline{u}_\eta^{t_i}}g_{r+1}(\overline{u}_\eta^{t_1},\cdots,\overline{u}_\eta^{t_{r+1}};\theta^0)\right)$$

$$= (I - \eta\Lambda^{t_r}) \left(\eta^{-1}\frac{\partial}{\partial \overline{u}_\eta^{t_i}} g_r(\overline{u}_\eta^{t_1}, \cdots, \overline{u}_\eta^{t_r}; \theta^0)\right) - \eta \sum_{j=i+1}^{r} \xi_{r,j} \left(\eta^{-1}\frac{\partial}{\partial \overline{u}_\eta^{t_i}} g_j(\overline{u}_\eta^{t_1}, \cdots, \overline{u}_\eta^{t_j}; \theta^0)\right) - \eta\xi_{r,i}\,. \tag{302}$$

Together with the induction hypothesis we then show for all $0 \le i \le r+1$,

$$R_\theta^\eta(t_{r+1}, t_i) = \mathbb{E}\left[\frac{\partial \theta_\eta^{t_{r+1}}}{\partial u_\eta^{t_i}}\right] = \mathbb{E}\left[\eta^{-1}\frac{\partial}{\partial \overline{u}_\eta^{t_i}} g_{r+1}(\overline{u}_\eta^{t_1}, \cdots, \overline{u}_\eta^{t_{r+1}}; \theta^0)\right] = \zeta_{r+1,i-1}/\eta\,. \tag{303}$$

Induction on Eqs. (294c) **and** (294d)**.** By Eq. (113b), for all $0 \le i \le r+1$,

$$\begin{aligned} r_\eta^{t_i} &= -\frac{1}{\delta}\int_0^{t_i} R_\theta^\eta(\lfloor t \rfloor, \lceil s \rceil)\ell_{\lfloor s \rfloor}(r_\eta^s; z)\mathrm{d}s + w_\eta^{t_i} \\ &= -\frac{1}{\delta}\sum_{j=0}^{i-1} \eta R_\theta^\eta(t_i, t_{j+1})\ell_{t_j}(r_\eta^{t_j}; z) + w_\eta^{t_i} \\ &= -\frac{1}{\delta}\sum_{j=0}^{i-1} \zeta_{i,j}\ell_{t_j}(r_\eta^{t_j}; z) + w_\eta^{t_i}\,, \end{aligned} \tag{304}$$

which further gives

$$\ell_{t_i}(r_\eta^{t_i}; z) = \ell_{t_i}\left(-\frac{1}{\delta}\sum_{j=0}^{i-1} \zeta_{i,j}\ell_{t_j}(r_\eta^{t_j}; z) + w_\eta^{t_i}; z\right)\,. \tag{305}$$

From Eq. (281) we get similarly

$$f_i(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_i}; z) = \ell_{t_i}\left(-\frac{1}{\delta}\sum_{j=0}^{i-1} \zeta_{i,j} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) + \overline{w}_\eta^{t_i}; z\right)\,. \tag{306}$$

Since Eq. (294a) holds for $r+1$, this implies we can assume without loss of generality that $(w_\eta^{t_0}, \cdots, w_\eta^{t_{r+1}}) = (\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_{r+1}})$. In this case, it always holds that $f_i(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_i}; z) = \ell_{t_i}(r_\eta^{t_i}; z)$ for $0 \le i \le r+1$. In particular

$$\begin{aligned} \nabla_r \ell_{t_i}(r_\eta^{t_i}; z) &= \nabla_r \ell_{t_i}\left(-\frac{1}{\delta}\sum_{j=0}^{i-1} \zeta_{i,j}\ell_{t_j}(r_\eta^{t_j}; z) + w_\eta^{t_i}; z\right) \\ &= \nabla_r \ell_{t_i}\left(-\frac{1}{\delta}\sum_{j=0}^{i-1} \zeta_{i,j} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) + \overline{w}_\eta^{t_i}; z\right) \\ &= \frac{\partial}{\partial \overline{w}_\eta^{t_i}} \ell_{t_i}\left(-\frac{1}{\delta}\sum_{j=0}^{i-1} \zeta_{i,j} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) + \overline{w}_\eta^{t_i}; z\right) \\ &= \frac{\partial}{\partial \overline{w}_\eta^{t_i}} f_i(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_i}; z)\,. \end{aligned} \tag{307}$$

Taking expectation on both sides and we obtain $\Gamma_\eta^{t_i} = \xi_{i,i}$ for $0 \le i \le r+1$. It then only remains to be shown that Eq. (294c) holds for $r+1$. From Eq. (114b), we have for all $0 \le i \le r$,

$$\begin{aligned} &\frac{\partial \ell_{t_{r+1}}(r_\eta^{t_{r+1}}; z)}{\partial w_\eta^{t_i}} \\ &= \nabla_r \ell_{t_{r+1}}(r_\eta^{t_{r+1}}; z) \cdot \left(-\frac{1}{\delta}\int_{t_{i+1}}^{t_{r+1}} R_\theta^\eta(t_{r+1}, \lceil s' \rceil)\frac{\partial \ell_{\lfloor s' \rfloor}(r_\eta^{s'}; z)}{\partial w_\eta^{t_i}}\mathrm{d}s' - \frac{1}{\delta}R_\theta^\eta(t_{r+1}, t_{i+1})\nabla_r \ell_{t_i}(r_\eta^{t_i}; z)\right) \end{aligned}$$

$$= \nabla_r \ell_{t_{r+1}}(r_\eta^{t_{r+1}}; z) \cdot \left(-\frac{1}{\delta} \sum_{j=i+1}^{r} \eta R_\theta^\eta(t_{r+1}, t_{j+1}) \frac{\partial \ell_{t_j}(r_\eta^{t_j}; z)}{\partial w_\eta^{t_i}} - \frac{1}{\delta} R_\theta^\eta(t_{r+1}, t_{i+1}) \nabla_r \ell_{t_i}(r_\eta^{t_i}; z) \right)$$

$$= \nabla_r \ell_{t_{r+1}}(r_\eta^{t_{r+1}}; z) \cdot \left(-\frac{1}{\delta} \sum_{j=i+1}^{r} \zeta_{r+1,j} \frac{\partial \ell_{t_j}(r_\eta^{t_j}; z)}{\partial w_\eta^{t_i}} - \frac{1}{\delta} \zeta_{r+1,i} \nabla_r \ell_{t_i}(r_\eta^{t_i}; z)/\eta \right) . \tag{308}$$

Since

$$\frac{\partial}{\partial \overline{w}_\eta^{t_i}} f_{r+1}(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_{r+1}}; z)$$

$$= \frac{\partial}{\partial \overline{w}_\eta^{t_i}} \ell_{t_{r+1}} \left(-\frac{1}{\delta} \sum_{j=0}^{r} \zeta_{r+1,j} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) + \overline{w}_\eta^{t_{r+1}}; z \right)$$

$$= \nabla_r \ell_{t_{r+1}} \left(-\frac{1}{\delta} \sum_{j=0}^{r} \zeta_{r+1,j} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) + \overline{w}_\eta^{t_{r+1}}; z \right) \cdot \left(-\frac{1}{\delta} \sum_{j=i}^{r} \zeta_{r+1,j} \frac{\partial}{\partial \overline{w}_\eta^{t_i}} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) \right)$$

$$= \nabla_r \ell_{t_{r+1}}(r_\eta^{t_{r+1}}; z) \cdot \left(-\frac{1}{\delta} \sum_{j=i}^{r} \zeta_{r+1,j} \frac{\partial}{\partial \overline{w}_\eta^{t_i}} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) \right)$$

$$= \nabla_r \ell_{t_{r+1}}(r_\eta^{t_{r+1}}; z) \cdot \left(-\frac{1}{\delta} \sum_{j=i+1}^{r} \zeta_{r+1,j} \frac{\partial}{\partial \overline{w}_\eta^{t_i}} f_j(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_j}; z) - \frac{1}{\delta} \zeta_{r+1,i} \nabla_r \ell_{t_i}(r_\eta^{t_i}; z) \right) , \tag{309}$$

where in the last line we use $\partial f_i(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_i}; z)/\partial \overline{w}_\eta^{t_i} = \nabla_r \ell_{t_i}(r_\eta^{t_i}; z)$. Comparing the above two equations, it then follows that

$$\frac{\partial \ell_{t_{r+1}}(r_\eta^{t_{r+1}}; z)}{\partial w_\eta^{t_i}} = \eta^{-1} \frac{\partial}{\partial \overline{w}_\eta^{t_i}} f_{r+1}(\overline{w}_\eta^{t_0}, \cdots, \overline{w}_\eta^{t_{r+1}}; z) , \tag{310}$$

which further implies Eq. (294c) for $r+1$ by taking expectation on both sides. This concludes the induction.
Finally, we invoke Lemma C.3 to show the solution of $\mathfrak{S}^\eta$ is in the space $\mathcal{S}$.

C.5 Proof of Lemma C.3

Since the covariance kernels C_ℓ^η and C_θ^η are piecewise constant, the continuity conditions are automatically satisfied by Definition 5.2 and 5.3. To show the lemma we only need to prove the upper bounds

$$\|R_\theta^\eta(t,s)\| \le \Phi_{R_\theta}(t-s) , \qquad \|R_\ell^\eta(t,s)\| \le \mathbb{E}\left[\left\| \frac{\partial \ell_{\lfloor t \rfloor}(r_\eta^t; z)}{\partial w_\eta^s} \right\| \right] \le \Phi_{R_\ell}(t-s) , \qquad 0 \le s \le t \le T , \tag{311a}$$

$$\|C_\theta^\eta(t,t)\| \le \Phi_{C_\theta}(t) , \qquad \|C_\ell^\eta(t,t)\| \le \Phi_{C_\ell}(t) , \qquad 0 \le t \le T , \tag{311b}$$

$$\left\| \Gamma_\eta^t \right\| \le M_\ell , \qquad 0 \le t \le T . \tag{311c}$$

Note that

$$\left\| \Gamma_\eta^t \right\| \le \mathbb{E}\left[\left\| \nabla_r \ell_{\lfloor t \rfloor}(r_\eta^t; z) \right\| \right] \le M_\ell , \tag{312}$$

which proves Eq. (311c).

Upper bounds for R_θ^η and R_ℓ^η. From the definition of R_ℓ^η in Eq. (113d) and that r_η^t and w_η^t are piecewise constant, we know $R_\ell^\eta(t,s) = R_\ell^\eta(\lfloor t \rfloor, \lfloor s \rfloor)$. Since $\max\{\lfloor t \rfloor - \lceil s \rceil, 0\} \le t - s$, it suffices to prove $\|R_\ell^\eta(\lfloor t \rfloor, \lfloor s \rfloor)\| \le \Phi_{R_\ell}(\max\{\lfloor t \rfloor - \lceil s \rceil, 0\})$. We prove Eq. (311a) for all $\lfloor t \rfloor - \lfloor s \rfloor \le m\eta$ for all $m \in \mathbb{Z}_{\ge 0}$. When $m = 0$, we have

$$\frac{\mathrm{d}}{\mathrm{d}t} \|R_\theta^\eta(t,s)\| \le (M_\Lambda + M_\ell) \|R_\theta^\eta(\lfloor t \rfloor, s)\| \le M_\Lambda + M_\ell , \tag{313}$$

$$\|R_\ell^\eta(t,s)\| = \left\| -\frac{1}{\delta}\mathbb{E}\left[\nabla_r \ell_{\lfloor t\rfloor}(r_\eta^t;z)\nabla_r \ell_{\lfloor t\rfloor}(r_\eta^s;z)\right]\right\| \le \frac{M_\ell^2}{\delta^2}\,. \tag{314}$$

Comparing to Eqs. (58b) and (58a) we see $\|R_\theta^\eta(t,s)\| \le \Phi_{R_\theta}(t-s), \|\Phi_{R_\ell}(t,s)\| \le \Phi_{R_\ell}(0)$ when $\lfloor t\rfloor = \lfloor s\rfloor$. Suppose Eq. (311a) holds for $\lfloor t\rfloor - \lfloor s\rfloor \le m\eta$, then by Eq. (114a) one has for $\lfloor t\rfloor - \lfloor s\rfloor = (m+1)\eta$,

$$\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\|R_\theta^\eta(t,s)\| &\le (M_\Lambda + M_\ell)\|R_\theta^\eta(\lfloor t\rfloor,s)\| + \int_s^{\lfloor t\rfloor}\|R_\ell^\eta(\lfloor t\rfloor,\lfloor s'\rfloor)\|\|R_\theta^\eta(\lfloor s'\rfloor,s)\|\,\mathrm{d}s' \\
&\overset{\text{(i)}}{=} \lim_{t'\uparrow\lfloor t\rfloor}\left\{(M_\Lambda+M_\ell)\|R_\theta^\eta(t',s)\| + \int_s^{t'}\|R_\ell^\eta(\lfloor t\rfloor,\lfloor s'\rfloor)\|\|R_\theta^\eta(\lfloor s'\rfloor,s)\|\right\}\mathrm{d}s' \\
&\le (M_\Lambda+M_\ell)\Phi_{R_\theta}(t-s) + \int_s^{\lfloor t\rfloor}\Phi_{R_\ell}(t-s')\Phi_{R_\theta}(s'-s)\mathrm{d}s' \\
&\overset{\text{(ii)}}{\le} \frac{\mathrm{d}}{\mathrm{d}t}\Phi_{R_\theta}(t-s)\,,
\end{aligned} \tag{315}$$

where in (i) we use that $R_\theta^\eta(t,s)$ is continuous in t and in (ii) we use Eq. (58a). We conclude $\|R_\theta^\eta(t,s)\| \le \Phi_{R_\theta}(t-s)$ when $\lfloor t\rfloor - \lfloor s\rfloor \le (m+1)\eta$. Similarly by Eq. (114b) when $\lfloor t\rfloor - \lfloor s\rfloor = (m+1)\eta$ it holds

$$\begin{aligned}
\mathbb{E}\left[\left\|\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t;z)}{\partial w_\eta^s}\right\|\right] &\le M_\ell\cdot\left(\frac{1}{\delta}\int_{\lceil s\rceil}^{\lfloor t\rfloor} R_\theta^\eta(\lfloor t\rfloor,\lceil s'\rceil)\mathbb{E}\left[\left\|\frac{\partial \ell_{\lfloor s'\rfloor}(r_\eta^{s'};z)}{\partial w_\eta^s}\right\|\right]\mathrm{d}s' + \frac{M_\ell}{\delta}\cdot R_\theta^\eta(\lfloor t\rfloor,\lceil s\rceil)\right) \\
&\overset{\text{(i)}}{\le} \frac{M_\ell}{\delta}\cdot\left\{M_\ell\Phi_{R_\theta}(t) + \int_{\lceil s\rceil}^{\lfloor t\rfloor}\Phi_{R_\theta}(t-s')\Phi_{R_\ell}(s'-s)\mathrm{d}s'\right\} \\
&\overset{\text{(ii)}}{\le} \Phi_{R_\ell}(t-s)\,,
\end{aligned} \tag{316}$$

where we use $\lfloor t\rfloor - \lceil s\rceil \le m\eta$ and the induction hypothesis at m in (i), in (ii) we invoke Eq. (58b). Finally, note that $\|R_\ell^\eta(t,s)\| \le \mathbb{E}\left[\left\|\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t;z)}{\partial w_\eta^s}\right\|\right]$ holds and we complete the proof by induction.

Upper bounds for C_θ^η and C_ℓ^η. Since C_θ^η and C_ℓ^η are piecewise constant we only need to show $\|C_\theta^\eta(\lfloor t\rfloor,\lfloor t\rfloor)\| \le \Phi_{C_\theta}(\lfloor t\rfloor)$ and $\|C_\ell^\eta(\lfloor t\rfloor,\lfloor t\rfloor)\| \le \Phi_{C_\ell}(\lfloor t\rfloor,\lfloor t\rfloor)$ and Eq. (311b) will follow by monotonicity of Φ_{C_θ} and Φ_{C_ℓ}. We show this by induction on $\lfloor t\rfloor = r\eta$ with hypotheses

$$\mathbb{E}\left[\left\|\theta_\eta^{\lfloor t\rfloor}\right\|^2\right] \le \Phi_{C_\theta}(\lfloor t\rfloor)\,, \tag{317}$$

When $r=0$, the initial condition holds at time 0. Suppose the inductive hypotheses hold for $\lfloor t\rfloor \le r\eta$, when $r\eta \le t < (r+1)\eta$, we can obtain from Eq. (113a) that

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\theta_\eta^t\|_2 = (M_\Lambda+M_\ell)\left\|\theta_\eta^{\lfloor t\rfloor}\right\|_2 + \int_0^{\lfloor t\rfloor}\Phi_{R_\ell}(t-s)\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2\mathrm{d}s + \left\|u_\eta^{\lfloor t\rfloor}\right\|_2\,. \tag{318}$$

By the same calculations in Eq. (162) we get

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\sqrt{\mathbb{E}\left[\|\theta_\eta^t\|_2^2\right]} \\
&\le \sqrt{3\cdot\left\{(M_\Lambda+M_\ell)^2\,\mathbb{E}\left[\left\|\theta_\eta^{\lfloor t\rfloor}\right\|_2^2\right] + \int_0^{\lfloor t\rfloor}(t-s+1)^2\Phi_{R_\ell}(t-s)^2\mathbb{E}\left[\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2^2\right]\mathrm{d}s + \frac{k}{\delta}\Phi_{C_\ell}(t)\right\}} \\
&\le \sqrt{3\cdot\left\{(M_\Lambda+M_\ell)^2\,\Phi_{C_\theta}(t)^2 + \int_0^{\lfloor t\rfloor}(t-s+1)^2\Phi_{R_\ell}(t-s)^2\Phi_{C_\theta}(s)^2\mathrm{d}s + \frac{k}{\delta}\Phi_{C_\ell}(t)\right\}}\,,
\end{aligned} \tag{319}$$

which together with Eq. (58c) implies $\sqrt{\mathbb{E}\left[\|\theta_\eta^t\|_2^2\right]} \le \Phi_{C_\theta}(t)$ when $t \le (r+1)\eta$. It then follows that

$$\|C_\theta^\eta(t,t)\| = \left\|\mathbb{E}\left[\theta_\eta^t {\theta_\eta^t}^\top\right]\right\| \le \mathbb{E}\left[\|\theta_\eta^t\|_2^2\right] \le \Phi_{C_\theta}(t)\,. \tag{320}$$

By Eq. (113b) we have

$$\begin{aligned}\left\|\ell_{\lfloor t\rfloor}\left(r_\eta^t; z\right)\right\|_2 &\le \left\|\ell_{\lfloor t\rfloor}(0;z)\right\|_2 + M_\ell \left\|r_\eta^t\right\|_2 \\ &\le \left\|\ell_{\lfloor t\rfloor}(0;z)\right\|_2 + \frac{M_\ell}{\delta}\int_0^{\lfloor t\rfloor} \|R_\theta^\eta(\lfloor t\rfloor, \lceil s\rceil)\| \left\|\ell_{\lfloor s\rfloor}\left(r_\eta^s; z\right)\right\|_2 \mathrm{d}s + M_\ell \left\|w_\eta^t\right\|_2 \\ &\le \left\|\ell_{\lfloor t\rfloor}(0;z)\right\|_2 + \frac{M_\ell}{\delta}\int_0^{\lfloor t\rfloor} \Phi_{R_\theta}(t-s)\left\|\ell_{\lfloor s\rfloor}\left(r_\eta^{\lfloor s\rfloor}; z\right)\right\|_2 \mathrm{d}s + M_\ell \left\|w_\eta^t\right\|_2\,, \end{aligned} \tag{321}$$

where in the last line we use the fact that r_η^t is piecewise constant. Repeat the same argument in Eq. (172), we obtain

$$\begin{aligned}&\mathbb{E}\left[\left\|\ell_{\lfloor t\rfloor}\left(r_\eta^t; z\right)\right\|_2^2\right] \\ &\le 3\left\{\mathbb{E}\left[\left\|\ell_{\lfloor t\rfloor}(0;z)\right\|_2^2\right] + \frac{M_\ell^2}{\delta^2}\int_0^{\lfloor t\rfloor}(t-s+1)^2\Phi_{R_\theta}(t-s)^2\mathbb{E}\left[\left\|\ell_{\lfloor s\rfloor}\left(r_\eta^{\lfloor s\rfloor}; z\right)\right\|_2^2\right]\mathrm{d}s + M_\ell^2\mathbb{E}\left[\left\|w_\eta^t\right\|_2^2\right]\right\} \\ &\le 3\left\{M_{\theta^0,z} + \frac{M_\ell^2}{\delta^2}\int_0^{\lfloor t\rfloor}(t-s+1)^2\Phi_{R_\theta}(t-s)^2\Phi_{C_\ell}(s)\mathrm{d}s + kM_\ell^2\Phi_{C_\theta}(t)\right\}\,, \end{aligned} \tag{322}$$

where we use $\mathbb{E}\left[\left\|w_\eta^t\right\|_2^2\right] \le k\left\|C_\theta^\eta(t,t)\right\| \le k\Phi_{C_\theta}(t)$. Comparing to Eq. (58d) we have $\mathbb{E}\left[\left\|\ell_{\lfloor t\rfloor}\left(r_\eta^t; z\right)\right\|_2^2\right] \le \Phi_{C_\ell}(t)$ and

$$\|C_\ell^\eta(t,t)\| = \left\|\mathbb{E}\left[\ell_{\lfloor t\rfloor}(r_\eta^t;z)\ell_{\lfloor t\rfloor}(r_\eta^t;z)^\top\right]\right\| \le \mathbb{E}\left[\left\|\ell_{\lfloor t\rfloor}\left(r_\eta^t; z\right)\right\|_2^2\right] \le \Phi_{C_\ell}(t)\,. \tag{323}$$

We conclude the proof by induction.

C.6 Proof of Lemma C.4

We first introduce a lemma for mappings $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}$ and $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}^\eta$ on X^η. The reader can find its proof in Appendix C.8.

Lemma C.6. *Under the same conditions of Lemma 6.3, suppose* $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}(X^\eta) = (\overline{C}_\theta^\eta, \overline{R}_\theta^\eta)$, $\mathcal{T}_{\mathcal{S}\to\overline{\mathcal{S}}}^\eta(X^\eta) = (C_\theta^\eta, R_\theta^\eta)$ *and define* $[R_\theta^\eta](t,s) := R_\theta^\eta(\lfloor t\rfloor, \lceil s\rceil)$ *when* $\lceil s\rceil \le \lfloor t\rfloor$, $[R_\theta^\eta](t,s) := I$ *when* $\lceil s\rceil > \lfloor t\rfloor$. *It then holds for all* $\lambda \ge \overline{\lambda}_6 := \overline{\lambda}_6(\mathcal{S}, \overline{\mathcal{S}})$.

$$\mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^\eta, C_\theta^\eta\right) \le \overline{h}(\eta)\,, \tag{324}$$

$$\mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^\eta, [R_\theta^\eta]\right) \le \overline{h}(\eta)\,, \tag{325}$$

for some nondecreasing function $\overline{h}(\eta)$ *with* $\overline{h}(0) = 0$. *Here the function* $\overline{h}$ *only depends on the spaces* $\mathcal{S}$ *and* $\overline{\mathcal{S}}$.

Suppose $\mathcal{T}(X^\eta) = (\overline{C}_\ell^\eta, \overline{R}_\ell^\eta, \overline{\Gamma}_\eta)$ and the fixed point equation $\mathcal{T}^\eta(X^\eta) = X^\eta = (C_\ell^\eta, R_\ell^\eta, \Gamma_\eta)$. Using the same notations in Lemma C.6 we can then write out the equations determining $(\overline{C}_\ell^\eta, \overline{R}_\ell^\eta, \overline{\Gamma}_\eta)$ and $(C_\ell^\eta, R_\ell^\eta, \Gamma_\eta)$ as

$$\overline{r}_\eta^t = -\frac{1}{\delta}\int_0^t \overline{R}_\theta^\eta(t,s)\ell_s(\overline{r}_\eta^s; z)\mathrm{d}s + \overline{w}_\eta^t\,, \qquad w^t \sim \mathsf{GP}(0, \overline{C}_\theta^\eta)\,,$$

$$\frac{\partial \ell_t(\overline{r}_\eta^t; z)}{\partial \overline{w}_\eta^s} = \nabla_r \ell_t(\overline{r}_\eta^t; z)\cdot\left(-\frac{1}{\delta}\int_s^t \overline{R}_\theta^\eta(t,s')\frac{\partial \ell_{s'}(\overline{r}_\eta^{s'}; z)}{\partial \overline{w}_\eta^s}\mathrm{d}s' - \frac{1}{\delta}\overline{R}_\theta^\eta(t,s)\nabla_r\ell_s(\overline{r}_\eta^s; z)\right)\,, \quad 0 \le s < t \le T\,,$$

$$\overline{C}_\ell^\eta(t,s) = \mathbb{E}\left[\ell_t(\overline{r}_\eta^t;z)\ell_s(\overline{r}_\eta^s;z)^\mathsf{T}\right], \qquad 0 \le s \le t \le T,$$

$$\overline{R}_\ell^\eta(t,s) = \mathbb{E}\left[\frac{\partial \ell_t(\overline{r}_\eta^t;z)}{\partial \overline{w}_\eta^s}\right], \qquad 0 \le s < t \le T,$$

$$\overline{\Gamma}_\eta^t = \mathbb{E}\left[\nabla_r \ell_t(\overline{r}_\eta^t;z)\right], \qquad 0 \le t \le T,$$

and

$$r_\eta^t = -\frac{1}{\delta}\int_0^{\lfloor t\rfloor} \left[R_\theta^\eta\right](t,s)\ell_{\lfloor s\rfloor}(r_\eta^s;z)\mathrm{d}s + w_\eta^t,$$

$$\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t;z)}{\partial w_\eta^s} = \nabla_r \ell_{\lfloor t\rfloor}(r_\eta^t;z)\cdot\left(-\frac{1}{\delta}\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left[R_\theta^\eta\right](t,s')\frac{\partial \ell_{\lfloor s'\rfloor}(r_\eta^{s'};z)}{\partial w_\eta^s}\mathrm{d}s' - \frac{1}{\delta}\left[R_\theta^\eta\right](t,s)\nabla_r\ell_{\lfloor s\rfloor}(r_\eta^s;z)\right),$$

$$C_\ell^\eta(t,s) = \mathbb{E}\left[\ell_{\lfloor t\rfloor}(r_\eta^{\lfloor t\rfloor};z)\ell_{\lfloor s\rfloor}(r_\eta^{\lfloor s\rfloor};z)^\mathsf{T}\right],$$

$$R_\ell^\eta(t,s) = \mathbb{E}\left[\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t;z)}{\partial w_\eta^s}\right],$$

$$\Gamma_\eta^t = \mathbb{E}\left[\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right],$$

where $w_\eta^t \sim \mathsf{GP}(0, C_\theta^\eta)$. Note that since we set $\left[R_\theta^\eta\right](t,s) = I$ when $\lceil s\rceil > \lfloor t\rfloor$, it is consistent with the definition in Eq. (115).

Controlling the distance between $\overline{C}_\ell^\eta$ and C_ℓ^η. By Lemma C.6, we can couple the Gaussian processes $\overline{w}_\eta^t$ and w_η^t such that for all $\lambda \ge \overline{\lambda}_6$,

$$\sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\overline{w}_\eta^t - w_\eta^t\right\|_2^2\right]} \le 2\cdot \mathsf{dist}_{\lambda,T}\left(\overline{C}_\theta^\eta, C_\theta^\eta\right) \le 2\overline{h}(\eta). \tag{326}$$

By definition of C_θ^η we know w_η^t is piecewise constant in the sense that $w_\eta^t = w_\eta^{\lfloor t\rfloor}$. Hence $r_\eta^t = r_\eta^{\lfloor t\rfloor}$ and

$$\begin{aligned}
\left\|\overline{r}_\eta^t - r_\eta^t\right\|_2 &= \left\|-\frac{1}{\delta}\int_0^t \overline{R}_\theta^\eta(t,s)\ell_s(\overline{r}_\eta^s;z)\mathrm{d}s + \overline{w}_\eta^t + \frac{1}{\delta}\int_0^{\lfloor t\rfloor}\left[R_\theta^\eta\right](t,s)\ell_{\lfloor s\rfloor}(r_\eta^s;z)\mathrm{d}s - w_\eta^t\right\|_2 \\
&\le \frac{1}{\delta}\int_0^{\lfloor t\rfloor}\left\|\overline{R}_\theta^\eta(t,s)\right\|\left\|\ell_s(\overline{r}_\eta^s;z) - \ell_{\lfloor s\rfloor}(r_\eta^s;z)\right\|_2\mathrm{d}s + \frac{1}{\delta}\int_0^{\lfloor t\rfloor}\left\|\overline{R}_\theta^\eta(t,s) - \left[R_\theta^\eta\right](t,s)\right\|\left\|\ell_{\lfloor s\rfloor}(r_\eta^s;z)\right\|_2\mathrm{d}s \\
&\quad + \frac{1}{\delta}\int_{\lfloor t\rfloor}^t\left\|\overline{R}_\theta^\eta(t,s)\right\|\left\|\ell_s(\overline{r}_\eta^s;z)\right\|_2\mathrm{d}s + \left\|\overline{w}_\eta^t - w_\eta^t\right\|_2 \\
&\le \frac{M_\ell}{\delta}\int_0^{\lfloor t\rfloor}\Phi_{R_\theta}(t-s)\left(\left\|\overline{r}_\eta^s - r_\eta^s\right\|_2 + |s-\lfloor s\rfloor|\right)\mathrm{d}s + \frac{1}{\delta}\int_0^{\lfloor t\rfloor}\left\|\overline{R}_\theta^\eta(t,s) - \left[R_\theta^\eta\right](t,s)\right\|\left\|\ell_{\lfloor s\rfloor}(r_\eta^s;z)\right\|_2\mathrm{d}s \\
&\quad + \frac{\Phi_{R_\theta}(T)}{\delta}\int_{\lfloor t\rfloor}^t\left\|\ell_s(\overline{r}_\eta^s;z)\right\|_2\mathrm{d}s + \left\|\overline{w}_\eta^t - w_\eta^t\right\|_2.
\end{aligned} \tag{327}$$

We choose λ large enough such that Lemma C.6 holds and we can get

$$\begin{aligned}
&e^{-\lambda t}\left\|\overline{r}_\eta^t - r_\eta^t\right\|_2 \\
&\le \frac{M_\ell}{\delta}\int_0^{\lfloor t\rfloor} e^{-\lambda(t-s)}\Phi_{R_\theta}(t-s)\cdot e^{-\lambda s}\left(\left\|\overline{r}_\eta^s - r_\eta^s\right\|_2 + \eta\right)\mathrm{d}s + \frac{\overline{h}(\eta)}{\delta}\int_0^{\lfloor t\rfloor}\left\|\ell_{\lfloor s\rfloor}(r_\eta^s;z)\right\|_2\mathrm{d}s \\
&\quad + \frac{e^{-\lambda t}\Phi_{R_\theta}(T)}{\delta}\int_{\lfloor t\rfloor}^t\left\|\ell_s(\overline{r}_\eta^s;z)\right\|_2\mathrm{d}s + e^{-\lambda t}\left\|\overline{w}_\eta^t - w_\eta^t\right\|_2.
\end{aligned} \tag{328}$$

Square both sides and take expectations. It follows by Cauchy-Schwarz inequality that

$$e^{-2\lambda t}\mathbb{E}\left[\left\|\overline{r}_\eta^t - r_\eta^t\right\|_2^2\right]$$

$$
\begin{aligned}
&\le \left\{2\int_0^{\lfloor t\rfloor}(t-s+1)^{-2}\mathrm{d}s+\int_0^t(t-s+1)^{-2}+1\right\}\\
&\quad\cdot\left\{\frac{M_\ell^2}{\delta^2}\int_0^{\lfloor t\rfloor}e^{-2\lambda(t-s)}\Phi_{R_\theta}(t-s)^2\cdot e^{-2\lambda s}\left(\mathbb{E}\left[\left\|\overline{r}_\eta^s-r_\eta^s\right\|_2^2\right]+\eta^2\right)\mathrm{d}s\right.\\
&\quad\left.+\frac{\overline{h}(\eta)^2}{\delta^2}\int_0^{\lfloor t\rfloor}\mathbb{E}\left[\left\|\ell_{\lfloor s\rfloor}(r_\eta^s;z)\right\|_2^2\right]\mathrm{d}s+\frac{\Phi_{R_\theta}(T)^2}{\delta^2}\int_{\lfloor t\rfloor}^t\mathbb{E}\left[\left\|\ell_s(\overline{r}_\eta^s;z)\right\|_2^2\right]\mathrm{d}s+e^{-2\lambda t}\mathbb{E}\left[\left\|\overline{w}_\eta^t-w_\eta^t\right\|_2^2\right]\right\}\\
&\le 4\cdot\left\{\frac{M_\ell^2}{\delta^2}\int_0^{\lfloor t\rfloor}e^{-2\lambda(t-s)}\Phi_{R_\theta}(t-s)^2\cdot e^{-2\lambda s}\left(\mathbb{E}\left[\left\|\overline{r}_\eta^s-r_\eta^s\right\|_2^2\right]+\eta^2\right)\mathrm{d}s+\frac{\overline{h}(\eta)^2kT\Phi_{C_\ell}(T)}{\delta^2}\right.\\
&\quad\left.+\frac{\eta k\Phi_{C_\ell}(T)\Phi_{R_\theta}(T)^2}{\delta^2}+4\overline{h}(\eta)^2\right\},
\end{aligned}\tag{329}
$$

where in the last line we invoke Eq. (326) and use $\mathbb{E}\left[\left\|\ell_s(\overline{r}_\eta^s;z)\right\|_2^2\right]\le k\left\|\overline{C}_\ell(t,t)\right\|\le k\Phi_{C_\ell}(T)$. Next, we take λ large enough such that

$$
\frac{M_\ell^2}{\delta^2}\int_0^\infty e^{-2\lambda t}\Phi_{R_\theta}(t)^2\le\frac{1}{8},
$$

which will further induce that

$$
\begin{aligned}
&e^{-2\lambda t}\mathbb{E}\left[\left\|\overline{r}_\eta^t-r_\eta^t\right\|_2^2\right]\\
&\le\frac{1}{2}\sup_{0\le s\le t}e^{-2\lambda s}\mathbb{E}\left[\left\|\overline{r}_\eta^s-r_\eta^s\right\|_2^2\right]+\frac{1}{2}\eta^2+\frac{4\overline{h}(\eta)^2kT\Phi_{C_\ell}(T)}{\delta^2}+\frac{4\eta k\Phi_{C_\ell}(T)\Phi_{R_\theta}(T)^2}{\delta^2}+16\overline{h}(\eta)^2,
\end{aligned}\tag{330}
$$

and taking supremum over $t\in[0,T]$ on both sides

$$
\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\overline{r}_\eta^t-r_\eta^t\right\|_2^2\right]}\le\sqrt{\eta^2+\frac{8\overline{h}(\eta)^2kT\Phi_{C_\ell}(T)}{\delta^2}+\frac{8\eta k\Phi_{C_\ell}(T)\Phi_{R_\theta}(T)^2}{\delta^2}+32\overline{h}(\eta)^2}.\tag{331}
$$

Further following the same coupling argument in Appendix B.4.1, we obtain

$$
\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\overline{C}_\ell^\eta,C_\ell^\eta\right)&\le\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\ell_t(\overline{r}_\eta^t;z)-\ell_{\lfloor t\rfloor}(r_\eta^{\lfloor t\rfloor};z)\right\|_2^2\right]}\\
&=\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\ell_t(\overline{r}_\eta^t;z)-\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right\|_2^2\right]}\\
&\le M_\ell\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\left(\left\|\overline{r}_\eta^t-r_\eta^t\right\|_2+\eta\right)^2\right]}.
\end{aligned}\tag{332}
$$

By triangle inequality, we then get

$$
\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\overline{C}_\ell^\eta,C_\ell^\eta\right)&\le M_\ell\sup_{t\in[0,T]}e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\overline{r}_\eta^t-r_\eta^t\right\|_2^2\right]}+M_\ell\eta\\
&\le M_\ell\sqrt{\eta^2+\frac{8\overline{h}(\eta)^2kT\Phi_{C_\ell}(T)}{\delta^2}+\frac{8\eta k\Phi_{C_\ell}(T)\Phi_{R_\theta}(T)^2}{\delta^2}+32\overline{h}(\eta)^2}+M_\ell\eta\\
&=:h_1(\eta).
\end{aligned}\tag{333}
$$

Clearly $h_1(\eta)\to 0$ as $\eta\to 0$.

Controlling the distances between $\overline{R}_\ell^\eta$ and R_ℓ^η, $\overline{\Gamma}_\eta$ and Γ_η. First we consider the distance between $\overline{\Gamma}_\eta$ and Γ_η. By Eq. (333), we can get

$$\begin{aligned}\mathsf{dist}_{\lambda,T}\left(\overline{\Gamma}_\eta,\Gamma_\eta\right) &= \sup_{t\in[0,T]} e^{-\lambda t}\left\|\mathbb{E}\left[\nabla_r\ell_t(\overline{r}_\eta^t;z)\right]-\mathbb{E}\left[\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right]\right\| \\ &\leq \sup_{t\in[0,T]} e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r\ell_t(\overline{r}_\eta^t;z)-\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right\|\right] \\ &\leq M_\ell \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\overline{r}_\eta^t-r_\eta^t\right\|_2^2\right]}+M_\ell\eta \\ &\leq h_1(\eta)\,. \end{aligned} \tag{334}$$

Now we only need to bound the distance between $\overline{R}_\ell^\eta$ and R_ℓ^η. To this end, we introduce two auxiliary functions

$$\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s} := -\frac{1}{\delta}\int_s^t \overline{R}_\theta^\eta(t,s')\frac{\partial \ell_{s'}(\overline{r}_\eta^{s'};z)}{\partial \overline{w}_\eta^s}\mathrm{d}s' - \frac{1}{\delta}\overline{R}_\theta^\eta(t,s)\nabla_r\ell_s(\overline{r}_\eta^s;z)\,, \qquad 0\leq s<t\leq T\,, \tag{335}$$

$$\frac{\partial r_\eta^t}{\partial w_\eta^s} := -\frac{1}{\delta}\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left[R_\theta^\eta\right](t,s')\frac{\partial \ell_{\lfloor s'\rfloor}(r_\eta^{s'};z)}{\partial w_\eta^s}\mathrm{d}s' - \frac{1}{\delta}\left[R_\theta^\eta\right](t,s)\nabla_r\ell_{\lfloor s\rfloor}(r_\eta^s;z)\,, \qquad 0\leq s<t\leq T\,. \tag{336}$$

We can then write

$$\begin{aligned}\frac{\partial \ell_t(\overline{r}_\eta^t;z)}{\partial \overline{w}_\eta^s} &= \nabla_r\ell_t(\overline{r}_\eta^t;z)\cdot\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s}\,, \\ \frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t;z)}{\partial w_\eta^s} &= \nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\cdot\frac{\partial r_\eta^t}{\partial w_\eta^s}\,.\end{aligned}$$

Therefore, we can derive

$$\begin{aligned}&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(\overline{r}_\eta^t;z)}{\partial \overline{w}_\eta^s}-\frac{\partial \ell_{\lfloor t\rfloor}(r_\eta^t;z)}{\partial w_\eta^s}\right\|\right] \\ &\leq e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r\ell_t(\overline{r}_\eta^t;z)-\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right\|\left\|\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s}\right\|\right]+e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right\|\left\|\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s}-\frac{\partial r_\eta^t}{\partial w_\eta^s}\right\|\right]\,.\end{aligned} \tag{337}$$

Since $(\overline{C}_\theta^\eta,\overline{R}_\theta^\eta)\in\overline{\mathcal{S}}$, we are allowed to invoke Eq. (223) that helps us bound the first term

$$\begin{aligned}&e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r\ell_t(\overline{r}_\eta^t;z)-\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right\|\left\|\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s}\right\|\right] \\ &\leq \frac{M_\ell\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)\cdot e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r\ell_t(\overline{r}_\eta^t;z)-\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right\|\right] \\ &\leq \frac{M_\ell\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)\cdot h_1(\eta)\,,\end{aligned} \tag{338}$$

where in the last line we apply Eq. (334) when $t\in[0,T]$. By Lipschitz property in Assumption 1, we can upper bound the second term by

$$\begin{aligned}&e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r\ell_{\lfloor t\rfloor}(r_\eta^t;z)\right\|\left\|\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s}-\frac{\partial r_\eta^t}{\partial w_\eta^s}\right\|\right] \\ &\leq M_\ell\cdot e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s}-\frac{\partial r_\eta^t}{\partial w_\eta^s}\right\|\right] \\ &\leq \frac{M_\ell}{\delta}\cdot e^{-\lambda t}\cdot\left\{\mathbb{E}\left[\left\|\int_s^t\overline{R}_\theta^\eta(t,s')\frac{\partial \ell_{s'}(\overline{r}_\eta^{s'};z)}{\partial \overline{w}_\eta^s}\mathrm{d}s'-\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left[R_\theta^\eta\right](t,s')\frac{\partial \ell_{\lfloor s'\rfloor}(r_\eta^{s'};z)}{\partial w_\eta^s}\mathrm{d}s'\right\|\right]\right.\end{aligned}$$

$$+\mathbb{E}\left[\left\|\overline{R}^{\eta}_{\theta}(t,s)\nabla_r\ell_s(\overline{r}^s_\eta;z)-\left[R^{\eta}_{\theta}\right](t,s)\nabla_r\ell_{\lfloor s\rfloor}(r^s_\eta;z)\right\|\right]\Bigg\}. \tag{339}$$

From Eq. (223) we can also get for any $0\le s<t\le T$,

$$\left\|\frac{\partial\ell_t(\overline{r}^t_\eta;z)}{\partial\overline{w}^s_\eta}\right\|\le M_\ell\cdot\left\|\frac{\partial\overline{r}^t_\eta}{\partial\overline{w}^s_\eta}\right\|\le\frac{M_\ell^2\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right), \tag{340}$$

and therefore

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\int_s^t\overline{R}^{\eta}_{\theta}(t,s')\frac{\partial\ell_{s'}(\overline{r}^{s'}_\eta;z)}{\partial\overline{w}^s_\eta}\mathrm{d}s'-\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left[R^{\eta}_{\theta}\right](t,s')\frac{\partial\ell_{\lfloor s'\rfloor}(r^{s'}_\eta;z)}{\partial w^s_\eta}\mathrm{d}s'\right\|\right]\\
&\le 2\eta\Phi_{R_\theta}(T)\cdot\frac{M_\ell^2\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)\\
&\quad+e^{-\lambda t}\mathbb{E}\left[\left\|\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\overline{R}^{\eta}_{\theta}(t,s')\frac{\partial\ell_{s'}(\overline{r}^{s'}_\eta;z)}{\partial\overline{w}^s_\eta}\mathrm{d}s'-\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left[R^{\eta}_{\theta}\right](t,s')\frac{\partial\ell_{\lfloor s'\rfloor}(r^{s'}_\eta;z)}{\partial w^s_\eta}\mathrm{d}s'\right\|\right]\\
&\le 2\eta\Phi_{R_\theta}(T)\cdot\frac{M_\ell^2\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)\\
&\quad+e^{-\lambda t}\mathbb{E}\left[\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left\|\overline{R}^{\eta}_{\theta}(t,s')-\left[R^{\eta}_{\theta}\right](t,s')\right\|\cdot\left\|\frac{\partial\ell_{s'}(\overline{r}^{s'}_\eta;z)}{\partial\overline{w}^s_\eta}\right\|\mathrm{d}s'\right]\\
&\quad+e^{-\lambda t}\mathbb{E}\left[\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left\|\left[R^{\eta}_{\theta}\right](t,s')\right\|\cdot\left\|\frac{\partial\ell_{s'}(\overline{r}^{s'}_\eta;z)}{\partial\overline{w}^s_\eta}-\frac{\partial\ell_{\lfloor s'\rfloor}(r^{s'}_\eta;z)}{\partial w^s_\eta}\right\|\mathrm{d}s'\right]\\
&\le 2\eta\Phi_{R_\theta}(T)\cdot\frac{M_\ell^2\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)+\mathsf{dist}_{\lambda,T}\left(\overline{R}^{\eta}_{\theta},\left[R^{\eta}_{\theta}\right]\right)\cdot T\cdot\frac{M_\ell^2\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)\\
&\quad+\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}e^{-\lambda(t-s')}\Phi_{R_\theta}(t-s')\cdot e^{-\lambda s'}\mathbb{E}\left[\left\|\frac{\partial\ell_{s'}(\overline{r}^{s'}_\eta;z)}{\partial\overline{w}^s_\eta}-\frac{\partial\ell_{\lfloor s'\rfloor}(r^{s'}_\eta;z)}{\partial w^s_\eta}\right\|\right]\mathrm{d}s'.
\end{aligned} \tag{341}$$

Take λ large enough such that Lemma C.6 holds and also

$$\int_0^\infty e^{-\lambda t}\Phi_{R_\theta}(t)\mathrm{d}t\le\frac{1}{2},$$

we can further get

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\int_s^t\overline{R}^{\eta}_{\theta}(t,s')\frac{\partial\ell_{s'}(\overline{r}^{s'}_\eta;z)}{\partial\overline{w}^s_\eta}\mathrm{d}s'-\int_{\min\{\lceil s\rceil,\lfloor t\rfloor\}}^{\lfloor t\rfloor}\left[R^{\eta}_{\theta}\right](t,s')\frac{\partial\ell_{\lfloor s'\rfloor}(r^{s'}_\eta;z)}{\partial w^s_\eta}\mathrm{d}s'\right\|\right]\\
&\le 2\eta\Phi_{R_\theta}(T)\cdot\frac{M_\ell^2\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)+\overline{h}(\eta)\cdot T\cdot\frac{M_\ell^2\Phi_{R_\theta}(T)}{\delta}\cdot\exp\left(\frac{M_\ell T\Phi_{R_\theta}(T)}{\delta}\right)\\
&\qquad+\frac{1}{2}\sup_{0\le s<t\le T}e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial\ell_t(\overline{r}^t_\eta;z)}{\partial\overline{w}^s_\eta}-\frac{\partial\ell_{\lfloor t\rfloor}(r^t_\eta;z)}{\partial w^s_\eta}\right\|\right]\\
&=:h_2(\eta)+\frac{1}{2}\sup_{0\le s<t\le T}e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial\ell_t(\overline{r}^t_\eta;z)}{\partial\overline{w}^s_\eta}-\frac{\partial\ell_{\lfloor t\rfloor}(r^t_\eta;z)}{\partial w^s_\eta}\right\|\right],
\end{aligned} \tag{342}$$

where $h_2(\eta)\to 0$ when η approaches 0. For the same λ, we also get

$$\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\overline{R}^{\eta}_{\theta}(t,s)\nabla_r\ell_s(\overline{r}^s_\eta;z)-\left[R^{\eta}_{\theta}\right](t,s)\nabla_r\ell_{\lfloor s\rfloor}(r^s_\eta;z)\right\|\right]\\
&\le e^{-\lambda t}\mathbb{E}\left[\left\|\overline{R}^{\eta}_{\theta}(t,s)-\left[R^{\eta}_{\theta}\right](t,s)\right\|\left\|\nabla_r\ell_s(\overline{r}^s_\eta;z)\right\|\right]+e^{-\lambda t}\mathbb{E}\left[\left\|\left[R^{\eta}_{\theta}\right](t,s)\right\|\left\|\nabla_r\ell_s(\overline{r}^s_\eta;z)-\nabla_r\ell_{\lfloor s\rfloor}(r^s_\eta;z)\right\|\right]
\end{aligned}$$

$$
\leq M_\ell \overline{h}(\eta) + \Phi_{R_\theta}(T) h_1(\eta) \,, \tag{343}
$$

in the last line we make use of Eq. (334). Taking Eqs. (342) and (343) into Eq. (339) yields

$$
\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\nabla_r \ell_{\lfloor t \rfloor}(r_\eta^t; z)\right\|\left\|\frac{\partial \overline{r}_\eta^t}{\partial \overline{w}_\eta^s} - \frac{\partial r_\eta^t}{\partial w_\eta^s}\right\|\right] \\
&\leq h_2(\eta) + M_\ell \overline{h}(\eta) + \Phi_{R_\theta}(T) h_1(\eta) + \frac{1}{2}\sup_{0\leq s<t\leq T} e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(\overline{r}_\eta^t; z)}{\partial \overline{w}_\eta^s} - \frac{\partial \ell_{\lfloor t \rfloor}(r_\eta^t; z)}{\partial w_\eta^s}\right\|\right] .
\end{aligned} \tag{344}
$$

Further with Eq. (338), substituting into Eq. (337) gives us

$$
\begin{aligned}
&e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(\overline{r}_\eta^t; z)}{\partial \overline{w}_\eta^s} - \frac{\partial \ell_{\lfloor t \rfloor}(r_\eta^t; z)}{\partial w_\eta^s}\right\|\right] \\
&\leq \frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right) \cdot h_1(\eta) + h_2(\eta) + M_\ell \overline{h}(\eta) + \Phi_{R_\theta}(T) h_1(\eta) \\
&\quad + \frac{1}{2}\sup_{0\leq s<t\leq T} e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(\overline{r}_\eta^t; z)}{\partial \overline{w}_\eta^s} - \frac{\partial \ell_{\lfloor t \rfloor}(r_\eta^t; z)}{\partial w_\eta^s}\right\|\right] .
\end{aligned} \tag{345}
$$

Taking supremum on both sides for $0 \leq s < t \leq T$ it then follows that

$$
\begin{aligned}
&\sup_{0\leq s<t\leq T} e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(\overline{r}_\eta^t; z)}{\partial \overline{w}_\eta^s} - \frac{\partial \ell_{\lfloor t \rfloor}(r_\eta^t; z)}{\partial w_\eta^s}\right\|\right] \\
&\leq 2\left\{\frac{M_\ell \Phi_{R_\theta}(T)}{\delta} \cdot \exp\left(\frac{M_\ell T \Phi_{R_\theta}(T)}{\delta}\right) \cdot h_1(\eta) + h_2(\eta) + M_\ell \overline{h}(\eta) + \Phi_{R_\theta}(T) h_1(\eta)\right\} =: h_3(\eta) \,,
\end{aligned} \tag{346}
$$

and $h_3(\eta) \to 0$ when $\eta \to 0$. Finally

$$
\begin{aligned}
\mathsf{dist}_{\lambda,T}\left(\overline{C}_\ell^\eta, C_\ell^\eta\right) &= \sup_{0\leq s<t\leq T} e^{-\lambda t}\left\|\mathbb{E}\left[\frac{\partial \ell_t(\overline{r}_\eta^t; z)}{\partial \overline{w}_\eta^s}\right] - \mathbb{E}\left[\frac{\partial \ell_{\lfloor t \rfloor}(r_\eta^t; z)}{\partial w_\eta^s}\right]\right\| \\
&\leq e^{-\lambda t}\mathbb{E}\left[\left\|\frac{\partial \ell_t(\overline{r}_\eta^t; z)}{\partial \overline{w}_\eta^s} - \frac{\partial \ell_{\lfloor t \rfloor}(r_\eta^t; z)}{\partial w_\eta^s}\right\|\right] \\
&\leq h_3(\eta) .
\end{aligned} \tag{347}
$$

By Eqs. (333), (334) and (347), we conclude the proof by taking $h(\eta) = \max\{h_1(\eta), h_3(\eta)\}$.

C.7 Proof of Lemma C.5

By Eqs. (12a) and (113a), we can write

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\theta^t &= -(\Lambda^t + \Gamma^t)\theta^t - \int_0^t R_\ell(t,s)\theta^s \mathrm{d}s + u^t \,, && u^t \sim \mathsf{GP}(0, C_\ell/\delta) \,, \\
\frac{\mathrm{d}}{\mathrm{d}t}\theta_\eta^t &= -(\Lambda^{\lfloor t \rfloor} + \Gamma_\eta^t)\theta_\eta^{\lfloor t \rfloor} - \int_0^{\lfloor t \rfloor} R_\ell^\eta(t,s)\theta_\eta^{\lfloor s \rfloor} \mathrm{d}s + u_\eta^t \,, && u_\eta^t \sim \mathsf{GP}(0, C_\ell^\eta/\delta) \,,
\end{aligned}
$$

where we use the fact that $R_\ell^\eta(\lfloor t \rfloor, \lfloor s \rfloor) = R_\ell^\eta(t,s)$ and $\Gamma_\eta^{\lfloor t \rfloor} = \Gamma_\eta^t$. We can couple u^t and u_η^t such that

$$
\sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|u^t - u_\eta^t\right\|_2^2\right]} \leq 2 \cdot \mathsf{dist}_{\lambda,T}\left(u^t, u_\eta^t\right) = 2 \cdot \mathsf{dist}_{\lambda,T}\left(C_\ell/\delta, C_\ell^\eta/\delta\right) = \frac{2}{\sqrt{\delta}}\mathsf{dist}_{\lambda,T}\left(C_\ell, C_\ell^\eta\right) . \tag{348}
$$

We can derive the upper bound

$$
e^{-\lambda t} \cdot \frac{\mathrm{d}}{\mathrm{d}t}\left\|\theta^t - \theta_\eta^t\right\|_2
$$

$$
\begin{aligned}
&\le e^{-\lambda t} \cdot \left\| \frac{\mathrm{d}}{\mathrm{d}t}\theta^t - \frac{\mathrm{d}}{\mathrm{d}t}\theta^t_\eta \right\|_2 \\
&\le e^{-\lambda t} \cdot \left\| -(\Lambda^t + \Gamma^t)\theta^t - \int_0^t R_\ell(t,s)\theta^s \mathrm{d}s + u^t + (\Lambda^{\lfloor t \rfloor} + \Gamma^t_\eta)\theta_\eta^{\lfloor t \rfloor} + \int_0^{\lfloor t \rfloor} R_\ell^\eta(t,s)\theta_\eta^{\lfloor s \rfloor}\mathrm{d}s - u^t_\eta \right\|_2 \\
&\le \underbrace{e^{-\lambda t} \cdot \left\| (\Lambda^t + \Gamma^t)\theta^t - (\Lambda^{\lfloor t \rfloor} + \Gamma^t_\eta)\theta_\eta^{\lfloor t \rfloor} \right\|_2}_{\text{(I)}} + \underbrace{e^{-\lambda t} \cdot \left\| \int_0^t R_\ell(t,s)\theta^s \mathrm{d}s - \int_0^{\lfloor t \rfloor} R_\ell^\eta(t,s)\theta_\eta^{\lfloor s \rfloor}\mathrm{d}s \right\|_2}_{\text{(II)}} \\
&\quad + e^{-\lambda t} \cdot \left\| u^t - u^t_\eta \right\|_2 . \qquad (349)
\end{aligned}
$$

We upper bound (I) and (II) respectively

$$
\begin{aligned}
\text{(I)} &\le e^{-\lambda t} \left\| (\Lambda^t + \Gamma^t)(\theta^t - \theta^t_\eta) \right\|_2 + e^{-\lambda t} \left\| (\Lambda^t + \Gamma^t)(\theta^t_\eta - \theta_\eta^{\lfloor t \rfloor}) \right\|_2 + e^{-\lambda t} \left\| (\Lambda^t + \Gamma^t - \Lambda^{\lfloor t \rfloor} - \Gamma^t_\eta)\theta_\eta^{\lfloor t \rfloor} \right\|_2 \\
&\le e^{-\lambda t} (M_\Lambda + M_\ell) \cdot \left(\left\| \theta^t - \theta^t_\eta \right\|_2 + \left\| \theta^t_\eta - \theta_\eta^{\lfloor t \rfloor} \right\|_2 \right) + e^{-\lambda t} \left(\left\| \Lambda^t - \Lambda^{\lfloor t \rfloor} \right\| + \left\| \Gamma^t - \Gamma^t_\eta \right\| \right) \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2 \\
&\le e^{-\lambda t} (M_\Lambda + M_\ell) \cdot \left(\left\| \theta^t - \theta^t_\eta \right\|_2 + \left\| \theta^t_\eta - \theta_\eta^{\lfloor t \rfloor} \right\|_2 \right) + e^{-\lambda t} \left(\eta M_\Lambda + \left\| \Gamma^t - \Gamma^t_\eta \right\| \right) \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2 , \qquad (350)
\end{aligned}
$$

and

$$
\begin{aligned}
\text{(II)} &\le e^{-\lambda t} \cdot \left\| \int_{\lfloor t \rfloor}^t R_\ell(t,s)\theta^s \mathrm{d}s \right\|_2 + e^{-\lambda t} \cdot \left\| \int_0^{\lfloor t \rfloor} \left(R_\ell(t,s) - R_\ell^\eta(t,s) \right) \theta^s \mathrm{d}s \right\|_2 \\
&\quad + e^{-\lambda t} \cdot \left\| \int_0^{\lfloor t \rfloor} R_\ell^\eta(t,s) \left(\theta^s - \theta^s_\eta \right) \mathrm{d}s \right\|_2 + e^{-\lambda t} \cdot \left\| \int_0^{\lfloor t \rfloor} R_\ell^\eta(t,s) \left(\theta^s_\eta - \theta_\eta^{\lfloor s \rfloor} \right) \mathrm{d}s \right\|_2 \\
&\le e^{-\lambda t} \Phi_{R_\ell}(T) \int_{\lfloor t \rfloor}^t \left\| \theta^s \right\|_2 \mathrm{d}s + e^{-\lambda t} \int_0^{\lfloor t \rfloor} \left\| R_\ell(t,s) - R_\ell^\eta(t,s) \right\| \left\| \theta^s \right\|_2 \mathrm{d}s \\
&\quad + \int_0^{\lfloor t \rfloor} e^{-\lambda(t-s)} \Phi_{R_\ell}(t-s) \cdot e^{-\lambda s} \left\| \theta^s - \theta^s_\eta \right\|_2 \mathrm{d}s + e^{-\lambda t} \Phi_{R_\ell}(T) \int_0^{\lfloor t \rfloor} \left\| \theta^s_\eta - \theta_\eta^{\lfloor s \rfloor} \right\|_2 \mathrm{d}s . \qquad (351)
\end{aligned}
$$

Suppose $\overline{\lambda}$ satisfies $\int_0^\infty e^{-\overline{\lambda} t} \Phi_{R_\ell}(t)\mathrm{d}t \le M_\Lambda + M_\ell$, combining inequalities above yields

$$
\begin{aligned}
& e^{-\overline{\lambda} t} \cdot \frac{\mathrm{d}}{\mathrm{d}t} \left\| \theta^t - \theta^t_\eta \right\|_2 \\
&\le e^{-\overline{\lambda} t} (M_\Lambda + M_\ell) \cdot \left(\left\| \theta^t - \theta^t_\eta \right\|_2 + \left\| \theta^t_\eta - \theta_\eta^{\lfloor t \rfloor} \right\|_2 \right) + e^{-\overline{\lambda} t} \left(\eta M_\Lambda + \left\| \Gamma^t - \Gamma^t_\eta \right\| \right) \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2 \\
&\quad + e^{-\overline{\lambda} t} \Phi_{R_\ell}(T) \int_{\lfloor t \rfloor}^t \left\| \theta^s \right\|_2 \mathrm{d}s + e^{-\overline{\lambda} t} \int_0^{\lfloor t \rfloor} \left\| R_\ell(t,s) - R_\ell^\eta(t,s) \right\| \left\| \theta^s \right\|_2 \mathrm{d}s \\
&\quad + (M_\Lambda + M_\ell) \cdot \sup_{0 \le s \le t} e^{-\overline{\lambda} s} \left\| \theta^s - \theta^s_\eta \right\|_2 + e^{-\overline{\lambda} t} \Phi_{R_\ell}(T) \int_0^{\lfloor t \rfloor} \left\| \theta^s_\eta - \theta_\eta^{\lfloor s \rfloor} \right\|_2 \mathrm{d}s + e^{-\overline{\lambda} t} \left\| u^t - u^t_\eta \right\|_2 \\
&\le 2(M_\Lambda + M_\ell) \cdot \sup_{0 \le s \le t} e^{-\overline{\lambda} s} \left\| \theta^s - \theta^s_\eta \right\|_2 + e^{-\overline{\lambda} t} \left\| u^t - u^t_\eta \right\|_2 + e^{-\overline{\lambda} t} (M_\Lambda + M_\ell) \left\| \theta^t_\eta - \theta_\eta^{\lfloor t \rfloor} \right\|_2 \\
&\quad + e^{-\overline{\lambda} t} \Phi_{R_\ell}(T) \int_0^{\lfloor t \rfloor} \left\| \theta^s_\eta - \theta_\eta^{\lfloor s \rfloor} \right\|_2 \mathrm{d}s + e^{-\overline{\lambda} t} \left(\eta M_\Lambda + \left\| \Gamma^t - \Gamma^t_\eta \right\| \right) \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2 \\
&\quad + e^{-\overline{\lambda} t} \Phi_{R_\ell}(T) \int_{\lfloor t \rfloor}^t \left\| \theta^s \right\|_2 \mathrm{d}s + e^{-\overline{\lambda} t} \int_0^{\lfloor t \rfloor} \left\| R_\ell(t,s) - R_\ell^\eta(t,s) \right\| \left\| \theta^s \right\|_2 \mathrm{d}s . \qquad (352)
\end{aligned}
$$

Similar to previous proofs in Eqs. (182) and (183), it follows that

$$
e^{-2(M_\Lambda + M_\ell)t - \overline{\lambda} t} \left\| \theta^t - \theta^t_\eta \right\|_2
$$

$$\leq \int_0^t e^{-2(M_\Lambda+M_\ell)s-\overline{\lambda}s}\cdot\Bigg\{\left\|u^s-u_\eta^s\right\|_2+(M_\Lambda+M_\ell)\left\|\theta_\eta^s-\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_0^{\lfloor s\rfloor}\left\|\theta_\eta^{s'}-\theta_\eta^{\lfloor s'\rfloor}\right\|_2\mathrm{d}s'$$
$$+\left(\eta M_\Lambda+\left\|\Gamma^s-\Gamma_\eta^s\right\|\right)\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_{\lfloor s\rfloor}^s\left\|\theta^{s'}\right\|_2\mathrm{d}s'+\int_0^{\lfloor s\rfloor}\left\|R_\ell(s,s')-R_\ell^\eta(s,s')\right\|\left\|\theta^{s'}\right\|_2\mathrm{d}s'\Bigg\}\mathrm{d}s\,. \tag{353}$$

This implies for any $\lambda\geq 2(M_\Lambda+M_\ell)+\overline{\lambda}$, one has

$$e^{-\lambda t}\left\|\theta^t-\theta_\eta^t\right\|_2$$
$$\leq \int_0^t e^{-\lambda s}\cdot\Bigg\{\left\|u^s-u_\eta^s\right\|_2+(M_\Lambda+M_\ell)\left\|\theta_\eta^s-\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_0^{\lfloor s\rfloor}\left\|\theta_\eta^{s'}-\theta_\eta^{\lfloor s'\rfloor}\right\|_2\mathrm{d}s'$$
$$+\left(\eta M_\Lambda+\left\|\Gamma^s-\Gamma_\eta^s\right\|\right)\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_{\lfloor s\rfloor}^s\left\|\theta^{s'}\right\|_2\mathrm{d}s'+\int_0^{\lfloor s\rfloor}\left\|R_\ell(s,s')-R_\ell^\eta(s,s')\right\|\left\|\theta^{s'}\right\|_2\mathrm{d}s'\Bigg\}\mathrm{d}s$$
$$\leq \int_0^t\Bigg(e^{-\lambda s}\left\|u^s-u_\eta^s\right\|_2+(M_\Lambda+M_\ell)\left\|\theta_\eta^s-\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_0^{\lfloor s\rfloor}\left\|\theta_\eta^{s'}-\theta_\eta^{\lfloor s'\rfloor}\right\|_2\mathrm{d}s'$$
$$+\left(\eta M_\Lambda+\mathsf{dist}_{\lambda,T}\left(\Gamma,\Gamma_\eta\right)\right)\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_{\lfloor s\rfloor}^s\left\|\theta^{s'}\right\|_2\mathrm{d}s'+\mathsf{dist}_{\lambda,T}\left(R_\ell,R_\ell^\eta\right)\cdot\int_0^{\lfloor s\rfloor}\left\|\theta^{s'}\right\|_2\mathrm{d}s'\Bigg)\mathrm{d}s\,. \tag{354}$$

Square both sides and take expectations, we have

$$e^{-2\lambda t}\mathbb{E}\left[\left\|\theta^t-\theta_\eta^t\right\|_2^2\right]$$
$$\leq \mathbb{E}\Bigg[\Bigg\{\int_0^t\Bigg(e^{-\lambda s}\left\|u^s-u_\eta^s\right\|_2+(M_\Lambda+M_\ell)\left\|\theta_\eta^s-\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_0^{\lfloor s\rfloor}\left\|\theta_\eta^{s'}-\theta_\eta^{\lfloor s'\rfloor}\right\|_2\mathrm{d}s'$$
$$+\left(\eta M_\Lambda+\mathsf{dist}_{\lambda,T}\left(\Gamma,\Gamma_\eta\right)\right)\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2+\Phi_{R_\ell}(T)\int_{\lfloor s\rfloor}^s\left\|\theta^{s'}\right\|_2\mathrm{d}s'+\mathsf{dist}_{\lambda,T}\left(R_\ell,R_\ell^\eta\right)\cdot\int_0^{\lfloor s\rfloor}\left\|\theta^{s'}\right\|_2\mathrm{d}s'\Bigg)\mathrm{d}s\Bigg\}^2\Bigg]$$
$$\overset{\text{(i)}}{\leq}\mathbb{E}\Bigg[\Bigg\{\int_0^t\Bigg(1+1+\int_0^{\lfloor s\rfloor}1\mathrm{d}s'+1+\int_{\lfloor s\rfloor}^s 1\mathrm{d}s'+\int_0^{\lfloor s\rfloor}1\mathrm{d}s'\Bigg)\mathrm{d}s\Bigg\}$$
$$\cdot\Bigg\{\int_0^t\Bigg(e^{-2\lambda s}\left\|u^s-u_\eta^s\right\|_2^2+(M_\Lambda+M_\ell)^2\left\|\theta_\eta^s-\theta_\eta^{\lfloor s\rfloor}\right\|_2^2+\Phi_{R_\ell}(T)^2\int_0^{\lfloor s\rfloor}\left\|\theta_\eta^{s'}-\theta_\eta^{\lfloor s'\rfloor}\right\|_2^2\mathrm{d}s'$$
$$+\left(\eta M_\Lambda+\mathsf{dist}_{\lambda,T}\left(\Gamma,\Gamma_\eta\right)\right)^2\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2^2+\Phi_{R_\ell}(T)^2\int_{\lfloor s\rfloor}^s\left\|\theta^{s'}\right\|_2^2\mathrm{d}s'+\mathsf{dist}_{\lambda,T}\left(R_\ell,R_\ell^\eta\right)^2\cdot\int_0^{\lfloor s\rfloor}\left\|\theta^{s'}\right\|_2^2\mathrm{d}s'\Bigg)\mathrm{d}s\Bigg\}\Bigg]$$
$$\leq\Bigg\{\int_0^t(2s+3)\mathrm{d}s\Bigg\}\cdot\Bigg\{\int_0^t\Bigg(e^{-2\lambda s}\mathbb{E}\left[\left\|u^s-u_\eta^s\right\|_2^2\right]+(M_\Lambda+M_\ell)^2\mathbb{E}\left[\left\|\theta_\eta^s-\theta_\eta^{\lfloor s\rfloor}\right\|_2^2\right]$$
$$+\Phi_{R_\ell}(T)^2\int_0^{\lfloor s\rfloor}\mathbb{E}\left[\left\|\theta_\eta^{s'}-\theta_\eta^{\lfloor s'\rfloor}\right\|_2^2\right]\mathrm{d}s'+\left(\eta M_\Lambda+\mathsf{dist}_{\lambda,T}\left(\Gamma,\Gamma_\eta\right)\right)^2\mathbb{E}\left[\left\|\theta_\eta^{\lfloor s\rfloor}\right\|_2^2\right]+\Phi_{R_\ell}(T)^2\int_{\lfloor s\rfloor}^s\mathbb{E}\left[\left\|\theta^{s'}\right\|_2^2\right]\mathrm{d}s'$$
$$+\mathsf{dist}_{\lambda,T}\left(R_\ell,R_\ell^\eta\right)^2\cdot\int_0^{\lfloor s\rfloor}\mathbb{E}\left[\left\|\theta^{s'}\right\|_2^2\right]\mathrm{d}s'\Bigg)\mathrm{d}s\Bigg\}, \tag{355}$$

where in (i) we use Cauchy-Schwarz inequality. Substituting in Eqs. (359), (348) and $\mathbb{E}\left[\left\|\theta\right\|_2^2\right]\leq k\left\|\mathbb{E}\left[\theta\theta^{\mathsf{T}}\right]\right\|$ yields

$$e^{-2\lambda t}\mathbb{E}\left[\left\|\theta^t-\theta_\eta^t\right\|_2^2\right]$$

$$
\begin{aligned}
&\leq (T^2+3T)\cdot\int_0^t \Bigg(\frac{4}{\delta}\mathsf{dist}_{\lambda,T}\left(C_\ell, C_\ell^\eta\right)^2 + (M_\Lambda+M_\ell)^2\overline{h}_1(\eta) + \Phi_{R_\ell}(T)^2\int_0^{\lfloor s\rfloor}\overline{h}_1(\eta)\mathrm{d}s' \\
&\quad + \left(\eta M_\Lambda + \mathsf{dist}_{\lambda,T}\left(\Gamma,\Gamma_\eta\right)\right)^2 k\Phi_{C_\theta}(T) + \Phi_{R_\ell}(T)^2\int_{\lfloor s\rfloor}^s k\Phi_{C_\theta}(T)\mathrm{d}s' + \mathsf{dist}_{\lambda,T}\left(R_\ell,R_\ell^\eta\right)^2\cdot\int_0^{\lfloor s\rfloor}k\Phi_{C_\theta}(T)\mathrm{d}s'\Bigg)\mathrm{d}s \\
&\leq (T^3+3T^2)\cdot\Bigg(\frac{4}{\delta}\mathsf{dist}_{\lambda,T}\left(C_\ell, C_\ell^\eta\right)^2 + (M_\Lambda+M_\ell)^2\overline{h}_1(\eta) + T\Phi_{R_\ell}(T)^2\overline{h}_1(\eta) + \left(\eta M_\Lambda + \mathsf{dist}_{\lambda,T}\left(\Gamma,\Gamma_\eta\right)\right)^2 k\Phi_{C_\theta}(T) \\
&\quad + \eta\Phi_{R_\ell}(T)^2 k\Phi_{C_\theta}(T) + \mathsf{dist}_{\lambda,T}\left(R_\ell,R_\ell^\eta\right)^2 kT\Phi_{C_\theta}(T)\Bigg)\,.
\end{aligned}
\tag{356}
$$

The proof is then completed by

$$
\begin{aligned}
&e^{-\lambda t}\sqrt{\mathbb{E}\left[\left\|\theta^t-\theta_\eta^t\right\|_2^2\right]} \\
&\leq (T^3+3T^2)^{\frac{1}{2}}\cdot\Bigg(\frac{4}{\delta}\mathsf{dist}_{\lambda,T}\left(X,X^\eta\right)^2 + (M_\Lambda+M_\ell)^2\overline{h}_1(\eta) + T\Phi_{R_\ell}(T)^2\overline{h}_1(\eta) \\
&\qquad + \left(\eta M_\Lambda + \mathsf{dist}_{\lambda,T}\left(X,X^\eta\right)\right)^2 k\Phi_{C_\theta}(T) + \eta\Phi_{R_\ell}(T)^2 k\Phi_{C_\theta}(T) + \mathsf{dist}_{\lambda,T}\left(X,X^\eta\right)^2 kT\Phi_{C_\theta}(T)\Bigg)^{\frac{1}{2}} \\
&=: H(\eta,\mathsf{dist}_{\lambda,T}\left(X,X^\eta\right))\,.
\end{aligned}
\tag{357}
$$

C.8 Proof of Lemma C.6

We write the equations that define $(\overline{C}_\theta^\eta, \overline{R}_\theta^\eta)$ and $(C_\theta^\eta, R_\theta^\eta)$ as

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\overline{\theta}_\eta^t &= -(\Lambda^t+\Gamma_\eta^t)\overline{\theta}_\eta^t - \int_0^t R_\ell^\eta(t,s)\overline{\theta}_\eta^s\mathrm{d}s + \overline{u}_\eta^t\,, && \overline{u}_\eta^t\sim\mathsf{GP}(0,C_\ell^\eta/\delta)\,,\\
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\overline{\theta}_\eta^t}{\partial\overline{u}_\eta^s} &= -(\Lambda^t+\Gamma_\eta^t)\frac{\partial\overline{\theta}_\eta^t}{\partial\overline{u}_\eta^s} - \int_s^t R_\ell^\eta(t,s')\frac{\partial\overline{\theta}_\eta^{s'}}{\partial\overline{u}_\eta^s}\mathrm{d}s'\,, && 0\leq s\leq t\leq T\,,\\
\overline{C}_\theta^\eta(t,s) &= \mathbb{E}\left[\overline{\theta}_\eta^t\overline{\theta}_\eta^{s\,\mathsf{T}}\right]\,, && 0\leq s\leq t\leq T\,,\\
\overline{R}_\theta^\eta(t,s) &= \mathbb{E}\left[\frac{\partial\overline{\theta}_\eta^t}{\partial\overline{u}_\eta^s}\right]\,, && 0\leq s\leq t\leq T\,,
\end{aligned}
$$

and

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\theta_\eta^t &= -(\Lambda^{\lfloor t\rfloor}+\Gamma_\eta^{\lfloor t\rfloor})\theta_\eta^{\lfloor t\rfloor} - \int_0^{\lfloor t\rfloor} R_\ell^\eta(\lfloor t\rfloor,\lfloor s\rfloor)\theta_\eta^{\lfloor s\rfloor}\mathrm{d}s + u_\eta^t\,, && u_\eta^t\sim\mathsf{GP}(0,C_\ell^\eta/\delta)\,,\\
\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\theta_\eta^t}{\partial u_\eta^s} &= -(\Lambda^{\lfloor t\rfloor}+\Gamma_\eta^{\lfloor t\rfloor})\frac{\partial\theta_\eta^{\lfloor t\rfloor}}{\partial u_\eta^s} - \int_s^{\lfloor t\rfloor} R_\ell^\eta(\lfloor t\rfloor,\lfloor s'\rfloor)\frac{\partial\theta_\eta^{\lfloor s'\rfloor}}{\partial u_\eta^s}\mathrm{d}s'\,, && 0\leq s\leq t\leq T\,,\\
C_\theta^\eta(t,s) &= \mathbb{E}\left[\theta_\eta^{\lfloor t\rfloor}\theta_\eta^{\lfloor s\rfloor\,\mathsf{T}}\right]\,, && 0\leq s\leq t\leq T\,,\\
R_\theta^\eta(t,s) &= \mathbb{E}\left[\frac{\partial\theta_\eta^t}{\partial u_\eta^s}\right]\,, && 0\leq s\leq t\leq T\,.
\end{aligned}
$$

Controlling the distance between $\overline{C}_\theta^\eta$ and C_θ^η. From Eq. (113g) we know that C_ℓ^η is piecewise constant, i.e. $C_\ell^\eta(t,s) = C_\ell^\eta(\lfloor t\rfloor,\lfloor s\rfloor)$ and $u_\eta^t = u_\eta^{\lfloor t\rfloor}$. Further since R_ℓ^η is piecewise constant from Eq. (113d), it follows that

$$
\theta_\eta^t - \theta_\eta^{\lfloor t\rfloor} = (t-\lfloor t\rfloor)\cdot\left(-(\Lambda^{\lfloor t\rfloor}+\Gamma_\eta^{\lfloor t\rfloor})\theta_\eta^{\lfloor t\rfloor} - \int_0^{\lfloor t\rfloor}R_\ell^\eta(t,s)\theta_\eta^{\lfloor s\rfloor}\mathrm{d}s + u_\eta^{\lfloor t\rfloor}\right)\,.
\tag{358}
$$

Hence using the fact that $(C_\ell^\eta, R_\ell^\eta, \Gamma_\eta) \in \mathcal{S}$,

$$
\begin{aligned}
\mathbb{E}\left[\left\|\theta_\eta^t - \theta_\eta^{\lfloor t \rfloor}\right\|_2^2\right] &\leq \eta^2 \cdot \mathbb{E}\left[\left\|-(\Lambda^{\lfloor t \rfloor} + \Gamma_\eta^{\lfloor t \rfloor})\theta_\eta^{\lfloor t \rfloor} - \int_0^{\lfloor t \rfloor} R_\ell^\eta(t,s)\theta_\eta^{\lfloor s \rfloor}\mathrm{d}s + u_\eta^{\lfloor t \rfloor}\right\|_2^2\right] \\
&\leq \eta^2 \cdot \mathbb{E}\left[\left((M_\Lambda + M_\ell)\cdot\left\|\theta_\eta^{\lfloor t \rfloor}\right\|_2 + \int_0^{\lfloor t \rfloor}\Phi_{R_\ell}(t-s)\left\|\theta_\eta^{\lfloor s \rfloor}\right\|_2 ds + \left\|u_\eta^{\lfloor t \rfloor}\right\|_2\right)^2\right] \\
&\leq \eta^2 \cdot \mathbb{E}\left[\left((M_\Lambda + M_\ell)^2\cdot\left\|\theta_\eta^{\lfloor t \rfloor}\right\|_2^2 + \int_0^{\lfloor t \rfloor}\Phi_{R_\ell}(t-s)^2\left\|\theta_\eta^{\lfloor s \rfloor}\right\|_2^2 ds + \left\|u_\eta^{\lfloor t \rfloor}\right\|_2^2\right)\cdot(1+T+1)\right] \\
&\leq \eta^2\cdot(T+2)\cdot\left\{(M_\Lambda+M_\ell)^2\cdot k\Phi_{C_\theta}(T) + T\Phi_{R_\ell}(T)^2\cdot k\Phi_{C_\theta}(T) + \frac{k}{\delta}\Phi_{C_\ell}(T)\right\} \\
&=: \overline{h}_1(\eta)\,. \qquad (359)
\end{aligned}
$$

where in the last line we use the inequality $\mathbb{E}\left[\|\theta\|_2^2\right] = \mathrm{Tr}\left(\mathbb{E}\left[\theta\theta^\mathsf{T}\right]\right) \leq k\left\|\mathbb{E}\left[\theta\theta^\mathsf{T}\right]\right\|$ for any k dimensional random vector θ. Note that $\overline{h}_1(\eta) \to 0$ as $\eta \to 0$. Next by using the coupling $\overline{u}_\eta^t = u_\eta^t$ and using $\Gamma_\eta^t = \Gamma_\eta^{\lfloor t \rfloor}$ by Eq. (113e),

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\left\|\overline{\theta}_\eta^t - \theta_\eta^t\right\|_2 &\leq \left\|\frac{\mathrm{d}}{\mathrm{d}t}\left(\overline{\theta}_\eta^t - \theta_\eta^t\right)\right\|_2 \\
&= \left\|-(\Lambda^t + \Gamma_\eta^t)\overline{\theta}_\eta^t - \int_0^t R_\ell^\eta(t,s)\overline{\theta}_\eta^s\mathrm{d}s + (\Lambda^{\lfloor t \rfloor} + \Gamma_\eta^t)\theta_\eta^{\lfloor t \rfloor} + \int_0^{\lfloor t \rfloor} R_\ell^\eta(t,s)\theta_\eta^{\lfloor s \rfloor}\mathrm{d}s\right\|_2 \\
&\leq \left\|-\left(\Lambda^t + \Gamma_\eta^t\right)\left(\overline{\theta}_\eta^t - \theta_\eta^t\right) - \left(\Lambda^t + \Gamma_\eta^t\right)\left(\theta_\eta^t - \theta_\eta^{\lfloor t \rfloor}\right) - \left(\Lambda^t - \Lambda^{\lfloor t \rfloor}\right)\theta_\eta^{\lfloor t \rfloor}\right\|_2 \\
&\quad + \left\|\int_0^t R_\ell^\eta(t,s)\left(\overline{\theta}_\eta^s - \theta_\eta^s\right)\mathrm{d}s + \int_0^t R_\ell^\eta(t,s)\left(\theta_\eta^s - \theta_\eta^{\lfloor s \rfloor}\right)\mathrm{d}s - \int_{\lfloor t \rfloor}^t R_\ell^\eta(t,s)\theta_\eta^{\lfloor s \rfloor}\mathrm{d}s\right\|_2 \\
&\leq (M_\Lambda + M_\ell)\left\|\overline{\theta}_\eta^t - \theta_\eta^t\right\|_2 + (M_\Lambda + M_\ell)\left\|\theta_\eta^t - \theta_\eta^{\lfloor t \rfloor}\right\|_2 + \eta M_\Lambda\left\|\theta_\eta^{\lfloor t \rfloor}\right\|_2 \\
&\quad + \int_0^t \Phi_{R_\ell}(t-s)\left\|\overline{\theta}_\eta^s - \theta_\eta^s\right\|_2\mathrm{d}s + \int_0^t \Phi_{R_\ell}(t-s)\left\|\theta_\eta^s - \theta_\eta^{\lfloor s \rfloor}\right\|_2\mathrm{d}s + \eta\Phi_{R_\ell}(T)\left\|\theta_\eta^{\lfloor t \rfloor}\right\|_2\,, \qquad (360)
\end{aligned}
$$

where in the last line we use Assumption 1 which gives $\left\|\Lambda^t - \Lambda^{\lfloor t \rfloor}\right\| \leq \eta M_\Lambda$, the fact $\theta_\eta^{\lfloor s \rfloor} = \theta_\eta^{\lfloor t \rfloor}$ when $\lfloor t \rfloor \leq s \leq t$ and also $\Phi_{R_\ell}(t)$ is a nondecreasing function in $t \in \mathbb{R}_{\geq 0}$. By taking $\overline{\lambda}$ such that $\int_0^\infty e^{-\overline{\lambda}t}\Phi_{R_\ell}(t)\mathrm{d} \leq M_\Lambda + M_\ell$ and repeating the argument in Eq. (181), we can have

$$
\begin{aligned}
&e^{-\overline{\lambda}t}\frac{\mathrm{d}}{\mathrm{d}t}\left\|\overline{\theta}_\eta^t - \theta_\eta^t\right\|_2 \\
&\leq 2\left(M_\Lambda + M_\ell\right)\cdot\sup_{0\leq s\leq t}\left\|\overline{\theta}_\eta^s - \theta_\eta^s\right\|_2 \\
&\quad + e^{-\overline{\lambda}t}\left\{(M_\Lambda + M_\ell)\left\|\theta_\eta^t - \theta_\eta^{\lfloor t \rfloor}\right\|_2 + \eta\left(M_\Lambda + \Phi_{R_\ell}(T)\right)\left\|\theta_\eta^{\lfloor t \rfloor}\right\|_2 + \Phi_{R_\ell}(T)\int_0^t\left\|\theta_\eta^s - \theta_\eta^{\lfloor s \rfloor}\right\|_2\mathrm{d}s\right\}\,. \qquad (361)
\end{aligned}
$$

Consequently

$$
\begin{aligned}
&e^{-2(M_\Lambda + M_\ell)t - \overline{\lambda}t}\left\|\overline{\theta}_\eta^t - \theta_\eta^t\right\|_2 \\
&\leq \int_0^t e^{-2(M_\Lambda + M_\ell)s - \overline{\lambda}s} \\
&\quad \cdot\left\{(M_\Lambda + M_\ell)\left\|\theta_\eta^s - \theta_\eta^{\lfloor s \rfloor}\right\|_2 + \eta\left(M_\Lambda + \Phi_{R_\ell}(T)\right)\left\|\theta_\eta^{\lfloor s \rfloor}\right\|_2 + \Phi_{R_\ell}(T)\int_0^s\left\|\theta_\eta^{s'} - \theta_\eta^{\lfloor s' \rfloor}\right\|_2\mathrm{d}s'\right\}\mathrm{d}t\,, \qquad (362)
\end{aligned}
$$

which further implies for any $\lambda > 2(M_\Lambda + M_\ell) + \overline{\lambda}$, we have

$$
\begin{aligned}
&e^{-\lambda t} \left\| \overline{\theta}_\eta^t - \theta_\eta^t \right\|_2 \\
&\le \int_0^t \left((M_\Lambda + M_\ell) \left\| \theta_\eta^s - \theta_\eta^{\lfloor s \rfloor} \right\|_2 + \eta \left(M_\Lambda + \Phi_{R_\ell}(T) \right) \left\| \theta_\eta^{\lfloor s \rfloor} \right\|_2 + \Phi_{R_\ell}(T) \int_0^s \left\| \theta_\eta^{s'} - \theta_\eta^{\lfloor s' \rfloor} \right\|_2 \mathrm{d}s' \right) \mathrm{d}t \\
&\le \int_0^T \left((M_\Lambda + M_\ell + T\Phi_{R_\ell}(T)) \left\| \theta_\eta^t - \theta_\eta^{\lfloor t \rfloor} \right\|_2 + \eta \left(M_\Lambda + \Phi_{R_\ell}(T) \right) \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2 \right) \mathrm{d}t \,.
\end{aligned}
\tag{363}
$$

By triangle inequality,

$$
\begin{aligned}
&e^{-\lambda t} \sqrt{\mathbb{E}\left[\left\| \overline{\theta}_\eta^t - \theta_\eta^t \right\|_2^2 \right]} \\
&\le \sqrt{\mathbb{E}\left[\left(\int_0^T \left((M_\Lambda + M_\ell + T\Phi_{R_\ell}(T)) \left\| \theta_\eta^t - \theta_\eta^{\lfloor t \rfloor} \right\|_2 + \eta \left(M_\Lambda + \Phi_{R_\ell}(T) \right) \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2 \right) \mathrm{d}t \right)^2 \right]} \\
&\le (M_\Lambda + M_\ell + T\Phi_{R_\ell}(T)) \cdot \sqrt{\mathbb{E}\left[\left(\int_0^T \left\| \theta_\eta^t - \theta_\eta^{\lfloor t \rfloor} \right\|_2 \mathrm{d}t \right)^2 \right]} + \eta \left(M_\Lambda + \Phi_{R_\ell}(T) \right) \cdot \sqrt{\mathbb{E}\left[\left(\int_0^T \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2 \mathrm{d}t \right)^2 \right]} \\
&\le (M_\Lambda + M_\ell + T\Phi_{R_\ell}(T)) \sqrt{T \cdot \mathbb{E}\left[\int_0^T \left\| \theta_\eta^t - \theta_\eta^{\lfloor t \rfloor} \right\|_2^2 \mathrm{d}t \right]} + \eta \left(M_\Lambda + \Phi_{R_\ell}(T) \right) \cdot \sqrt{T \cdot \mathbb{E}\left[\int_0^T \left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2^2 \mathrm{d}t \right]} \,,
\end{aligned}
$$

where we invoke Cauchy-Schwarz inequality in the last line. Substituting in Eq. (359) and $\mathbb{E}\left[\left\| \theta_\eta^{\lfloor t \rfloor} \right\|_2^2 \right] \le k\Phi_{C_\theta}(T)$, we get

$$
e^{-\lambda t} \sqrt{\mathbb{E}\left[\left\| \overline{\theta}_\eta^t - \theta_\eta^t \right\|_2^2 \right]} \le T \left(M_\Lambda + M_\ell + T\Phi_{R_\ell}(T) \right) \overline{h}_1(\eta) + \eta T \left(M_\Lambda + \Phi_{R_\ell}(T) \right) \sqrt{k\Phi_{C_\theta}(T)} \,. \tag{364}
$$

Following the same coupling argument in Appendix B.4.1, we obtain

$$
\begin{aligned}
&\mathsf{dist}_{\lambda,T} \left(\overline{C}_\theta^\eta, C_\theta^\eta \right) \\
&\le \sup_{t \in [0,T]} e^{-\lambda t} \sqrt{\mathbb{E}\left[\left\| \overline{\theta}_\eta^t - \theta_\eta^{\lfloor t \rfloor} \right\|_2^2 \right]} \\
&\le \sup_{t \in [0,T]} e^{-\lambda t} \sqrt{\mathbb{E}\left[\left\| \overline{\theta}_\eta^t - \theta_\eta^t \right\|_2^2 \right]} + \sup_{t \in [0,T]} e^{-\lambda t} \sqrt{\mathbb{E}\left[\left\| \theta_\eta^t - \theta_\eta^{\lfloor t \rfloor} \right\|_2^2 \right]} \\
&\le T \left(M_\Lambda + M_\ell + T\Phi_{R_\ell}(T) \right) \overline{h}_1(\eta) + \eta T \left(M_\Lambda + \Phi_{R_\ell}(T) \right) \sqrt{k\Phi_{C_\theta}(T)} + \overline{h}_1(\eta) =: \overline{h}_2(\eta) \,,
\end{aligned}
\tag{365}
$$

where $\overline{h}_2(\eta) \to 0$ as $\eta \to 0$.

Controlling the distance between $\overline{R}_\theta^\eta$ and $[R_\theta^\eta]$. Since $\partial \overline{\theta}_\eta^t / \partial \overline{u}_\eta^s$ and $\partial \theta_\eta^t / \partial u_\eta^s$ are not random, we can write

$$
\frac{\mathrm{d}}{\mathrm{d}t} \overline{R}_\theta^\eta(t,s) = -(\Lambda^t + \Gamma_\eta^t) \overline{R}_\theta^\eta(t,s) - \int_s^t R_\ell^\eta(t,s') \overline{R}_\theta^\eta(s',s) \mathrm{d}s' \,, \qquad 0 \le s \le t \le T \,, \tag{366}
$$

$$
\frac{\mathrm{d}}{\mathrm{d}t} R_\theta^\eta(t,s) = -(\Lambda^{\lfloor t \rfloor} + \Gamma_\eta^{\lfloor t \rfloor}) R_\theta^\eta(\lfloor t \rfloor, s) - \int_s^{\lfloor t \rfloor} R_\ell^\eta(t,s') R_\theta^\eta(\lfloor s' \rfloor, s) \mathrm{d}s' \,, \qquad 0 \le s \le t \le T \,. \tag{367}
$$

with the same boundary conditions $\overline{R}_\theta^\eta(s,s) = R_\theta^\eta(s,s) = I$ and the convention that $R_\theta^\eta(t,s) = 0$ when $t < s$.

First we try to control the error $\|R^\eta_\theta(t,s) - [R^\eta_\theta](t,s)\|$. By definition, we have $[R^\eta_\theta](t,s) = I$ when $\lceil s\rceil > \lfloor t\rfloor$. In this case, we have $\lfloor t\rfloor \le s$ and therefore $\left\|\frac{\mathrm{d}}{\mathrm{d}t}R^\eta_\theta(t,s)\right\| \le (M_\Lambda + M_\ell)\|R^\eta_\theta(\lfloor t\rfloor, s)\| \le (M_\Lambda + M_\ell)$. We can then control

$$\|R^\eta_\theta(t,s) - [R^\eta_\theta](t,s)\| = \|R^\eta_\theta(t,s) - R^\eta_\theta(s,s)\| \le \eta\,(M_\Lambda + M_\ell)\,. \tag{368}$$

We can then assume $\lceil s\rceil \le \lfloor t\rfloor$. One then can derive

$$\begin{aligned}\|R^\eta_\theta(t,s) - R^\eta_\theta(\lfloor t\rfloor, s)\| &\le \eta \sup_{\lfloor t\rfloor \le s' \le t}\left\|\frac{\mathrm{d}}{\mathrm{d}s'}R^\eta_\theta(s',s)\right\| \\ &\le \eta\cdot\left\|-(\Lambda^{\lfloor t\rfloor} + \Gamma^{\lfloor t\rfloor}_\eta)R^\eta_\theta(\lfloor t\rfloor, s) - \int_s^{\lfloor t\rfloor} R^\eta_\ell(t,s')R^\eta_\theta(\lfloor s'\rfloor, s)\mathrm{d}s'\right\| \\ &\le \eta\cdot\{(M_\Lambda + M_\ell)\,\Phi_{R_\theta}(T) + T\Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\} =: \overline{h}_3(\eta)\,.\end{aligned} \tag{369}$$

In particular, take $t = \lceil s\rceil - \epsilon$ and let $\epsilon \to 0$, it follows then

$$\|R^\eta_\theta(\lceil s\rceil, s) - R^\eta_\theta(s,s)\| = \|R^\eta_\theta(\lceil s\rceil, s) - R^\eta_\theta(\lceil s\rceil, \lceil s\rceil)\| \le \overline{h}_3(\eta)\,. \tag{370}$$

Note that for all $t \ge \lceil s\rceil$,

$$\begin{aligned}&\frac{\mathrm{d}}{\mathrm{d}t}\|R^\eta_\theta(t,s) - R^\eta_\theta(t,\lceil s\rceil)\| \\ &\le \left\|\frac{\mathrm{d}}{\mathrm{d}t}R^\eta_\theta(t,s) - \frac{\mathrm{d}}{\mathrm{d}t}R^\eta_\theta(t,\lceil s\rceil)\right\| \\ &= \Bigg\| -(\Lambda^{\lfloor t\rfloor} + \Gamma^{\lfloor t\rfloor}_\eta)\,(R^\eta_\theta(\lfloor t\rfloor, s) - R^\eta_\theta(\lfloor t\rfloor, \lceil s\rceil)) - \int_{\lceil s\rceil}^{\lfloor t\rfloor} R^\eta_\ell(t,s')\,(R^\eta_\theta(\lfloor s'\rfloor, s) - R^\eta_\theta(\lfloor s'\rfloor, \lceil s\rceil))\,\mathrm{d}s' \\ &\qquad - \int_s^{\lceil s\rceil} R^\eta_\ell(t,s')R^\eta_\theta(\lfloor s'\rfloor, s)\mathrm{d}s'\Bigg\| \\ &\le (M_\Lambda + M_\ell)\,\|R^\eta_\theta(\lfloor t\rfloor, s) - R^\eta_\theta(\lfloor t\rfloor, \lceil s\rceil)\| + \int_{\lceil s\rceil}^{\lfloor t\rfloor}\Phi_{R_\ell}(t-s')\,\|R^\eta_\theta(\lfloor s'\rfloor, s) - R^\eta_\theta(\lfloor s'\rfloor, \lceil s\rceil)\|\,\mathrm{d}s' \\ &\qquad + \eta\Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\,.\end{aligned} \tag{371}$$

Similar to what we did in Eq. (181), taking $\overline{\lambda}$ such that $\int_0^\infty e^{-\overline{\lambda}t}\Phi_{R_\ell}(t)\mathrm{d} \le M_\Lambda + M_\ell$ gives us

$$\begin{aligned}&e^{-\overline{\lambda}t}\frac{\mathrm{d}}{\mathrm{d}t}\|R^\eta_\theta(t,s) - R^\eta_\theta(t,\lceil s\rceil)\| \\ &\le (M_\Lambda + M_\ell)e^{-\overline{\lambda}t}\,\|R^\eta_\theta(\lfloor t\rfloor, s) - R^\eta_\theta(\lfloor t\rfloor, \lceil s\rceil)\| + \int_{\lceil s\rceil}^{\lfloor t\rfloor} e^{-\overline{\lambda}(t-s')}\Phi_{R_\ell}(t-s')\cdot e^{-\overline{\lambda}s'}\,\|R^\eta_\theta(\lfloor s'\rfloor, s) - R^\eta_\theta(\lfloor s'\rfloor, \lceil s\rceil)\|\,\mathrm{d}s' \\ &\qquad + e^{-\overline{\lambda}t}\eta\Phi_{R_\ell}(T)\Phi_{R_\theta}(T) \\ &\le (M_\Lambda + M_\ell)e^{-\overline{\lambda}\lfloor t\rfloor}\,\|R^\eta_\theta(\lfloor t\rfloor, s) - R^\eta_\theta(\lfloor t\rfloor, \lceil s\rceil)\| + \int_{\lceil s\rceil}^{\lfloor t\rfloor} e^{-\overline{\lambda}(t-s')}\Phi_{R_\ell}(t-s')\cdot e^{-\overline{\lambda}\lfloor s'\rfloor}\,\|R^\eta_\theta(\lfloor s'\rfloor, s) - R^\eta_\theta(\lfloor s'\rfloor, \lceil s\rceil)\|\,\mathrm{d}s' \\ &\qquad + e^{-\overline{\lambda}t}\eta\Phi_{R_\ell}(T)\Phi_{R_\theta}(T) \\ &\le 2(M_\Lambda + M_\ell)\cdot\sup_{\lceil s\rceil \le s' \le t} e^{-\overline{\lambda}s'}\,\|R^\eta_\theta(s',s) - R^\eta_\theta(s',\lceil s\rceil)\| + e^{-\overline{\lambda}t}\eta\Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\,.\end{aligned} \tag{372}$$

Further, this allows us to derive by taking in Eq. (370)

$$\begin{aligned}&e^{-2(M_\Lambda+M_\ell)t-\overline{\lambda}t}\,\|R^\eta_\theta(t,s) - R^\eta_\theta(t,\lceil s\rceil)\| \\ &\le \|R^\eta_\theta(\lceil s\rceil, s) - R^\eta_\theta(\lceil s\rceil, \lceil s\rceil)\| + \int_0^t e^{-2(M_\Lambda+M_\ell)s-\overline{\lambda}s}\eta\Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\mathrm{d}s\end{aligned}$$

$$\leq \overline{h}_3(\eta) + \eta T \Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\,. \tag{373}$$

This implies for any $\lambda > 2(M_\Lambda + M_\ell) + \overline{\lambda}$,

$$e^{-\lambda t}\left\|R_\theta^\eta(t,s) - R_\theta^\eta(t,\lceil s\rceil)\right\| \leq \overline{h}_3(\eta) + \eta T \Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\,. \tag{374}$$

Combining Eqs. (368), (369) and (374), we obtain for any $0 \leq s \leq t \leq T$,

$$e^{-\lambda t}\left\|R_\theta^\eta(t,s) - \left[R_\theta^\eta\right](t,s)\right\| \leq \max\left\{\eta\left(M_\Lambda + M_\ell\right), 2\overline{h}_3(\eta) + \eta T\Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\right\} =: \overline{h}_4(\eta)\,. \tag{375}$$

Next we control the term $\left\|\overline{R}_\theta^\eta(t,s) - R_\theta^\eta(t,s)\right\|$, by definition and the fact that $\Gamma_\eta^{\lfloor t\rfloor} = \Gamma_\eta^t$ we have

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left\|\overline{R}_\theta^\eta(t,s) - R_\theta^\eta(t,s)\right\| \\
&\leq \left\|\frac{\mathrm{d}}{\mathrm{d}t}\overline{R}_\theta^\eta(t,s) - \frac{\mathrm{d}}{\mathrm{d}t}R_\theta^\eta(t,s)\right\| \\
&= \left\|-(\Lambda^t + \Gamma_\eta^t)\overline{R}_\theta^\eta(t,s) - \int_s^t R_\ell^\eta(t,s')\overline{R}_\theta^\eta(s',s)\mathrm{d}s' + (\Lambda^{\lfloor t\rfloor} + \Gamma_\eta^t)R_\theta^\eta(\lfloor t\rfloor,s) + \int_s^{\lfloor t\rfloor} R_\ell^\eta(t,s')R_\theta^\eta(\lfloor s'\rfloor,s)\mathrm{d}s'\right\| \\
&\leq \left\|-(\Lambda^t + \Gamma_\eta^t)\left(\overline{R}_\theta^\eta(t,s) - R_\theta^\eta(t,s)\right) - (\Lambda^t + \Gamma_\eta^t)\left(R_\theta^\eta(t,s) - R_\theta^\eta(\lfloor t\rfloor,s)\right) - \left(\Lambda^t - \Lambda^{\lfloor t\rfloor}\right)R_\theta^\eta(\lfloor t\rfloor,s)\right\| \\
&\quad + \left\| - \int_s^{\lfloor t\rfloor} R_\ell^\eta(t,s')\left(\overline{R}_\theta^\eta(s',s) - R_\theta^\eta(s',s)\right)\mathrm{d}s' - \int_s^{\lfloor t\rfloor} R_\ell^\eta(t,s')\left(R_\theta^\eta(s',s) - R_\theta^\eta(\lfloor s'\rfloor,s)\right)\mathrm{d}s'\right. \\
&\quad \left. - \int_{\lfloor t\rfloor}^t R_\ell^\eta(t,s')\overline{R}_\theta^\eta(s',s)\mathrm{d}s'\right\| \\
&\leq (M_\Lambda + M_\ell)\left\|\overline{R}_\theta^\eta(t,s) - R_\theta^\eta(t,s)\right\| + (M_\Lambda + M_\ell)\left\|R_\theta^\eta(t,s) - R_\theta^\eta(\lfloor t\rfloor,s)\right\| + \eta M_\Lambda\Phi_{R_\theta}(T) \\
&\quad + \int_s^{\lfloor t\rfloor}\Phi_{R_\ell}(t-s')\left\|\overline{R}_\theta^\eta(s',s) - R_\theta^\eta(s',s)\right\|\mathrm{d}s' + T\Phi_{R_\ell}(T)\sup_{s\leq s'\leq t}\left\|R_\theta^\eta(s',s) - R_\theta^\eta(\lfloor s'\rfloor,s)\right\| \\
&\quad + \eta\Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\,.
\end{aligned} \tag{376}$$

Further by Eq. (369), we can obtain

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}t}\left\|\overline{R}_\theta^\eta(t,s) - R_\theta^\eta(t,s)\right\| \\
&\leq (M_\Lambda + M_\ell)\left\|\overline{R}_\theta^\eta(t,s) - R_\theta^\eta(t,s)\right\| + \int_s^{\lfloor t\rfloor}\Phi_{R_\ell}(t-s')\left\|\overline{R}_\theta^\eta(s',s) - R_\theta^\eta(s',s)\right\|\mathrm{d}s' \\
&\quad + \eta\left(M_\Lambda\Phi_{R_\theta}(T) + \Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\right) + \left(M_\Lambda + M_\ell + T\Phi_{R_\ell}(T)\right)\overline{h}_3(\eta)\,.
\end{aligned} \tag{377}$$

We get the exact same type of inequality as in Eq. (360) and we can repeat the same argument and get for all $\lambda > 2(M_\Lambda + M_\ell) + \overline{\lambda}$

$$e^{-\lambda t}\left\|\overline{R}_\theta^\eta(t,s) - R_\theta^\eta(t,s)\right\| \leq \eta T\left(M_\Lambda\Phi_{R_\theta}(T) + \Phi_{R_\ell}(T)\Phi_{R_\theta}(T)\right) + T\left(M_\Lambda + M_\ell + T\Phi_{R_\ell}(T)\right)\overline{h}_3(\eta) =: \overline{h}_5(\eta)\,. \tag{378}$$

Putting together Eqs. (375) and (378) yields

$$\mathsf{dist}_{\lambda,T}\left(\overline{R}_\theta^\eta, \left[R_\theta^\eta\right]\right) \leq \sup_{0\leq s\leq t\leq T} e^{-\lambda t}\left\|\overline{R}_\theta^\eta(t,s) - \left[R_\theta^\eta\right](t,s)\right\| \leq \overline{h}_4(\eta) + \overline{h}_5(\eta) =: \overline{h}_6(\eta)\,. \tag{379}$$

Clearly $\eta \to 0$ we have $\overline{h}_6(\eta) \to 0$. The proof is completed by taking $\overline{h}(\eta) = \max\left\{\overline{h}_2(\eta), \overline{h}_6(\eta)\right\}$.

C.9 Proof of Lemma 6.5

The claim of this lemma follows by establishing separately the following two statements (possibly after adjusting the constants $M()\varepsilon)$):

$$\mathbb{P}\Big(\widehat{\mu}^{(n)}\big(\|\theta^0\| > M(\varepsilon)\big) \geq \varepsilon \text{ for infinitely many } n\Big) = 0\,, \tag{380}$$

$$\mathbb{P}\Big(\widehat{\mu}^{(n)}\big(\|(\theta)_0^T\|_{C^{0,\alpha}} > M(\varepsilon)\big) \geq \varepsilon \text{ for infinitely many } n\Big) = 0\,. \tag{381}$$

We begin by Eq. (380):

$$\widehat{\mu}^{(n)}\big(\|\theta^0\| > M(\varepsilon)\big) = \frac{1}{d}\sum_{i=1}^d \mathbf{1}_{\{\|\theta_i^0\|>M\}} \leq \frac{1}{dM^2}\sum_{i=1}^d \|\theta_i^0\|^2 = \frac{1}{dM^2}\|\boldsymbol{\theta}^0\|_F^2\,.$$

Since by assumption $\mathbb{E}_{\widehat{\mu}_{\theta^0}}\Big[\|\theta^0\|^2\Big] \to \mathbb{E}_{\mu_{\theta^0}}\Big[\|\theta^0\|^2\Big] < \infty$, there exists a constant C such that $\|\boldsymbol{\theta}^0\|_F^2/d \leq C$ for all n large enough. Therefore

$$\widehat{\mu}^{(n)}\big(\|\theta^0\| > M(\varepsilon)\big) \leq \frac{C}{M^2}$$

for all but finitely many values of n, which yields the claim (380).

Next, to prove Eq. (381), we begin by noting that, for any differentiable function $f : [0,T] \to \mathbb{R}^k$, ans any $0 \leq s \leq t \leq T$, we have

$$\begin{aligned}\big\|f(t) - f(s)\big\| &= \Big\|\int_0^T f'(u)\mathbf{1}_{[t,s]}(u)\,\mathrm{d}u\Big\| \\ &\leq (t-s)^{1/2}\Big(\int_0^T \|f'(u)\|^2 \mathrm{d}u\Big)^{1/2}\,,\end{aligned}$$

which implies $\|f\|_{C^{0,1/2}} \leq \|f'\|_{L^2}$. Therefore

$$\begin{aligned}\widehat{\mu}^{(n)}\big(\|(\theta)_0^T\|_{C^{0,1/2}} > M\big) \leq \widehat{\mu}^{(n)}\big(\|(\dot{\theta})_0^T\|_{L^2} > M\big) &= \frac{1}{d}\sum_{i=1}^d \mathbf{1}_{\{\|(\dot{\theta}_i)_0^T\|_{L^2}>M\}} \\ &\leq \frac{1}{M^2 d}\sum_{i=1}^d \|(\dot{\theta}_i)_0^T\|_{L^2}^2 = \frac{1}{M^2 d}\sum_{i=1}^d \int_0^T \|\dot{\theta}_i^t\|^2 \mathrm{d}t \\ &= \frac{1}{M^2 d}\int_0^T \Big\|\boldsymbol{\theta}^t\Lambda^t + \frac{1}{\delta}\boldsymbol{X}\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t;\boldsymbol{z})\Big\|_F^2 \mathrm{d}t\,,\end{aligned}$$

where in the last step we used the definition of the flow, per Eq. (11).

By the Bai-Yin law, there exists $C = C(\delta)$ such that almost surely $\|\boldsymbol{X}\| \leq C(\delta)$ for all but finitely man values of n. Using the conditions on Λ^t, $\boldsymbol{\ell}^t$ in Assumption 1, we deduce that, for all but finitely many values of n,

$$\widehat{\mu}^{(n)}\big(\|(\theta)_0^T\|_{C^{0,1/2}} > M\big) \leq \frac{C}{M^2 d}\int_0^T \big(\|\boldsymbol{\theta}^t\|_F^2 + \|\boldsymbol{\ell}_t(\mathbf{0};\boldsymbol{z})\|_F^2\big)\,\mathrm{d}t\,. \tag{382}$$

It is therefore sufficient to bound $\|\boldsymbol{\theta}^t\|_F^2$. We established already such a bound in Eq. (263), which implies that, for all but finitely many values of n, and for all t,

$$\|\boldsymbol{\theta}^t\| \leq Ce^{Ct}\big(\|\boldsymbol{\theta}^0\|_F + \|\boldsymbol{\ell}_t(\mathbf{0};\boldsymbol{z})\|_F\big)\,.$$

By the assumptions on ℓ_t and $\boldsymbol{z}$, we have $\|\boldsymbol{\ell}_t(\underline{0};\boldsymbol{z})\|^2 \leq C(\|\boldsymbol{\ell}_t(\mathbf{0};\mathbf{0})\|^2 + \|\boldsymbol{z}\|^2) \leq C'd$. Substituting these bounds in Eq. (382), we obtain

$$\widehat{\mu}^{(n)}\big(\|(\theta)_0^T\|_{C^{0,1/2}} > M\big) \leq \frac{Ce^{CT}}{M^2 d}\big(\|\boldsymbol{\theta}^0\|_F^2 + \|\boldsymbol{\ell}_t(\mathbf{0};\boldsymbol{z})\|_F^2\big)\,\mathrm{d}t \leq \frac{C'}{M^2}\,, \tag{383}$$

where the last step follows for all n large enough from the assumptions $\mathbb{E}_{\widehat{\mu}_{\theta^0}}\Big[\|\theta^0\|^2\Big] \to \mathbb{E}_{\mu_{\theta^0}}\Big[\|\theta^0\|^2\Big] < \infty$ and $\mathbb{E}_{\widehat{\mu}_z}\Big[\|z\|^2\Big] \to \mathbb{E}_{\mu_z}\Big[\|z\|^2\Big] < \infty$. This concludes the proof of Eq. (381).

D Proofs for fixed-point equations

D.1 Proof of Corollary 4.1

We write the flow Eq. (28) as

$$\frac{\mathrm{d}\bar{\boldsymbol{\theta}}^t}{\mathrm{d}t} = -\bar{\boldsymbol{\theta}}^t\bar{\Lambda}^t - \frac{1}{\delta}\boldsymbol{X}^\top\bar{\boldsymbol{\ell}}_t(\boldsymbol{X}\bar{\boldsymbol{\theta}}^t;\boldsymbol{z}), \tag{384}$$

where

$$\bar{\Lambda}^t = \mathrm{diag}(\bar{\Lambda}^t_{11}, 0), \qquad \bar{\ell}_t(\boldsymbol{X}\bar{\boldsymbol{\theta}}^t;\boldsymbol{z}) = \begin{pmatrix} \bar{\ell}_t(\boldsymbol{X}\bar{\boldsymbol{\theta}}^t;\boldsymbol{z})_1 \\ 0 \end{pmatrix}, \tag{385}$$

initialized at $\bar{\boldsymbol{\theta}}^0 = (\boldsymbol{\theta}^0, \boldsymbol{\theta}^*)$. Here, we have identified $\Lambda^t = \bar{\Lambda}^t_{11}$ and $\boldsymbol{\ell}_t(\boldsymbol{X}\boldsymbol{\theta}^t, \boldsymbol{X}\boldsymbol{\theta}^*;\boldsymbol{z}) = \bar{\boldsymbol{\ell}}_t(\boldsymbol{X}\boldsymbol{\theta}^t, \boldsymbol{X}\boldsymbol{\theta}^*;\boldsymbol{z})_1$. Corollary 4.1 will follow from applying Theorem 2 to the special case that $\bar{\Lambda}^t$ and $\bar{\ell}_t$ take the special form given above, namely, that they contain zeros in certain coordinates. We will show then that in this case, the unique solution to the integro-differential equations (3.1) and (13) are of the form

$$\begin{gathered}
\bar{\theta}^t = \begin{pmatrix} \bar{\theta}^t_1 \\ \bar{\theta}^0_2 \end{pmatrix}, \qquad \bar{r}^t = \begin{pmatrix} \bar{r}^t_1 \\ \bar{w}^0_2 \end{pmatrix}, \qquad \bar{u}^t = \begin{pmatrix} \bar{u}^t_1 \\ 0 \end{pmatrix}, \qquad \bar{w}^t = \begin{pmatrix} \bar{w}^t_1 \\ \bar{w}^0_2 \end{pmatrix}, \\
\bar{R}_\theta(t,s) = \begin{pmatrix} \bar{R}_\theta(t,s)_{11} & \bar{R}_\theta(t,s)_{12} \\ 0 & I_k \end{pmatrix}, \qquad \bar{R}_\ell(t,s) = \begin{pmatrix} \bar{R}_\ell(t,s)_{11} & \bar{R}_\ell(t,s)_{12} \\ 0 & 0 \end{pmatrix}, \qquad \bar{\Gamma}^t = \begin{pmatrix} \bar{\Gamma}^t_{11} & \bar{\Gamma}^t_{12} \\ 0 & 0 \end{pmatrix}, \\
\frac{\partial\bar{\theta}^t}{\partial\bar{u}^s} = \begin{pmatrix} \left(\frac{\partial\bar{\theta}^t}{\partial\bar{u}^s}\right)_{11} & \left(\frac{\partial\bar{\theta}^t}{\partial\bar{u}^s}\right)_{12} \\ 0 & I_k \end{pmatrix}, \qquad \frac{\partial\bar{\ell}_t(\bar{r}^t;z)}{\partial\bar{w}^s} = \begin{pmatrix} \left(\frac{\partial\bar{\ell}_t(\bar{r}^t;z)}{\partial\bar{w}^s}\right)_{11} & \left(\frac{\partial\bar{\ell}_t(\bar{r}^t;z)}{\partial\bar{w}^s}\right)_{12} \\ 0 & 0 \end{pmatrix}, \\
\nabla_r\bar{\ell}_t(\bar{r}^t;z) = \begin{pmatrix} \nabla_r\bar{\ell}_t(\bar{r}^t;z)_{11} & \nabla_r\bar{\ell}_t(\bar{r}^t;z)_{12} \\ 0 & 0 \end{pmatrix}, \\
\bar{C}_\theta(t,s) = \begin{pmatrix} \bar{C}_\theta(t,s)_{11} & \bar{C}_\theta(t,0)_{12} \\ \bar{C}_\theta(0,t)_{21} & \bar{C}_\theta(0,0)_{22} \end{pmatrix}, \qquad \bar{C}_\ell(t,s) = \begin{pmatrix} \bar{C}_\ell(t,s)_{11} & 0 \\ 0 & 0 \end{pmatrix}.
\end{gathered} \tag{386}$$

Indeed, it is immediate that $\nabla_r\bar{\ell}_t(\bar{r}^t;z)$ is of the claimed form. Then, by Eqs. (12e), (12g), and (13b), we must have that $\frac{\partial\bar{\ell}_t(\bar{r}^t;z)}{\partial\bar{w}^s}$, $\bar{\Gamma}^t$, $\bar{C}_\ell$, and $\bar{R}_\ell$ are of the claimed form (i.e., they have zeros in the locations specified by the preceding display). Thus, $\bar{u}^t_2 = 0$ for all t. Moreover, by Eq. (13a), we must have that the final k-rows of $\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\bar{\theta}^t}{\partial\bar{u}^s}$ are 0 for all t, whence by the intial condition $\frac{\partial\bar{\theta}^t}{\partial\bar{u}^s} = I_{2k}$ we have that $\frac{\partial\bar{\theta}^t}{\partial\bar{u}^s}$ and thus also $\bar{R}_\theta(t,s)$ are of the claimed form. Then, by Eq. (12a), $\left(\frac{\mathrm{d}}{\mathrm{d}t}\bar{\theta}^t\right)_2$ is equal to zero for all t, whence $\bar{\theta}^t_2 = \bar{\theta}^0_2$, so that $\bar{\theta}^t$ is of the claimed form. Then, by Eq. (12f), $\bar{C}_\theta(t,s)$ is of the claimed form (because $\bar{\theta}^t_2 = \bar{\theta}^0_2$). Because $\bar{w}^t$ has covariance kernel $\bar{C}_\theta$, we have that $\bar{w}^t_2 = \bar{w}^0_2$ for all t. Then, by Eq. (12b), we have that $\bar{r}^t_2 = \bar{w}^t_2 = \bar{w}^0_2$, so that $\bar{w}^t$ is of the claimed form. We thus conclude that the unique solution to Eqs. (3.1) and (13) is of the form given in the preceding display.

To complete the proof of Corollary 4.1, we must show that $\bar{\theta}^t, \bar{r}^t, \bar{u}^t, \bar{w}^t, \bar{R}_\theta, \bar{R}_\ell, \bar{\Gamma}^t, \bar{C}_\theta, \bar{C}_\ell$ of the form (386) solves Eqs. (3.1) and (13) if and only if it solves Eqs. (31) and (32). Indeed, plugging (386) into

Eqs. (3.1) and (13) and simplifying where possible gives

$$
\begin{aligned}
\frac{\mathrm{d}}{\mathrm{d}t}\bar{\theta}_1^t &= -(\bar{\Lambda}_{11}^t + \bar{\Gamma}_{11}^t)\bar{\theta}_1^t - \int_0^t \bar{R}_\ell(t,s)_{11}\bar{\theta}_1^s \mathrm{d}s - \Big(\bar{\Gamma}_{12}^t + \int_0^t \bar{R}_\ell(t,s)_{12}\mathrm{d}s\Big)\bar{\theta}_2^s + \bar{u}_1^t,\\
\frac{\mathrm{d}}{\mathrm{d}t}\bar{\theta}_2^0 &= 0,\\
\bar{r}_1^t &= -\frac{1}{\delta}\int_0^t \bar{R}_\theta(t,s)_{11}\bar{\ell}_s(\bar{r}_1^s, \bar{w}_2^0; z)_1 \mathrm{d}s + \bar{w}_1^t,\\
\bar{r}_2^t &= \bar{w}_2^t = \bar{w}_2^0,\\
\bar{R}_\theta(t,s)_{11} &= \mathbb{E}\Big[\Big(\frac{\partial \bar{\theta}^t}{\partial \bar{u}^s}\Big)_{11}\Big],\\
\bar{R}_\theta(t,s)_{12} &= \mathbb{E}\Big[\Big(\frac{\partial \bar{\theta}^t}{\partial \bar{u}^s}\Big)_{12}\Big],\\
\bar{R}_\ell(t,s)_{11} &= \mathbb{E}\Big[\Big(\frac{\partial \bar{\ell}_t(\bar{r}^t;z)}{\partial \bar{w}^s}\Big)_{11}\Big],\\
\bar{R}_\ell(t,s)_{12} &= \mathbb{E}\Big[\Big(\frac{\partial \bar{\ell}_t(\bar{r}^t;z)}{\partial \bar{w}^s}\Big)_{12}\Big],\\
\bar{\Gamma}_{11}^t &= \mathbb{E}\Big[\nabla_{r_1}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1\Big],\\
\bar{\Gamma}_{12}^t &= \mathbb{E}\Big[\nabla_{w_2^0}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1\Big],\\
\bar{C}_\theta(t,s)_{11} &= \mathbb{E}[\bar{\theta}_1^t(\bar{\theta}_1^t)^\top], \qquad \bar{C}_\theta(t,0)_{12} = \mathbb{E}[\bar{\theta}_1^t(\bar{\theta}_2^0)^\top], \qquad \bar{C}_\theta(0,0)_{22} = \mathbb{E}[\bar{\theta}_2^0(\bar{\theta}_2^0)^\top],\\
\bar{C}_\ell(t,s)_{11} &= \mathbb{E}[\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)^\top]\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\partial \bar{\theta}^t}{\partial \bar{u}^s}\Big)_{11} &= -(\bar{\Gamma}_{11}^t + \bar{\Lambda}_{11}^t)\Big(\frac{\partial \bar{\theta}^t}{\partial \bar{u}^s}\Big)_{11} - \int_s^t \bar{R}_\ell(t,s')_{11}\Big(\frac{\partial \bar{\theta}^{s'}}{\partial \bar{u}^s}\Big)_{11}\mathrm{d}s',\\
\frac{\mathrm{d}}{\mathrm{d}t}\Big(\frac{\partial \bar{\theta}^t}{\partial \bar{u}^s}\Big)_{12} &= -(\bar{\Gamma}_{11}^t + \bar{\Lambda}_{11}^t)\Big(\frac{\partial \bar{\theta}^t}{\partial \bar{u}^s}\Big)_{12} - \int_s^t \bar{R}_\ell(t,s')_{11}\Big(\frac{\partial \bar{\theta}^{s'}}{\partial \bar{u}^s}\Big)_{12}\mathrm{d}s' - \int_s^t \bar{R}_\ell(t,s')_{12}\mathrm{d}s',\\
\Big(\frac{\partial \bar{\ell}_t(\bar{r}^t;z)}{\partial \bar{w}^s}\Big)_{11} &= \nabla_{r_1}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 \cdot \left(-\frac{1}{\delta}\int_s^t \bar{R}_\theta(t,s')_{11}\Big(\frac{\partial \bar{\ell}_{s'}(\bar{r}^{s'};z)}{\partial \bar{w}^s}\Big)_{11}\mathrm{d}s' - \frac{1}{\delta}\bar{R}_\theta(t,s)_{11}\nabla_{r_1}\bar{\ell}_s(\bar{r}_1^s, \bar{w}_2^0; z)_{11}\right),\\
\Big(\frac{\partial \bar{\ell}_t(\bar{r}^t;z)}{\partial \bar{w}^s}\Big)_{12} &= \nabla_{r_1}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 \cdot \left(-\frac{1}{\delta}\int_s^t \bar{R}_\theta(t,s')_{11}\Big(\frac{\partial \bar{\ell}_{s'}(\bar{r}^{s'};z)}{\partial \bar{w}^s}\Big)_{12}\mathrm{d}s' - \frac{1}{\delta}\bar{R}_\theta(t,s)_{11}\nabla_{w_2^0}\bar{\ell}_s(\bar{r}_1^s, \bar{w}_2^0; z)_1\right).
\end{aligned}
\tag{387}
$$

By Theorem 1, there exists a unique solution to the equations in the previous display.

We can simplify the above equations. In particular, integrating the last line and adding $\nabla_{w_2^0}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1$ to both sides gives

$$
\begin{aligned}
&\nabla_{w_2^0}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 + \int_0^t \Big(\frac{\partial \bar{\ell}_t(\bar{r}^t;z)}{\partial \bar{w}^s}\Big)_{12}\mathrm{d}s\\
&= \nabla_{r_1}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 + \int_0^t \nabla_{r_1}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 \cdot \left(-\frac{1}{\delta}\int_s^t \bar{R}_\theta(t,s')_{11}\Big(\frac{\partial \bar{\ell}_{s'}(\bar{r}^{s'};z)}{\partial \bar{w}^s}\Big)_{12}\mathrm{d}s' - \frac{1}{\delta}\bar{R}_\theta(t,s)_{11}\nabla_{w_2^0}\bar{\ell}_s(\bar{r}_1^s, \bar{w}_2^0; z)_1\right)\mathrm{d}s\\
&= \nabla_{w_2^0}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 - \frac{1}{\delta}\nabla_{r_1}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 \cdot \int_0^t \left(\bar{R}_\theta(t,s)_{11}\left(\nabla_{w_2^0}\bar{\ell}_s(\bar{r}_1^s, \bar{w}_2^0; z)_1 + \int_0^s \Big(\frac{\partial \bar{\ell}_s(\bar{r}^s;z)}{\partial \bar{w}^{s'}}\Big)_{12}\mathrm{d}s'\right)\right)\mathrm{d}s.
\end{aligned}
\tag{388}
$$

Define

$$
\frac{\partial \bar{\ell}(\bar{r}_1^t, \bar{w}^*; z)}{\partial \bar{w}^*} := \nabla_{w_2^0}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 + \int_0^t \Big(\frac{\partial \bar{\ell}_t(\bar{r}^t;z)}{\partial \bar{w}^s}\Big)_{12}\mathrm{d}s. \tag{389}
$$

Then we get

$$
\frac{\partial \bar{\ell}(\bar{r}_1^t, \bar{w}^*; z)}{\partial \bar{w}^*} = -\frac{1}{\delta}\nabla_{r_1}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1 \cdot \int_0^t \bar{R}_\theta(t,s)_{11}\frac{\partial \bar{\ell}(\bar{r}_1^s, \bar{w}^*; z)}{\partial \bar{w}^*}\mathrm{d}s + \nabla_{w_2^0}\bar{\ell}_t(\bar{r}_1^t, \bar{w}_2^0; z)_1. \tag{390}
$$

Taking expectations, we get

$$\bar{\Gamma}^t_{12} + \int_0^t \bar{R}_\ell(t,s)_{12}\mathrm{d}s = \mathbb{E}\Big[\frac{\partial \bar{\ell}(\bar{r}^t_1, \bar{w}^*; z)}{\partial \bar{w}^*}\Big]. \tag{391}$$

We thus see that the state evolution equations (3.1) and (13) applied to $\bar{\Lambda}^t$ and $\bar{\ell}_t$ as in Eq. (385) gives the planted state evolution Eqs. (31) and (32) under the change of variables (with the notation appearing in Eq. (31) and (32) on the right)

$$\begin{gathered}
\theta^t = \bar{\theta}^t_1, \qquad \Lambda^t = \bar{\Lambda}^t_{11}, \qquad \Gamma^t = \bar{\Gamma}^t_{11}, \qquad R_\ell(t,s) = \bar{R}_\ell(t,s)_{11} \\
R_\ell(t,*) = \bar{\Gamma}^t_{12} + \int_0^t \bar{R}_\ell(t,s)_{12}\mathrm{d}s, \qquad u^t = \bar{u}^t_1, \qquad r^t = \bar{r}^t_1, \qquad R_\theta(t,s) = \bar{R}_\theta(t,s)_{11}, \\
w^* = w^0_2, \qquad w^t = w^t_1, \\
\frac{\partial \theta^t}{\partial u^s} = \Big(\frac{\partial \bar{\theta}^t}{\partial \bar{u}^s}\Big)_{11}, \qquad \ell_s(r^s, w^*; z) = \bar{\ell}_s(r^s, w^*; z)_1, \qquad R_\ell(t,s) = \bar{R}_\ell(t,s)_{11}, \\
\frac{\partial \ell_t(r^t, w^*; z)}{\partial w^s} = \Big(\frac{\partial \bar{\ell}_t(\bar{r}^t; z)}{\partial \bar{w}^s}\Big)_{11}, \qquad C_\theta(t,s) = \bar{C}_\theta(t,s)_{11}, \quad 0 \le s \le t, \\
C_\theta(t,*) = \bar{C}_\theta(t,0)_{12}, \qquad C_\theta(*,*) = \bar{C}_\theta(0,0)_{22}.
\end{gathered} \tag{392}$$

Note that $\bar{R}_\theta(t,s)_{12}$, though defined by the state evolution equations (3.1) and (13), plays no role in the dynamics of Eqs. (31) and (32), so is omitted.

In summary, we have shown that the unique solution to Eqs. (3.1) and (13)with inputs (385) gives a solution to Eqs. (31) and (32). Thus, we have shown existence of a solution to these equations. Uniqueness requires a few more steps of argumentation. We have already shown that (387) have a unique solution. Note that any solution to Eqs. (31) and (32) generates a solution to (387) using the change of variables in the previous display, as well as setting

$$\begin{gathered}
\bar{\Gamma}^t_{12} = \mathbb{E}\Big[\nabla_{w^*}\ell_t(r^t, w^*; z)\Big], \qquad \bar{R}_\ell(t,s)_{12} = \mathbb{E}\Big[\Big(\frac{\partial \bar{\ell}_t(\bar{r}^t; z)}{\partial \bar{w}^s}\Big)_{12}\Big], \\
\Big(\frac{\partial \bar{\ell}_t(\bar{r}^t; z)}{\partial \bar{w}^s}\Big)_{12} = \nabla_{r_1}\bar{\ell}_t(\bar{r}^t_1, \bar{w}^0_2; z)_1 \cdot \Bigg(-\frac{1}{\delta}\int_s^t \bar{R}_\theta(t,s')_{11}\Big(\frac{\partial \bar{\ell}_{s'}(\bar{r}^{s'}; z)}{\partial \bar{w}^s}\Big)_{12}\mathrm{d}s' - \frac{1}{\delta}\bar{R}_\theta(t,s)_{11}\nabla_{w^0_2}\bar{\ell}_s(\bar{r}^s_1, \bar{w}^0_2; z)_1\Bigg).
\end{gathered} \tag{393}$$

By (388), we have that Eq. (32c) is satisfied with $\nabla_{w^*}\ell_t(r^s, w^*; z) + \int_0^s \Big(\frac{\partial \bar{\ell}_t(\bar{r}^{s'}; z)}{\partial \bar{w}^{s'}}\Big)_{12}\mathrm{d}s'$ in place of $\frac{\partial \ell(r^t, w^*; z)}{\partial w^*}$. This implies that

$$\frac{\partial \ell(r^s, w^*; z)}{\partial w^*} = \nabla_{w^*}\ell_t(r^s, w^*; z) + \int_0^s \Big(\frac{\partial \bar{\ell}_t(\bar{r}^{s'}; z)}{\partial \bar{w}^{s'}}\Big)_{12}\mathrm{d}s', \tag{394}$$

and that $R_\ell(t,*) = \bar{\Gamma}^t_{12} + \int_0^t \bar{R}_\ell(t,s)_{12}\mathrm{d}s$. We have thus generated from Eqs. (31) and (32) a solution to Eqs. (387). Because distinct solutions to Eqs. (31) and (32) will generate distinct solutions to Eqs. (387), and the solution to Eqs. (387) is unique, we conlude the solution to Eqs. (31) and (32) is unique.

D.2 Proof of Theorem 3: convergence to fixed points

Proof of Theorem 3. Throughout the proof, we will repeatedly use that for $\mathsf{x} \in \{\theta, \ell\}$,

$$\lim_{t\to\infty}\int_0^t \|R_\mathsf{x}(t, t-s) - R_\mathsf{x}(s)\|\mathrm{d}s = 0. \tag{395}$$

Indeed, for any $t \ge \Delta \ge 0$, we have $\int_0^t \|R_\mathsf{x}(t,t-s) - R_\mathsf{x}(s)\|\mathrm{d}s \le \Delta\|R_\mathsf{x}(t,t-\cdot) - R_\mathsf{x}(\cdot)\|_\infty + 2Ce^{-c\Delta}$. The previous display follows by taking $t \to \infty$ followed by $\Delta \to \infty$.

Theorem 3 will hold for

$$R^\infty_\ell = \Gamma + \int_0^\infty R_\ell(s)\mathrm{d}s, \qquad R^\infty_\theta = \int_0^\infty R_\theta(s)\mathrm{d}s. \tag{396}$$

We begin by establishing Eq. (38). Note that as $t \to \infty$,

$$\|(\Lambda + \Gamma^t)\theta^t - (\Lambda + \Gamma^\infty)\theta^\infty\|_{L^2} \le \|\Lambda + \Gamma^\infty\| \|\theta^t - \theta^\infty\|_{L^2} + \|\Gamma - \Gamma^t\| \|\theta^t\|_{L^2} \to 0, \tag{397}$$

and

$$\begin{aligned} &\Big\| \int_0^t R_\ell(t, t-s)\theta^{t-s}\mathrm{d}s - \int_0^t R_\ell(s)\theta^\infty \mathrm{d}s \Big\|_{L^2} \\ &\quad \le \int_0^t \|R_\ell(t,t-s)\| \|\theta^{t-s} - \theta^\infty\|_{L^2}\mathrm{d}s + \int_0^t \|R_\ell(t,t-s) - R_\ell(s)\|\mathrm{d}s\, \|\theta^\infty\|_{L^2} \\ &\quad \le Ce^{-cs}Ce^{-c(t-s)} + \int_0^t \|R_\ell(t,t-s) - R_\ell(s)\|\mathrm{d}s\, \|\theta^\infty\|_{L^2} \to 0, \end{aligned} \tag{398}$$

and

$$\|R_\ell(t,*)\theta^* - R_\ell^*\theta^*\|_{L^2} \to 0, \qquad \|u^t - u^\infty\|_{L^2} \to 0. \tag{399}$$

Combining these bounds, we conclude that

$$\Big\| \frac{\mathrm{d}}{\mathrm{d}t}\theta^t - \Big(-(\Lambda + \Gamma^\infty)\theta^\infty - \int_0^t R_\ell(s)\theta^\infty \mathrm{d}s - R_\ell^*\theta^* + u^\infty \Big) \Big\|_{L^2} \to 0. \tag{400}$$

Because $\Gamma^\infty + \int_0^t R_\ell(s)\theta^\infty \mathrm{d}s \xrightarrow{L^2} R_\ell^\infty \theta^\infty$, we conclude that

$$\frac{\mathrm{d}}{\mathrm{d}t}\theta^t \xrightarrow{\mathrm{L}^2} -(\Lambda + R_\ell^\infty)\theta^\infty - R_\ell^*\theta^* + u^\infty. \tag{401}$$

Because θ^t stays bounded in L^2 as $t \to \infty$, we must have that

$$0 = -(\Lambda + R_\ell^\infty)\theta^\infty - R_\ell^*\theta^* + u^\infty. \tag{402}$$

Similarly, as $t \to \infty$

$$\begin{aligned} &\Big\| \int_0^t R_\theta(t,t-s)\ell(r^{t-s}, w^*; z)\mathrm{d}s - \int_0^t R_\theta(s)\ell(r^\infty, w^*; z)\mathrm{d}s \Big\|_{L^2} \\ &\quad \le \int_0^t \|R_\theta(t,t-s)\| \|\ell(r^{t-s}, w^*; z) - \ell(r^\infty, w^*; z)\|_{L^2}\mathrm{d}s + \int_0^t \|R_\theta(t,t-s) - R_\theta(s)\|\mathrm{d}s\, \|\ell(r^\infty, w^*; z)\|_{L^2} \\ &\quad \le CLe^{-cs}Ce^{-c(t-s)} + \int_0^t \|R_\theta(t,t-s) - R_\theta(s)\|\mathrm{d}s\, \|\ell(r^\infty, w^*; z)\|_{L^2} \to 0. \end{aligned} \tag{403}$$

Because $\int_0^t R_\theta(t,t-s)\mathrm{d}s \to R_\theta(s)\mathrm{d}s$, we conclude that

$$r^\infty = -\frac{1}{\delta}R_\theta^\infty r^\infty + w^\infty. \tag{404}$$

Because $u^t \xrightarrow{L^2} u^\infty$ and $r^t \xrightarrow{L^2} r^\infty$, we have $u^\infty \sim \mathsf{N}(0, C_\ell^\infty/\delta)$, where

$$C_\ell^\infty = \lim_{t\to\infty} C_\ell(t,t) = \lim_{t\to\infty} \mathbb{E}[\ell(r^t, w^*; z)\ell(r^t, w^*; z)^\top] = \mathbb{E}[\ell(r^\infty, w^*; z)\ell(r^\infty, w^*; z)^\top]. \tag{405}$$

Likewise, because $(w^t, w^*) \xrightarrow{L^2} (w^\infty, w^*)$ and $(\theta^t, \theta^*) \xrightarrow{L^2} (\theta^\infty, \theta^*)$, we have $(w^\infty, w^*) \sim \mathsf{N}(0, C_\theta^\infty)$, where

$$C_\theta^\infty = \lim_{t\to\infty} C_\theta(\{t,*\},\{t,*\}) = \lim_{t\to\infty} \mathbb{E}[(\theta^{t\top}, \theta^{*\top})^\top (\theta^{t\top}, \theta^{*\top})] = \mathbb{E}[(\theta^{\infty\top}, \theta^{*\top})^\top (\theta^{\infty\top}, \theta^{*\top})]. \tag{406}$$

We have finished the proof of Eq. (38) and that $u^\infty \sim \mathsf{N}(0, C_\ell/\delta)$ and $(w^\infty, w^*) \sim \mathsf{N}(0, C_\theta)$.

We now show that $R_\ell^\infty, R_\theta^\infty, R_\ell^*$ satisfy Eq. (39). By Eq. (13a), $\frac{\partial \theta^t}{\partial u^s}$ is deterministic, so $R_\theta(t,s) = \frac{\partial \theta^t}{\partial u^s}$. First compute

$$\frac{\mathrm{d}}{\mathrm{d}s}\int_0^s R_\theta(t+s, t+s')\mathrm{d}s' = I_k + \int_0^s \Big(-(\Lambda + \Gamma^t)R_\theta(t+s, t+s') - \int_{t+s'}^{t+s} R_\ell(t+s, s'')R_\theta(s'', t+s')\mathrm{d}s'' \Big)\mathrm{d}s', \tag{407}$$

where we have used Eq. (13a), and that we may exchanged differentiation and integration by Definition 4.2 (using the boundendess of the derivative). Taking $s \to \infty$, the right-hand side converges to

$$I_k - (\Lambda + \Gamma^\infty) \int_0^\infty R_\theta(s')\mathrm{d}s' - \int_0^\infty R_\ell(s)\mathrm{d}s \int_0^\infty R_\theta(s)\mathrm{d}s, \tag{408}$$

where to get the second term, we have used that

$$\int_0^s \int_{t+s'}^{t+s} R_\ell(t+s,s'')R_\theta(s'',t+s')\mathrm{d}s''\mathrm{d}s' = \int_0^s \int_0^{s-s'} R_\ell(t+s,t+s-s')R_\theta(t+s-s',t+s-s'-s'')\mathrm{d}s''\mathrm{d}s', \tag{409}$$

and taken $t \to \infty$ followed by $s \to \infty$, and used the convergence and decay conditions of R_ℓ, R_θ. Beecause $R_\theta(t+s,t+s') \le Ce^{-cs'}$, we must have that $\int_0^s R_\theta(t+s,t+s')\mathrm{d}s'$ converges as $s \to \infty$, which implies that

$$0 = I_k - (\Lambda + \Gamma^\infty) \int_0^\infty R_\theta(s')\mathrm{d}s' - \int_0^\infty R_\ell(s)\mathrm{d}s \int_0^\infty R_\theta(s)\mathrm{d}s = I_k - \Lambda R_\theta^\infty - R_\ell^\infty R_\theta^\infty. \tag{410}$$

(because otherwise, we would have that $\int_0^s R_\theta(t+s,t+s')\mathrm{d}s'$ diverges). This gives us the second equation in Eq. (39).

Now define

$$\widehat{R}_\ell^{(1)}(s) = \nabla_r \ell(r^\infty, w^*; z) \cdot \left(-\frac{1}{\delta} \int_0^s R_\theta(s-s')\widehat{R}_\ell(s')\mathrm{d}s' - \frac{1}{\delta} R_\theta(s) \nabla_r \ell(r^\infty, w^*; z) \right). \tag{411}$$

We bound

$$\begin{aligned}
&\left\| \int_t^{t+s} R_\theta(t+s,s') \frac{\partial \ell(r^{s'}, w^*; z)}{\partial w^s} \mathrm{d}s' - \int_0^s R_\theta(s-s')\widehat{R}_\ell(s')\mathrm{d}s' \right\|_{L_2} \\
&\quad \le \int_t^{t+s} \|R_\theta(t+s,s')\| \left\| \frac{\partial \ell(r^{s'}, w^*; z)}{\partial w^s} - \widehat{R}_\ell(s') \right\|_{L_2} \mathrm{d}s' + \int_t^{t+s} \|\widehat{R}_\ell(s')\|_{L_2} \|R_\theta(t+s,s') - R_\theta(s-s')\|\mathrm{d}s' \\
&\quad \le Cs \int_t^{t+s} \left\| \frac{\partial \ell(r^{s'}, w^*; z)}{\partial w^s} - \widehat{R}_\ell(s') \right\|_{L_2} \mathrm{d}s' + C\|R_\theta(t+s,t+s-\cdot) - R_\theta(\cdot)\|_\infty \to 0.
\end{aligned} \tag{412}$$

where the limit is for s fixed and $t \to \infty$. One can likewise show that as $t \to \infty$, $R_\theta(t+s,t)\nabla_r \ell(r^t, w^*; z) \xrightarrow{L_2} R_\theta(s)\nabla_r \ell(r^\infty, w^*; z)$ and $\nabla_r \ell(r^{t+s}, w^*; z) \xrightarrow{L_2} \nabla_r \ell(r^\infty, w^*; z)$. Because each of these terms is also bounded, using Eq. (411) we conclude that $\frac{\partial \ell(r^{t+s}, w^*; z)}{\partial w^t} \xrightarrow{L_2} \widehat{R}_\ell^{(1)}(s)$. Then, we must have that $\widehat{R}_\ell^{(1)}(s) = \widehat{R}_\ell(s)$. In particular, $\widehat{R}_\ell(s)$ satisfies the equation

$$\widehat{R}_\ell(s) = \nabla_r \ell(r^\infty, w^*; z) \cdot \left(-\frac{1}{\delta} \int_0^s R_\theta(s-s')\widehat{R}_\ell(s')\mathrm{d}s' - \frac{1}{\delta} R_\theta(s) \nabla_r \ell(r^\infty, w^*; z) \right), \tag{413}$$

and moreover,

$$\mathbb{E}[\widehat{R}_\ell(s)] = \lim_{t \to \infty} \mathbb{E}\left[\frac{\partial \ell(r^{t+s}, w^*; z)}{\partial w^t} \right] = \lim_{t \to \infty} R_\ell(t+s,s) = R_\ell(s). \tag{414}$$

Because $R_\theta(s), \widehat{R}_\ell(s) \le Ce^{-cs}$, we may integrate Eq. (413) and apply Fubini's theorem to get

$$\int_0^\infty \widehat{R}_\ell(s)\mathrm{d}s = \nabla_r \ell(r^\infty, w^*; z) \cdot \left(-\frac{1}{\delta} \int_0^\infty R_\theta(s)\mathrm{d}s \int_0^\infty \widehat{R}_\ell(s)\mathrm{d}s - \frac{1}{\delta} \int_0^\infty R_\theta(s)\mathrm{d}s \nabla_r \ell(r^\infty, w^*; z) \right). \tag{415}$$

Recalling the definition of R_θ (Eq. (396)), this can be rearrnaged to

$$\begin{aligned}
\nabla_r \ell(r^\infty, w^*; z) + \int_0^\infty \widehat{R}_\ell(t)\mathrm{d}t &= \left(I_k + \frac{1}{\delta} \nabla_r \ell(r^\infty, w^*; z) R_\theta^\infty \right)^{-1} \nabla_r \ell(r^\infty, w^*; z) \\
&= \delta\left(I_k - \left(I_k + \frac{1}{\delta} \nabla_r \ell(r^\infty, w^*; z) R_\theta^\infty \right)^{-1} \right) (R_\theta^\infty)^{-1}.
\end{aligned} \tag{416}$$

Taking expectations and using Eq. (396) gives the first equation in (39).

Now consider Eq. (32c). Recall

$$\frac{\partial \ell(r^t, w^*; z)}{\partial w^*} = -\frac{1}{\delta}\nabla_r \ell(r^t, w^*; z) \int_0^t R_\theta(t, t-s) \frac{\partial \ell(r^{t-s}, w^*; z)}{\partial w^*} \mathrm{d}s + \nabla_{w^*} \ell(r^t, w^*; z). \tag{417}$$

Because $\nabla_r \ell(r^t, w^*; z) \overset{L_2}{\nabla_r} \ell(r^\infty, w^*; z)$, $\|R_\theta(t, t-\cdot) - R_\theta(\cdot)\|_\infty \to 0$, $\frac{\partial \ell(r^{t-s}, w^*; z)}{\partial w^*} \overset{L_2}{\to} \widehat{R}_\ell^*$, $\nabla_{w^*} \ell(r^t, w^*; z) \overset{L_2}{\nabla_{w^*}} \ell(r^\infty, w^*; z)$, and we have that $\|\nabla_r \ell(r^t, w^*; z)\|, \|\nabla_r \ell(r^\infty, w^*; z)\|, \|\nabla_{w^*} \ell(r^t, w^*; z)\|, \|\nabla_{w^*} \ell(r^\infty, w^*; z)\| \leq M_\ell$, and $\|R_\theta(t, t-s)\|, \|R_\theta(s)\| \leq Ce^{-ts}$, we can take the limit of the previous display as $t \to \infty$ to get

$$\widehat{R}_\ell^* = -\frac{1}{\delta}\nabla_r \ell(r^\infty, w^*; z) \int_0^\infty R_\theta(s) \mathrm{d}s \widehat{R}_\ell^* + \nabla_{w^*} \ell(r^\infty, w^*; z). \tag{418}$$

This can be rearranged to

$$\widehat{R}_\ell^* = \Big(I_k + \frac{1}{\delta}\nabla_r \ell(r^\infty, w^*; z) R_\theta^\infty\Big)^{-1} \nabla_{w^*} \ell(r^\infty, w^*; z). \tag{419}$$

Taking expectations gives the third equation in Eq. (39). This completes the proof. □

D.3 Fixed-point equations for logistic regression

Note that under the change of variables in (51), we have

$$\eta_{(R_\theta)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty) \to \eta[\lambda \mathcal{L}(\cdot, w^*; z)](w^\infty), \tag{420}$$

where the absence of a subscript on η denotes taking the subscript to be 1. In establishing the correspondence to [SC19], we will frequently use the identities that $\mathbb{P}(y|w^*) = \rho'(yw^*) = 1 - \rho'(-yw^*)$ and $\eta[\lambda\rho](w^\infty) = -\eta[\lambda\rho(-\,\cdot)](-w^\infty)$, which are straightforward to check.

Note that in this case, η is not almost differentiabily in w^* because φ is not continuous in w^*, z. Nevertheless, we will formally apply Gaussian integration by parts. For example, we will write

$$\begin{aligned} &\mathbb{E}[w^* \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty)] \\ &\; = \mathrm{Var}(w^*) \mathbb{E}[\partial_{w^*} \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty)] + \mathrm{Cov}(w^*, w^\infty) \mathbb{E}[\partial_{w^\infty} \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty)], \end{aligned} \tag{421}$$

even though the derivative with respect to w^* is not well defined. Our goal is simply to show the formal equivalence between the fixed-point equations Eq. (38) and (39), even though Theorem 3 does not directly apply to the case of logistic regression. One could extend our result to logistic regression via differentiable approximations of φ, though we do not pursue this here. Alternatively, the work of [SC19] and our calculations below show that by formally applying Gaussian integration by parts to Eqs. (38) and (39), we arrive at fixed point equations which correctly characterize the behavior of the logistic regression estimate.

With this discussion in mind, we now pursue establishing the equivalence between Eqs. (38) and (39) of the current work and Eq. (5) of [SC19]. By Eq. (38),

$$\begin{aligned} C_\ell &= \mathbb{E}[\rho'(-yr^\infty)^2] \\ &= \mathbb{E}[\mathbb{P}(y = 1|w^*)\rho'(-\eta[\lambda\rho(-\,\cdot)](w^\infty))^2 + \mathbb{P}(y = -1|w^*)\rho'(\eta[\lambda\rho](w^\infty))^2] \\ &= \mathbb{E}[\lambda\rho'(w^*)\rho'(-\eta[\lambda\rho(-\,\cdot)](w^\infty))^2 + \rho'(-w^*)\rho'(\eta[\lambda\rho](w^\infty))^2] \\ &= \mathbb{E}[\lambda\rho'(w^*)\rho'(\eta[\lambda\rho](-w^\infty))^2 + \rho'(-w^*)\rho'(\eta[\lambda\rho](w^\infty))^2] \\ &= 2\mathbb{E}[\lambda\rho'(w^*)\rho'(\eta[\lambda\rho](-w^\infty))^2]. \end{aligned} \tag{422}$$

This agrees with the first equation in Eq. (5) of [SC19] under the change of variables above.

By the first two equations in Eq. (39),

$$
\begin{aligned}
1 &= \delta(1 - \mathbb{E}[\partial_{w^\infty} \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty)) \\
&= \delta - \delta \mathbb{E}\Big[\frac{1}{1 + \ell'(r^\infty, w^*; z)/(\delta R_\ell^\infty)}\Big] \\
&= \delta - \delta \mathbb{E}\Big[\rho'(w^*) \frac{1}{1 + \lambda\rho''(-\eta[\lambda\rho(-\cdot)](w^\infty))} - \rho'(-w^*) \frac{1}{1 + \lambda\rho''(\eta[\lambda\rho](w^\infty))}\Big] \\
&= \delta - \delta \mathbb{E}\Big[\rho'(w^*) \frac{1}{1 + \lambda\rho''(\eta[\lambda\rho](-w^\infty))} - \rho'(-w^*) \frac{1}{1 + \lambda\rho''(\eta[\lambda\rho](w^\infty))}\Big] \\
&= \delta - \delta \mathbb{E}\Big[\frac{2\rho'(w^*)}{1 + \lambda\rho''(\eta[\lambda\rho](-w^\infty))}\Big],
\end{aligned} \tag{423}
$$

which agrees with the last equation in Eq. (5) of [SC19] under the change of variables above.

By Theorem 3, we have $\mathrm{Var}(w^*) = \mathrm{Var}(\theta^*)$ and $\mathrm{Cov}(w^*, w^\infty) = \mathrm{Cov}(\theta^*, \theta^\infty) = -(R_\ell^*/R_\ell^\infty)\mathrm{Var}(\theta^*)$, where in the last equality we use that $\theta^\infty = (u^\infty - R_\ell^*\theta^*)/R_\ell^\infty$. Recall also by the first equation in Eq. (39) that $\mathbb{E}[\partial_{w^\infty} \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty)] = 1 - 1/\delta$. Thus, we have[3]

$$
\begin{aligned}
&\mathbb{E}[w^*(w^\infty - \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty))] \\
&= \mathrm{Var}(\theta^*)\Big(-(R_\ell^*/R_\ell^\infty) - \mathbb{E}[\partial_{w^*} \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty)] + (R_\ell^*/R_\ell^\infty)(1 - 1/\delta)\Big).
\end{aligned} \tag{424}
$$

We see that the third fixed-point equation is equivalent to

$$
\mathbb{E}[w^*(w^\infty - \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty))] = 0. \tag{425}
$$

By a standard proximal operator identity, we may rewrite this as

$$
\begin{aligned}
0 &= \mathbb{E}[w^*(w^\infty - \eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty))] \\
&= \mathbb{E}[w^* \ell(\eta_{(R_\theta^\infty)^{-1}}[\mathcal{L}(\cdot, w^*; z)/\delta](w^\infty), w^*; z)]/(R_\ell^\infty \delta) \\
&= \lambda \mathbb{E}[-w^*\rho'(w^*)\rho'(-\eta[\lambda\rho(-\cdot)](w^\infty)) + w^*\rho'(-w^*)\rho'(\eta[\lambda\rho](w^\infty))] \\
&= \lambda \mathbb{E}[-w^*\rho'(w^*)\rho'(\eta[\lambda\rho](-w^\infty)) + w^*\rho'(-w^*)\rho'(\eta[\lambda\rho](w^\infty))] \\
&= -2\lambda \mathbb{E}[w^*\rho'(w^*)\rho'(\eta[\lambda\rho](-w^\infty))].
\end{aligned} \tag{426}
$$

Thus, we get the second equation in Eq. (5) of [SC19] under the change of variables (51).

[3] Recall that we apply Gaussian integration by parts formally