arXiv:2504.15556v1 [math.ST] 22 Apr 2025

Dynamical mean-field analysis of adaptive Langevin diffusions: Propagation-of-chaos and convergence of the linear response

Zhou Fan[*], Justin Ko[†], Bruno Loureiro[‡], Yue M. Lu[§], Yandi Shen[¶]

Abstract

Motivated by an application to empirical Bayes learning in high-dimensional regression, we study a class of Langevin diffusions in a system with random disorder, where the drift coefficient is driven by a parameter that continuously adapts to the empirical distribution of the realized process up to the current time. The resulting dynamics take the form of a stochastic interacting particle system having both a McKean-Vlasov type interaction and a pairwise interaction defined by the random disorder. We prove a propagation-of-chaos result, showing that in the large system limit over dimension-independent time horizons, the empirical distribution of sample paths of the Langevin process converges to a deterministic limit law that is described by dynamical mean-field theory. This law is characterized by a system of dynamical fixed-point equations for the limit of the drift parameter and for the correlation and response kernels of the limiting dynamics. Using a dynamical cavity argument, we verify that these correlation and response kernels arise as the asymptotic limits of the averaged correlation and linear response functions of single coordinates of the system. These results enable an asymptotic analysis of an empirical Bayes Langevin dynamics procedure for learning an unknown prior parameter in a linear regression model, which we develop in a companion paper.

Contents

[*]Department of Statistics and Data Science, Yale University
[†]Department of Statistics and Actuarial Science, University of Waterloo
[‡]Departement d'Informatique, École Normale Supérieure, PSL & CNRS
[§]Departments of Electrical Engineering and Applied Mathematics, Harvard University
[¶]Department of Statistics and Data Science, Carnegie Mellon University

1 Introduction

Let $\boldsymbol{\theta} = (\theta_1, \ldots, \theta_d) \in \mathbb{R}^d$ be a system of d interacting particles, evolving according to a stochastic dynamics

$$\mathrm{d}\boldsymbol{\theta}^t = \Big[-\beta \mathbf{X}^\top \mathbf{X}\boldsymbol{\theta}^t + \big(s(\theta_j^t, \widehat{\alpha}^t)\big)_{j=1}^d\Big]\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{b}^t, \qquad \frac{\mathrm{d}}{\mathrm{d}t}\widehat{\alpha}^t = \mathcal{G}\Big(\widehat{\alpha}^t, \frac{1}{d}\sum_{j=1}^d \delta_{\theta_j^t}\Big). \tag{1}$$

Here $\mathbf{X} \in \mathbb{R}^{n\times d}$ is a matrix of random disorder, and $s(\,\cdot\,, \widehat{\alpha}^t) : \mathbb{R} \to \mathbb{R}$ in the drift coefficient is a nonlinear function driven by a stochastic time-dependent parameter $\widehat{\alpha}^t \in \mathbb{R}^K$ that adapts to the past history $\{\boldsymbol{\theta}^s\}_{s\in[0,t]}$. (We defer formal definitions and conditions for the functions $s(\cdot)$ and $\mathcal{G}(\cdot)$ to Section 2.) We will study the pathwise convergence of the empirical measure $\frac{1}{d}\sum_{j=1}^d \delta_{\theta_j^t}$ and of the parameter $\widehat{\alpha}^t$ to deterministic limits as $n, d \to \infty$ at a fixed rate, in this model (1) and in a closely related statistical model.

In the setting of $\beta = 0$, i.e. with no random disorder, the dynamics (1) take a pathwise-exchangeable form as studied classically by [1,2], where the evolution $\mathrm{d}\theta_j^t$ of each j^{th} particle depends on the remaining particles only via the empirical law $\frac{1}{d}\sum_{j=1}^d \delta_{\{\theta_j^s\}_{s\in[0,t]}}$. The convergence of this law in the asymptotic limit $d \to \infty$, together with a resulting asymptotic decoupling of low-dimensional marginals of $\{\boldsymbol{\theta}^s\}_{s\in[0,t]}$, is commonly referred to as propagation-of-chaos. We refer to the classical monographs [3, 4] for a detailed treatment of such models, and to [5, 6] and [7, 8] for modern surveys and examples of recent quantitative convergence results.

The study of propagation-of-chaos for dynamics with random disorder ($\beta \neq 0$) has also a separate and rich development in the literature, using techniques of dynamical mean-field theory (DMFT). DMFT was initially developed to study Langevin dynamics in the soft Sherrington-Kirkpatrick (SK) model [9, 10] and related spherical p-spin models in spin glass theory [11–13], and relied on deep but non-rigorous techniques of the dynamical cavity method [14, 15] and generating functional methods [16–18] of statistical physics. In recent years, DMFT has been applied to shed insight into the learning dynamics in an increasingly wide range of statistical and machine learning models, including matrix and tensor PCA [19–23], phase retrieval and generalized linear models [15, 24–26], Gaussian mixture classification [27, 28], and deep neural networks [29–33].

Pioneering work of [34–37] established the first mathematical formalizations of DMFT in variants of the SK model, in the forms of large deviations principles for the empirical distributions of sample paths. Mathematical results for spherical models were subsequently obtained in [38, 39], and universality of such results with respect to the law of the disorder in [40, 41]. Recently, [42] developed a different and innovative new approach to formalizing DMFT via time discretization and reduction to Approximate Message Passing schemes [43–45], and applied this to derive a DMFT limit for gradient flow dynamics in statistical multi-index models. A related strategy via iterative Gaussian conditioning was developed in [46], which extended the results of [42] to a class of discrete-time Langevin and stochastic gradient dynamics. Non-asymptotic analyses of the entrywise behavior of such dynamics were obtained in [47].

In this work, we will prove a DMFT approximation for the dynamics (1), which has both the above elements of a pathwise-exchangeable interaction driven by the empirical law, as well as a pairwise interaction driven by random disorder. Our motivation is the study of a variant of Langevin dynamics for posterior sampling in a statistical linear model

$$\mathbf{y} = \mathbf{X}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}, \qquad \theta_j^* \overset{iid}{\sim} g(\,\cdot\,, \alpha^*), \qquad \varepsilon_i \overset{iid}{\sim} \mathcal{N}(0, \sigma^2),$$

where the regression coefficients of interest $\theta_1^*, \ldots, \theta_d^*$ are distributed according to a prior law $g(\,\cdot\,, \alpha^*)$ that has an unknown parameter $\alpha^* \in \mathbb{R}^K$. To implement empirical Bayes learning of α^* [48–50], a Langevin dynamics method was introduced in [51][1] which, from an initial estimate or guess $\widehat{\alpha}^0$, evolves a prior parameter estimate

$$\frac{\mathrm{d}}{\mathrm{d}t}\widehat{\alpha}^t = \frac{1}{d}\sum_{j=1}^d \nabla_\alpha \log g(\theta_j^t, \widehat{\alpha}^t) \tag{2}$$

based on the coordinates of a coupled Langevin diffusion

$$\mathrm{d}\boldsymbol{\theta}^t = \nabla_{\boldsymbol{\theta}}\Big(-\frac{1}{2\sigma^2}\|\mathbf{y}-\mathbf{X}\boldsymbol{\theta}^t\|_2^2 + \sum_{j=1}^d \log g(\theta_j^t, \widehat{\alpha}^t)\Big)\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{b}^t \tag{3}$$

that samples from the posterior law $\mathsf{P}(\boldsymbol{\theta} \mid \mathbf{X}, \mathbf{y})$. Such dynamics comprise a minor extension of (1) (which motivates our choice to study a disorder matrix having the covariance form $\mathbf{X}^\top\mathbf{X}$), and we state in (4–5) of Section 2 an extended general dynamics that encompasses this application. We defer a more detailed discussion and analysis of this specific empirical Bayes procedure to our companion paper [52], focusing in this work on the formalization of the limiting dynamics in a general context.

We summarize the main contributions of our paper as follows:

1. Adapting and building upon the methods of [42], we prove a DMFT limit for the dynamics (1) (with a natural extension to the dynamics (4–5) to follow). This will take the form of almost-sure convergence to certain deterministic limits, as $n, d \to \infty$,

$$\{\widehat{\alpha}^t\}_{t\in[0,T]} \to \{\alpha^t\}_{t\in[0,T]}, \quad \frac{1}{d}\sum_{j=1}^d \delta_{\{\theta_j^t\}_{t\in[0,T]}} \overset{W_2}{\to} \mathsf{P}(\{\theta^t\}_{t\in[0,T]}), \quad \frac{1}{n}\sum_{i=1}^n \delta_{\{\eta_i^t\}_{t\in[0,T]}} \overset{W_2}{\to} \mathsf{P}(\{\eta^t\}_{t\in[0,T]})$$

for the sample path of $\widehat{\alpha}^t$ and for the empirical laws of sample paths of $\boldsymbol{\theta}^t$ and $\boldsymbol{\eta}^t = \mathbf{X}\boldsymbol{\theta}^t$.

Each limit $\mathsf{P}(\{\theta^t\}_{t\in[0,T]})$ and $\mathsf{P}(\{\eta^t\}_{t\in[0,T]})$ represents the law of a univariate stochastic process, which is driven by the above limit $\{\alpha^t\}_{t\in[0,T]}$ for the evolving drift parameter, an additional Gaussian process representing the mean field, and an integrated response. These Gaussian processes and integrated responses are described by correlation and response kernels $C_\theta, C_\eta, R_\theta, R_\eta$, where

$$\{\alpha^t\}, C_\theta, C_\eta, R_\theta, R_\eta$$

are defined self-consistently from the laws $\mathsf{P}(\{\theta^t\}_{t\in[0,T]})$ and $\mathsf{P}(\{\eta^t\}_{t\in[0,T]})$ via a system of dynamical fixed-point equations. We establish that this dynamical fixed point is unique in a certain domain of functions with exponential growth.

2. We show that the dynamics (1) admit a well-defined notion of a linear response function $R_{AB}(t,s)$ for a class of observables $A, B : \mathbb{R}^d \to \mathbb{R}$, where $R_{AB}(t,s)$ represents a linear response of $A(\boldsymbol{\theta}^t)$ to a perturbation of the drift coefficient at a previous time s by $\nabla B(\boldsymbol{\theta}^s)$.

We then verify that the above DMFT correlation and response kernels $C_\theta, C_\eta, R_\theta, R_\eta$ arise as the mean-field limits of averages of the correlation and linear response functions for the "single-particle" coordinate observables $A(\boldsymbol{\theta}), B(\boldsymbol{\theta}) = \theta_j$ and $A(\boldsymbol{\theta}), B(\boldsymbol{\theta}) = [\mathbf{X}\boldsymbol{\theta} - \mathbf{y}]_i$ of the high-dimensional system.

Our methods and analyses in the first contribution above follow the approach of [42]. We incorporate into the dynamical fixed-point system a deterministic limit $\{\alpha^t\}_{t\in[0,T]}$ for the trajectory of the stochastic drift parameter $\{\widehat{\alpha}^t\}_{t\in[0,T]}$, extend the analyses to encompass processes with more irregular sample paths resulting from the additional Brownian diffusion term $\mathrm{d}\mathbf{b}^t$ in the dynamics, and simplify the approach in [42] for embedding a discrete-time DMFT system into a continuous-time limit.

Our second contribution above is, to our knowledge, novel in the mathematical literature on DMFT (although anticipated by statistical physics derivations of the DMFT equations). To understand R_θ, R_η as asymptotic limits of averaged single-particle linear response functions, we formalize a dynamical cavity calculation that analyzes the response of a single coordinate θ_j^t to a perturbation of θ_j^s at a preceding time s, via a DMFT approximation of the cavity system with this coordinate left out. This result will be

[1] [51] proposed a nonparametric variant of this method, and we simplify our discussion here to a parametric formulation

important to our companion work [52], allowing us to transfer a fluctuation-dissipation theorem [53] from the high-dimensional dynamics to the DMFT correlation and response kernels $C_\theta, C_\eta, R_\theta, R_\eta$. This will then allow us to carry out an analysis of the long-time behavior of the DMFT equations in an approximately time-translation-invariant setting, and to show convergence of the prior parameter estimate $\widehat{\alpha}^t$ in the above empirical Bayes dynamics to a stationary point of a replica-symmetric limit for the model free energy.

Acknowledgments

This research was initiated during the "Huddle on Learning and Inference from Structured Data" at ICTP Trieste in 2023. We'd like to thank the huddle organiers Jean Barbier, Manuel Sáenz, Subhabrata Sen, and Pragya Sur for their hospitality and many helpful discussions. We are also grateful to Pierfrancesco Urbani, Francesca Mignacco, Emanuele Troiani, and Andrea Montanari for useful discussions. Z. Fan was supported in part by NSF DMS-2142476 and a Sloan Research Fellowship. J. Ko was supported in part by the Natural Sciences and Engineering Research Council of Canada (NSERC), the Canada Research Chairs programme, and the Ontario Research Fund [RGPIN-2020-04597, DGECR-2020-00199]. B. Loureiro was supported by the French government, managed by the National Research Agency (ANR), under the France 2030 program with the reference "ANR-23-IACL-0008" and the Choose France - CNRS AI Rising Talents program. Y. M. Lu was supported in part by the Harvard FAS Dean's Competitive Fund for Promising Scholarship and by a Harvard College Professorship.

Notational conventions

Constants $C, C', c, c' > 0$ are independent of the dimensions n, d unless otherwise specified, and may depend on the time horizon T, dimension K of the drift parameter, and scalar parameter $\beta \in \mathbb{R}$.

In a separable and complete normed vector space $(\mathcal{M}, \|\cdot\|)$, for any $p \geq 1$, $\mathcal{P}_p(\mathcal{M})$ is the space of probability distributions P on $(\mathcal{M}, \|\cdot\|)$ such that $\mathbb{E}_{\xi\sim\mathsf{P}}\|\xi\|^p < \infty$, $W_p(\cdot)$ is the Wasserstein-p metric on $\mathcal{P}_p(\mathcal{M})$, and $\mathsf{P}_n \overset{W_p}{\to} \mathsf{P}$ denotes $W_p(\mathsf{P}_n, \mathsf{P}) \to 0$ as $n \to \infty$. For a random variable ξ in $\mathcal{M}$, we will use $\mathsf{P}(\xi)$ to denote its law. For a vector $\mathbf{x} \in \mathcal{M}^n$, $\widehat{\mathsf{P}}(\mathbf{x}) = \frac{1}{n}\sum_{i=1}^n \delta_{x_i} \in \mathcal{P}_p(\mathcal{M})$ (for any $p \geq 1$) denotes the empirical distribution of the coordinates $x_1, \ldots, x_n$ of $\mathbf{x}$.

On a Euclidean space $\mathbb{R}^d$, $\|\cdot\|$ without subscript is, by convention, the Euclidean (i.e. ℓ_2) norm. $C([0,T], \mathbb{R})$ is the space of continuous functions $f : [0,T] \to \mathbb{R}$ equipped with the norm of uniform convergence $\|f\|_\infty = \sup_{t\in[0,T]} |f(t)|$. $\mathbb{Z}_+ = \{0, 1, 2, \ldots\}$ denotes the nonnegative integers, and $\mathbb{R}_+ = [0, \infty)$ denotes the nonnegative reals. For a function $f : \mathbb{R}^d \to \mathbb{R}$, $\nabla f(\mathbf{x}) \in \mathbb{R}^d$ and $\nabla^2 f(\mathbf{x}) \in \mathbb{R}^{d\times d}$ are its gradient and Hessian at $\mathbf{x}$. For $f : \mathbb{R}^d \to \mathbb{R}^m$, $\mathrm{d}f(\mathbf{x}) \in \mathbb{R}^{d\times m}$ is its derivative at $\mathbf{x}$. $\operatorname{Tr} M$, $\|M\|_{\mathrm{op}}$, and $\|M\|_F$ are the matrix trace, Euclidean operator norm, and Frobenius norm.

2 Model and main results

2.1 Model and dynamics

Let $\mathbf{X} \in \mathbb{R}^{n\times d}$ be a random matrix with independent entries, and $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon} \in \mathbb{R}^n$ the observations of a linear model with regression design $\mathbf{X}$, regression coefficients $\boldsymbol{\theta}^* \in \mathbb{R}^d$, and noise $\boldsymbol{\varepsilon} \in \mathbb{R}^n$.

Let $s : \mathbb{R} \times \mathbb{R}^K \to \mathbb{R}$ be a Lipschitz drift function, $\mathcal{G} : \mathbb{R}^K \times \mathcal{P}_2(\mathbb{R}) \to \mathbb{R}^K$ a Lipschitz gradient map (where $\mathcal{P}_2(\mathbb{R})$ is the space of probability measures on $\mathbb{R}$ with finite second moment), $\{\mathbf{b}^t\}_{t\geq 0}$ a standard Brownian motion on $\mathbb{R}^d$, and $\beta \in \mathbb{R}$ a scalar parameter. We will study the dynamics

$$\mathrm{d}\boldsymbol{\theta}^t = \Big[-\beta\,\mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}^t - \mathbf{y}) + \big(s(\theta_j^t, \widehat{\alpha}^t)\big)_{j=1}^d\Big]\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{b}^t \tag{4}$$

$$\mathrm{d}\widehat{\alpha}^t = \mathcal{G}\Big(\widehat{\alpha}^t, \frac{1}{d}\sum_{j=1}^d \delta_{\theta_j^t}\Big)\mathrm{d}t \tag{5}$$

with initial conditions

$$(\boldsymbol{\theta}^0, \widehat{\alpha}^0) \in \mathbb{R}^d \times \mathbb{R}^K. \tag{6}$$

This encompasses the general dynamics (1) and the application (2–3) under a unified model: Specializing to $\boldsymbol{\theta}^* = 0$ and $\boldsymbol{\varepsilon} = 0$ (hence $\mathbf{y} = 0$) recovers (1), while specializing to $\beta = \sigma^{-2}$, $s(\theta, \alpha) = \partial_\theta \log g(\theta, \alpha)$, and $\mathcal{G}(\alpha, \mathsf{P}) = \mathbb{E}_{\theta \sim \mathsf{P}}[\nabla_\alpha \log g(\theta, \alpha)]$ recovers (2–3). We will refer to these general dynamics (4–5) as an adaptive Langevin diffusion.

We impose the following assumptions on the components of the above model and dynamics throughout this work.

Assumption 2.1 (Model and initial conditions)**.**

(a) (Asymptotic scaling) $\lim_{n,d\to\infty} \frac{n}{d} = \delta \in (0, \infty)$.

(b) (Random design) $\mathbf{X} = (x_{ij}) \in \mathbb{R}^{n \times d}$ has independent entries satisfying $\mathbb{E}x_{ij} = 0$, $\mathbb{E}x_{ij}^2 = \frac{1}{d}$, and $\|\sqrt{d} x_{ij}\|_{\psi_2} \le C$ for a constant $C > 0$ where $\| \cdot \|_{\psi_2}$ is the sub-gaussian norm.

(c) (Linear model and initial conditions) $\boldsymbol{\theta}^0, \boldsymbol{\theta}^*, \boldsymbol{\varepsilon}$ are independent of $\mathbf{X}$, and $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$. For some probability distributions $\mathsf{P}(\theta^*, \theta^0)$ and $\mathsf{P}(\varepsilon)$ having finite moment generating functions in a neighborhood of 0, and for each fixed $p \ge 1$, the entries of $\boldsymbol{\theta}^0, \boldsymbol{\theta}^*, \boldsymbol{\varepsilon}$ satisfy the Wasserstein-p convergence almost surely as $n, d \to \infty$,

$$\frac{1}{d} \sum_{j=1}^{d} \delta_{(\theta_j^*, \theta_j^0)} \overset{W_p}{\to} \mathsf{P}(\theta^*, \theta^0), \qquad \frac{1}{n} \sum_{i=1}^{n} \delta_{\varepsilon_i} \overset{W_p}{\to} \mathsf{P}(\varepsilon). \tag{7}$$

For a deterministic parameter $\alpha^0 \in \mathbb{R}^K$, almost surely $\lim_{n,d\to\infty} \widehat{\alpha}^0 = \alpha^0$.

Assumption 2.2 (Drift function)**.** $s : \mathbb{R} \times \mathbb{R}^K \to \mathbb{R}$ is twice continuously-differentiable, and for some constant $C > 0$ and all $(\theta, \alpha) \in \mathbb{R} \times \mathbb{R}^K$,

$$|s(\theta, \alpha)| \le C(1 + |\theta| + \|\alpha\|_2), \quad \|\nabla_{(\theta,\alpha)} s(\theta, \alpha)\|_2, \|\nabla^2_{(\theta,\alpha)} s(\theta, \alpha)\|_{\text{op}} \le C. \tag{8}$$

Assumption 2.3 (Gradient map)**.** Let $\widehat{\mathsf{P}}(\boldsymbol{\theta}) = d^{-1} \sum_{j=1}^{d} \delta_{\theta_j}$ denote the empirical distribution of coordinates of $\boldsymbol{\theta} \in \mathbb{R}^d$, and let $\mathcal{G}_k : \mathbb{R}^K \to \mathbb{R}$ be the k^{th} component of $\mathcal{G}$.

(a) For some constant $C > 0$ and all $(\alpha, \mathsf{P}) \in \mathbb{R}^K \times \mathcal{P}_2(\mathbb{R})$,

$$\|\mathcal{G}(\alpha, \mathsf{P})\|_2 \le C(1 + \|\alpha\|_2 + \mathbb{E}_{\mathsf{P}}[\theta^2]^{1/2}), \quad \|\mathcal{G}(\alpha, \mathsf{P}) - \mathcal{G}(\alpha', \mathsf{P}')\|_2 \le C\big(\|\alpha - \alpha'\|_2 + W_2(\mathsf{P}, \mathsf{P}')\big). \tag{9}$$

(b) For each $k = 1, \ldots, K$, $(\boldsymbol{\theta}, \alpha) \mapsto \mathcal{G}_k(\alpha, \widehat{\mathsf{P}}(\boldsymbol{\theta}))$ is twice continuously-differentiable, and for some constant $C > 0$ and all $(\boldsymbol{\theta}, \alpha) \in \mathbb{R}^d \times \mathbb{R}^K$,

$$\|\nabla_\alpha \mathcal{G}_k(\alpha, \widehat{\mathsf{P}}(\boldsymbol{\theta}))\|_2 \le C, \qquad \sqrt{d}\, \|\nabla_{\boldsymbol{\theta}} \mathcal{G}_k(\alpha, \widehat{\mathsf{P}}(\boldsymbol{\theta}))\|_2 \le C, \tag{10}$$

$$\max\Big(d\, \|\nabla^2_{\boldsymbol{\theta}} \mathcal{G}_k(\alpha, \widehat{\mathsf{P}}(\boldsymbol{\theta}))\|_{\text{op}},\, \sqrt{d}\, \|\nabla_{\boldsymbol{\theta}} \nabla_\alpha \mathcal{G}_k(\alpha, \widehat{\mathsf{P}}(\boldsymbol{\theta}))\|_{\text{op}},\, \|\nabla^2_\alpha \mathcal{G}_k(\alpha, \widehat{\mathsf{P}}(\boldsymbol{\theta}))\|_{\text{op}}\Big) \le C. \tag{11}$$

Viewing (4–5) as a joint diffusion of $(\boldsymbol{\theta}^t, \widehat{\alpha}^t)$ on $\mathbb{R}^{d+K}$, we remark that (8) and (10) imply that the drift function of this joint diffusion is Lipschitz with respect to the Euclidean norm. Then there exists a unique solution $\{\boldsymbol{\theta}^t, \widehat{\alpha}^t\}_{t\ge0}$ to (4–5) with initial condition (6) that is adapted to the filtration $\mathcal{F}_t := \mathcal{F}(\{\mathbf{b}^s\}_{s\in[0,t]}, \boldsymbol{\theta}^0, \widehat{\alpha}^0, \mathbf{X}, \boldsymbol{\theta}^*, \boldsymbol{\varepsilon})$, which will be the process of interest in our main results.

2.2 Existence and uniqueness of the DMFT fixed point

In this section we define the DMFT limit for the preceding dynamics (4–5). Let $\delta = \lim_{n,d\to\infty} \frac{n}{d}$ and $\alpha^0 = \lim_{n,d\to\infty} \widehat{\alpha}^0$ be as in Assumption 2.1. Let

$$(\theta^*, \theta^0) \sim \mathsf{P}(\theta^*, \theta^0)$$

denote scalar variables with the distribution (7). Let

$$\{b^t\}_{t\ge0}, \qquad \{u^t\}_{t\ge0}$$

be univariate mean-zero Gaussian processes independent of each other and of (θ^*, θ^0), where $\{b^t\}_{t\geq 0}$ is a standard Brownian motion and $\{u^t\}_{t\geq 0}$ has a correlation kernel $C_\eta(\cdot)$ on $[0,\infty)$, defined self-consistently below. Let

$$\{\alpha^t\}_{t\geq 0}$$

be a deterministic continuous process on $\mathbb{R}^K$, also defined self-consistently below. We consider univariate processes $\{\theta^t\}_{t\geq 0}$ and $\{\frac{\partial \theta^t}{\partial u^s}\}_{t\geq s\geq 0}$ adapted to the filtration

$$\mathcal{F}_t^\theta := \mathcal{F}(\{b^s\}_{s\in[0,t]}, \{u^s\}_{s\in[0,t]}, \theta^*, \theta^0)$$

(in the sense that θ^t and $\frac{\partial \theta^t}{\partial u^s}$ for all $s \in [0,t]$ are $\mathcal{F}_t^\theta$-measurable), defined by the stochastic differential equations

$$\mathrm{d}\theta^t = \Big[-\delta\beta(\theta^t - \theta^*) + s(\theta^t, \alpha^t) + \int_0^t R_\eta(t,s)(\theta^s - \theta^*)\mathrm{d}s + u^t\Big]\mathrm{d}t + \sqrt{2}\,\mathrm{d}b^t \text{ with } \theta^t\big|_{t=0} = \theta^0, \quad (12)$$

$$\mathrm{d}\Big(\frac{\partial \theta^t}{\partial u^s}\Big) = \Big[-\Big(\delta\beta - \partial_\theta s(\theta^t, \alpha^t)\Big)\frac{\partial \theta^t}{\partial u^s} + \int_s^t R_\eta(t,s')\frac{\partial \theta^{s'}}{\partial u^s}\mathrm{d}s'\Big]\mathrm{d}t \text{ with } \frac{\partial \theta^t}{\partial u^s}\Big|_{t=s} = 1. \quad (13)$$

We clarify that $\frac{\partial \theta^t}{\partial u^s}$ is a notation for a univariate process on $t \in [s,\infty)$, defined via (13) for each $s \geq 0$.

Consider also

$$\varepsilon \sim \mathsf{P}(\varepsilon), \qquad (w^*, \{w^t\}_{t\geq 0}),$$

where ε is a scalar variable with the distribution (7), and $(w^*, \{w^t\}_{t\geq 0})$ is a univariate mean-zero Gaussian process indexed by $\{*\} \cup [0,\infty)$, independent of ε and with a correlation kernel $C_\theta(\cdot)$ on $\{*\}\cup[0,\infty)$ also defined self-consistently below. We consider univariate processes $\{\eta^t\}_{t\geq 0}$ and $\{\frac{\partial \eta^t}{\partial w^s}\}_{t\geq s\geq 0}$ adapted to the filtration

$$\mathcal{F}_t^\eta := \mathcal{F}(\{w^s\}_{s\leq t}, w^*, \varepsilon),$$

defined by the integral equations

$$\eta^t = -\beta\int_0^t R_\theta(t,s)\big(\eta^s + w^* - \varepsilon\big)\mathrm{d}s - w^t, \quad (14)$$

$$\frac{\partial \eta^t}{\partial w^s} = \beta\Big[-\int_s^t R_\theta(t,s')\frac{\partial \eta^{s'}}{\partial w^s}\mathrm{d}s' + R_\theta(t,s)\Big]. \quad (15)$$

Again $\frac{\partial \eta^t}{\partial u^s}$ is a notation for a univariate process on $t \in [s,\infty)$, defined by (15) for each $s \geq 0$.

The centered Gaussian processes $\{u^t\}_{t\geq 0}$ and $(w^*, \{w^t\}_{t\geq 0})$ above have correlation kernels

$$\mathbb{E}[u^t u^s] = C_\eta(t,s), \qquad \mathbb{E}[w^t w^s] = C_\theta(t,s), \qquad \mathbb{E}[w^t w^*] = C_\theta(t,*), \qquad \mathbb{E}[(w^*)^2] = C_\theta(*,*). \quad (16)$$

Denoting by $\mathsf{P}(\theta^t)$ the law of θ^t solving (12), the above deterministic process $\{\alpha^t\}_{t\geq 0}$ and correlation/response kernels $C_\eta, C_\theta, R_\eta, R_\theta$ are defined for all $t \geq s \geq 0$ self-consistently by

$$\frac{\mathrm{d}}{\mathrm{d}t}\alpha^t = \mathcal{G}(\alpha^t, \mathsf{P}(\theta^t)) \text{ with } \alpha^t\big|_{t=0} = \alpha^0, \quad (17)$$

$$\begin{gathered} C_\theta(t,s) = \mathbb{E}[\theta^t\theta^s], \quad C_\theta(t,*) = \mathbb{E}[\theta^t\theta^*], \quad C_\theta(*,*) = \mathbb{E}[(\theta^*)^2], \\ C_\eta(t,s) = \delta\beta^2\,\mathbb{E}[(\eta^t + w^* - \varepsilon)(\eta^s + w^* - \varepsilon)], \\ R_\theta(t,s) = \mathbb{E}\Big[\frac{\partial \theta^t}{\partial u^s}\Big], \quad R_\eta(t,s) = \delta\beta\,\mathbb{E}\Big[\frac{\partial \eta^t}{\partial w^s}\Big]. \end{gathered} \quad (18)$$

We note that the above process $\{\frac{\partial \eta^t}{\partial w^s}\}_{t\geq s\geq 0}$ defined by (15) is in fact deterministic, but we keep the expectation defining $R_\eta(t,s)$ for symmetry of notation.

The equations (17–18) should be understood as fixed-point equations for $\alpha, C_\theta, C_\eta, R_\theta, R_\eta$, where the laws of the processes $\{\theta^t, \frac{\partial \theta^t}{\partial u^s}, u^t\}_{t\geq s\geq 0}$ and $\{\eta^t, \frac{\partial \eta^t}{\partial w^s}, w^t\}_{t\geq s\geq 0}$ defining (17–18) are in turn defined by $\alpha, C_\theta, C_\eta, R_\theta, R_\eta$ via (12–16). For each fixed time horizon $T > 0$, let $\mathcal{S}(T)$ be a space of functions

$$(\alpha, C_\theta, C_\eta, R_\theta, R_\eta) \equiv \{\alpha^t, C_\theta(t,s), C_\theta(t,*), C_\theta(*,*), C_\eta(t,s), R_\theta(t,s), R_\eta(t,s)\}_{0\leq s\leq t\leq T}$$

having at most exponential growth, and $\mathcal{S}(T)^{\rm cont}$ a subset of continuous such functions, whose precise definitions we defer to Section 3.1 to follow. The following result establishes existence and uniqueness of a fixed point to (17–18) in this space $\mathcal{S}(T)^{\rm cont}$.

Theorem 2.4. *Under Assumptions 2.1, 2.2, and 2.3, for any fixed $T > 0$:*

(a) For any $(\alpha, C_\theta, C_\eta, R_\theta, R_\eta) \in \mathcal{S}(T)$ and any realization of the mean-zero Gaussian processes $\{u^t\}_{t\geq 0}$ and $(w^*, \{w^t\}_{t\geq 0})$ satisfying (16) (independent of $(\theta^*, \theta^0, \{b^t\}_{t\geq 0})$ and ε respectively), there exist unique solutions to (12–13) and (14–15) adapted to $\{\mathcal{F}_t^\theta\}_{t\in[0,T]}$ and $\{\mathcal{F}_t^\eta\}_{t\in[0,T]}$ for times $0 \leq s \leq t \leq T$.

(b) There exists a unique fixed point $(\alpha, C_\theta, C_\eta, R_\theta, R_\eta) \in \mathcal{S}(T)$ satisfying (17–18) for the solution of part (a). This fixed point belongs to $\mathcal{S}(T)^{cont}$, and in particular $\{\alpha^t\}_{t\geq 0}$ is a deterministic continuous process on $\mathbb{R}^K$.

The proof of Theorem 2.4 is given in Section 3. We will call the components of Theorem 2.4(a–b) the unique solution of the DMFT system (12–18).

2.3 The dynamical mean-field approximation

The following is the first main result of our work, showing that the preceding solution to the DMFT system describes the limit of $\{\widehat{\alpha}^t\}_{t\in[0,T]}$ and empirical distributions of coordinates of $\{\boldsymbol{\theta}^t\}_{t\in[0,T]}$ and $\{\boldsymbol{\eta}^t\}_{t\in[0,T]} \equiv \{\mathbf{X}\boldsymbol{\theta}^t\}_{t\in[0,T]}$ solving (4–5), for fixed time horizons $T > 0$ in the limit $n, d \to \infty$.

Theorem 2.5. *Suppose Assumptions 2.1, 2.2, and 2.3 hold. Denote*

$$\boldsymbol{\eta}^t = \mathbf{X}\boldsymbol{\theta}^t, \qquad \boldsymbol{\eta}^* = \mathbf{X}\boldsymbol{\theta}^*$$

let θ^*, ε, $\eta^* = -w^*$, and $\{\theta^t, \eta^t, \alpha^t\}_{t\in[0,T]}$ be the components of the unique solution to the DMFT system (12–18) given by Theorem 2.4, and let $\mathsf{P}(\cdot)$ denote the law of these components. Then for each fixed $T > 0$, almost surely as $n, d \to \infty$,

(a) $(\widehat{\alpha}^t)_{t\in[0,T]} \to (\alpha^t)_{t\in[0,T]}$ in $C([0,T], \mathbb{R}^K)$.

(b) In the sense of Wasserstein-2 convergence over $\mathbb{R} \times C([0,T], \mathbb{R})$ and $\mathbb{R} \times \mathbb{R} \times C([0,T], \mathbb{R})$,

$$\frac{1}{d}\sum_{j=1}^d \delta_{\theta_j^*, \{\theta_j^t\}_{t\in[0,T]}} \overset{W_2}{\to} \mathsf{P}(\theta^*, \{\theta^t\}_{t\in[0,T]}), \qquad \frac{1}{n}\sum_{i=1}^n \delta_{\eta_i^*, \varepsilon_i, \{\eta_i^t\}_{t\in[0,T]}} \overset{W_2}{\to} \mathsf{P}(\eta^*, \varepsilon, \{\eta^t\}_{t\in[0,T]}).$$

The proof of Theorem 2.5 is given in Section 4. For ease of interpretation, we record here two corollaries of this result. The first clarifies an implication of the above Wasserstein-2 convergence in terms of the convergence of pseudo-Lipschitz test functions of finite-dimensional marginals of the processes.

Corollary 2.6. *In the setting of Theorem 2.5, for any fixed $m \geq 1$ and times $t_1, \ldots, t_m \in [0,T]$, and for any pseudo-Lipschitz test functions $f_\theta : \mathbb{R}^{m+1} \to \mathbb{R}$ and $f_\eta : \mathbb{R}^{m+2} \to \mathbb{R}$ (i.e. satisfying $|f(x) - f(y)| \leq C\|x-y\|_2(1 + \|x\|_2 + \|y\|_2)$), almost surely as $n, d \to \infty$,*

$$\begin{aligned} \frac{1}{d}\sum_{j=1}^d f_\theta(\theta_j^*, \theta_j^{t_1}, \ldots, \theta_j^{t_m}) &\to \mathbb{E} f_\theta(\theta^*, \theta^{t_1}, \ldots, \theta^{t_m}) \\ \frac{1}{n}\sum_{i=1}^n f_\eta(\eta_i^*, \varepsilon_i, \eta_i^{t_1}, \ldots, \eta_i^{t_m}) &\to \mathbb{E} f_\eta(\eta^*, \varepsilon, \eta^{t_1}, \ldots, \eta^{t_m}) \end{aligned} \tag{19}$$

where the expectations on the right side are under the joint laws of the solution to the DMFT system.

Proof. Any pseudo-Lipschitz function $(\theta^*, \theta^{t_1}, \ldots, \theta^{t_m}) \mapsto f_\theta(\theta^*, \theta^{t_1}, \ldots, \theta^{t_m})$ is also a pseudo-Lipschitz function of the full sample path $(\theta^*, \{\theta^t\}_{t\in[0,T]}) \in \mathbb{R} \times C([0,T], \mathbb{R})$. Thus the first statement of (19) follows from Theorem 2.5(b) and the characterization of Wasserstein-p convergence in [54, Definition 6.8 and Theorem 6.9], and the second statement follows similarly. □

The second corollary asserts an asymptotic decoupling of the finite-dimensional marginal distributions of $(\boldsymbol{\theta}^*, \{\boldsymbol{\theta}^t\}_{t\in[0,T]})$ in a coordinate-exchangeable setting, which is the usual notion of propagation-of-chaos for interacting particle systems.

Corollary 2.7. *In the setting of Theorem 2.5, suppose in addition that* $(\boldsymbol{\theta}^*, \boldsymbol{\theta}^0) \in \mathbb{R}^{d\times 2}$ *and* $\mathbf{X} \in \mathbb{R}^{n\times d}$ *are both invariant in law under permutations of the coordinates* $\{1, \ldots, d\}$.

Fix any $J \geq 1$, *and let* $\mathsf{P}(\theta^*_{1:J}, \{\theta^t_{1:J}\}_{t\in[0,T]})$ *denote the joint law of sample paths* $(\theta^*_j, \{\theta^t_j\}_{t\in[0,T]}) \in \mathbb{R} \times C([0,T], \mathbb{R})$ *for* $j = 1, \ldots, J$. *Let* $\mathsf{P}(\theta^*, \{\theta^t\}_{t\in[0,T]})^{\otimes J}$ *denote the* J*-fold product of the limit law in Theorem 2.5(b). Then as* $n, d \to \infty$, *in the sense of weak convergence,*

$$\mathsf{P}(\theta^*_{1:J}, \{\theta^t_{1:J}\}_{t\in[0,T]}) \to \mathsf{P}(\theta^*, \{\theta^t\}_{t\in[0,T]})^{\otimes J}.$$

Proof. Under the stated assumptions and the definition of the process (4–5), the law of $(\boldsymbol{\theta}^*, \{\boldsymbol{\theta}^t\}_{t\in[0,T]}) \in (\mathbb{R} \times C([0,T], \mathbb{R}))^d$ remains invariant under permutations of the coordinates $\{1, \ldots, d\}$. Then the stated result is equivalent to convergence of the empirical law $\frac{1}{d}\sum_{j=1}^d \delta_{\theta^*_j, \{\theta^t_j\}_{t\in[0,T]}}$ to $\mathsf{P}(\theta^*, \{\theta^t\}_{t\in[0,T]})$ weakly in probability (c.f. [4, Proposition 2.2]), and this is implied by Theorem 2.5(b). □

We clarify that $\mathsf{P}(\theta^*_{1:J}, \{\theta^t_{1:J}\}_{t\in[0,T]})$ in this statement refers to the law over all randomness including that of $\boldsymbol{\theta}^*, \boldsymbol{\theta}^0$ and the disorder $\mathbf{X}$. It would be interesting to also study propagation-of-chaos phenomena conditional on parts of this randomness, and we leave such investigations to future work.

2.4 Interpretation of the DMFT correlation and response

Fixing $\mathbf{X}, \boldsymbol{\theta}^*, \boldsymbol{\varepsilon}$ and $\mathbf{y} = \mathbf{X}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon}$, define the coordinate observables

$$e_j(\boldsymbol{\theta}) = \theta_j, \qquad x_i(\boldsymbol{\theta}) = \sqrt{\delta}\beta([\mathbf{X}\boldsymbol{\theta}]_i - y_i). \tag{20}$$

Fixing also the initial conditions $\mathbf{x} = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$, for each pair $A, B \in \{e_1, \ldots, e_d, x_1, \ldots, x_n\}$, define

$$\{R^{\mathbf{x}}_{AB}(t,s)\}_{0\leq s\leq t}$$

as a response function for the joint dynamics (4–5) that satisfies the following condition: For any continuous bounded function $h : [0, \infty) \to \mathbb{R}$ and any $\varepsilon > 0$, consider the perturbed dynamics

$$\begin{aligned}
\mathrm{d}\boldsymbol{\theta}^{t,\varepsilon} &= \Big[-\beta\mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}^{t,\varepsilon} - \mathbf{y}) + \varepsilon h(t)\nabla_{\boldsymbol{\theta}} B(\boldsymbol{\theta}^{t,\varepsilon}, \widehat{\alpha}^{t,\varepsilon}) + \big(s(\theta^{t,\varepsilon}_j, \widehat{\alpha}^{t,\varepsilon})\big)_{j=1}^d\Big]\mathrm{d}t + \sqrt{2}\,\mathrm{d}\mathbf{b}^t \\
\mathrm{d}\widehat{\alpha}^{t,\varepsilon} &= \mathcal{G}\Big(\widehat{\alpha}^{t,\varepsilon}, \frac{1}{d}\sum_{j=1}^d \delta_{\theta^{t,\varepsilon}_j}\Big)\mathrm{d}t
\end{aligned}$$

with the same initial condition $(\boldsymbol{\theta}^{0,\varepsilon}, \widehat{\alpha}^{0,\varepsilon}) = \mathbf{x}$. Denote the expectation conditional on $\mathbf{X}, \boldsymbol{\theta}^*, \boldsymbol{\varepsilon}$ and $\mathbf{x} = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$ as $\langle f(\{\boldsymbol{\theta}^t, \widehat{\alpha}^t\}_{t\geq 0})\rangle_{\mathbf{x}}$. Then for any $t > 0$,

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big(\langle A(\boldsymbol{\theta}^{t,\varepsilon}, \widehat{\alpha}^{t,\varepsilon})\rangle_{\mathbf{x}} - \langle A(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\rangle_{\mathbf{x}}\Big) = \int_0^t R^{\mathbf{x}}_{AB}(t,s)h(s)\,ds. \tag{21}$$

Thus $R^{\mathbf{x}}_{AB}(t,s)$ may be understood as the linear response of the observable $A(\boldsymbol{\theta})$ at time t to a perturbation of the Langevin potential by $B(\boldsymbol{\theta})$ at a preceding time s. Existence of such a response function for smooth bounded observables in uniformly elliptic and hypoelliptic diffusions has been shown in [55, 56]. We verify in Proposition A.1 that the arguments of [56] may be extended to show also the existence of a response function $R^{\mathbf{x}}_{AB}(t,s)$ satisfying (21) in our adaptive Langevin diffusion, for a class of unbounded and Lipschitz observables including all $A, B \in \{e_1, \ldots, e_d, x_1, \ldots, x_n\}$.

Let $\{\boldsymbol{\theta}^t, \widehat{\alpha}^t\}_{t\geq 0}$ be the solution to (4–5) with the given initial condition $\mathbf{x} = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$ of Assumption 2.1, and define the corresponding correlation and response matrices

$$\begin{gathered}
\mathbf{C}_\theta(t,s) = \Big(\big\langle e_j(\boldsymbol{\theta}^t)e_k(\boldsymbol{\theta}^s)\big\rangle_{\mathbf{x}}\Big)_{j,k=1}^d, \quad \mathbf{C}_\theta(t,*) = \Big(\big\langle e_j(\boldsymbol{\theta}^t)e_k(\boldsymbol{\theta}^*)\big\rangle_{\mathbf{x}}\Big)_{j,k=1}^d, \quad \mathbf{R}_\theta(t,s) = \Big(R^{\mathbf{x}}_{e_je_k}(t,s)\Big)_{j,k=1}^d, \\
\mathbf{C}_\eta(t,s) = \Big(\big\langle x_j(\boldsymbol{\theta}^t)x_k(\boldsymbol{\theta}^s)\big\rangle_{\mathbf{x}}\Big)_{j,k=1}^n, \quad \mathbf{R}_\eta(t,s) = \Big(R^{\mathbf{x}}_{x_jx_k}(t,s)\Big)_{j,k=1}^n
\end{gathered} \tag{22}$$

for the above coordinate observables $e_1,\ldots,e_d,x_1,\ldots,x_n$. The following is the second main result of our work, showing that the correlation and response kernels $C_\theta, C_\eta, R_\theta, R_\eta$ defining the DMFT limit in Theorem 2.5 are the almost-sure limits of the normalized traces of these matrices, i.e. the correlation and self-responses of the observables e_j and x_i averaged across coordinates $j=1,\ldots,d$ and $i=1,\ldots,n$.

Theorem 2.8. *Suppose Assumptions 2.1, 2.2, and 2.3 hold, and* $(\theta,\alpha)\mapsto\nabla^2_{(\theta,\alpha)}s(\theta,\alpha)$ *and* $(\boldsymbol{\theta},\alpha)\mapsto\nabla^2_{(\boldsymbol{\theta},\alpha)}\mathcal{G}_k(\alpha,\widehat{\mathsf{P}}(\boldsymbol{\theta}))$ *are uniformly Hölder-continuous for each* $k=1,\ldots,K$. *Let* $C_\theta,C_\eta,R_\theta,R_\eta$ *be the correlation and response kernels of the solution to the DMFT system (12–18) given by Theorem 2.4. Then for any fixed* $t\geq s\geq 0$, *almost surely as* $n,d\to\infty$,

$$d^{-1}\operatorname{Tr}\mathbf{C}_\theta(t,s)\to C_\theta(t,s),\qquad d^{-1}\operatorname{Tr}\mathbf{C}_\theta(t,*)\to C_\theta(t,*),\qquad n^{-1}\operatorname{Tr}\mathbf{C}_\eta(t,s)\to C_\eta(t,s),$$
$$d^{-1}\operatorname{Tr}\mathbf{R}_\theta(t,s)\to R_\theta(t,s),\qquad n^{-1}\operatorname{Tr}\mathbf{R}_\eta(t,s)\to R_\eta(t,s).$$

The proof of Theorem 2.8 is provided in Section 5. We note that the convergence of $d^{-1}\operatorname{Tr}\mathbf{C}_\theta$ and $n^{-1}\operatorname{Tr}\mathbf{C}_\eta$ is an immediate consequence of Corollary 2.6. The additional content of this theorem is the convergence of $d^{-1}\operatorname{Tr}\mathbf{R}_\theta$ and $n^{-1}\operatorname{Tr}\mathbf{R}_\eta$, which relies on an inductive analysis of dynamics at a single particle level using a dynamical cavity argument.

Remark 2.9. By an argument similar to our proof of Theorem 2.8, one may show that the DMFT response kernels $R_\theta(t,s)$ and $R_\eta(t,s)$ also represent the limits of $d^{-1}\operatorname{Tr}\mathbf{R}_\theta(t,s)$ and $n^{-1}\operatorname{Tr}\mathbf{R}_\eta(t,s)$ defined for a non-adaptive version of the dynamics

$$\mathrm{d}\tilde{\boldsymbol{\theta}}^t=\Big[-\beta\mathbf{X}^\top(\mathbf{X}\tilde{\boldsymbol{\theta}}^t-\mathbf{y})+\big(s(\tilde{\theta}^t_j,\alpha^t)\big)_{j=1}^d\Big]\mathrm{d}t+\sqrt{2}\,\mathrm{d}\mathbf{b}^t$$

which replaces the adaptively-evolving drift parameter $\{\widehat{\alpha}^t\}_{t\geq0}$ by its deterministic DMFT limit $\{\alpha^t\}_{t\geq0}$. The response matrices $\mathbf{R}_\theta,\mathbf{R}_\eta$ for this non-adaptive dynamics $\{\tilde{\boldsymbol{\theta}}^t\}_{t\geq0}$ are different from those for the adaptive dynamics (4–5), in that a perturbation in the adaptive system affects $\{\widehat{\alpha}^t\}_{t\geq s}$ whereas it does not change $\{\alpha^t\}_{t\geq s}$ in the non-adaptive system. However, our result implies that the almost-sure limits of $d^{-1}\operatorname{Tr}\mathbf{R}_\theta$ and $n^{-1}\operatorname{Tr}\mathbf{R}_\eta$ coincide for these two dynamics, i.e. the propagation of the effect of the perturbation through $\{\widehat{\alpha}^t\}$ is negligible in the large-(n,d) limit.

The remainder of this paper will prove the preceding results of Theorems 2.4, 2.5, and 2.8.

3 Existence and uniqueness of the DMFT fixed point

In this section we prove Theorem 2.4. We assume throughout Assumptions 2.1, 2.2, and 2.3. Section 3.1 defines the spaces $\mathcal{S}(T)$ and $\mathcal{S}(T)^{\mathrm{cont}}$ and proves Theorem 2.4(a) on existence and uniqueness of the processes (12–15). Section 3.2 then proves Theorem 2.4(b) on existence and uniqueness of the dynamical fixed point via a contractive mapping argument similar to that of [42].

3.1 The function spaces $\mathcal{S}(T)$ and $\mathcal{S}(T)^{\mathrm{cont}}$

Let $\tau_*^2 = \mathbb{E}\theta^{*2}$ and $\sigma^2 = \mathbb{E}\varepsilon^2$, and let $C_0 > 0$ denote a constant larger than the constants $C > 0$ of (8) and (9). Consider the following system of equations for functions $\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta}, \Phi_{R_\theta}, \Phi_{R_\eta}$ on $[0, \infty)$:

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_\alpha(t) = 4.1C_0(1 + \Phi_{C_\theta}(t)) + 3C_0\Phi_\alpha(t) \text{ with } \Phi_\alpha(0) = \|\alpha^0\|^2, \tag{23}$$

$$\begin{aligned}\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{C_\theta}(t) &= (6\delta^2\beta^2 + 18C_0^2 + 1.1)\Phi_{C_\theta}(t) + 6\int_0^t (t-s+1)^2\Phi_{R_\eta}^2(t-s)\Phi_{C_\theta}(s)\mathrm{d}s \\ &\quad + 6\Big(\delta^2\beta^2\tau_*^2 + 3C_0^2 + 3C_0^2\Phi_\alpha(t) + \int_0^t (t-s+1)^2\Phi_{R_\eta}^2(t-s)\mathrm{d}s \cdot \tau_*^2 + \Phi_{C_\eta}(t)\Big) + 2 \\ &\quad \text{with } \Phi_{C_\theta}(0) = \mathbb{E}(\theta^0)^2,\end{aligned} \tag{24}$$

$$\Phi_{C_\eta}(t) = 2\delta\beta^2\Big[\frac{1}{\delta}\int_0^t (t-s+1)^2 \cdot \Phi_{R_\theta}^2(t-s)\Phi_{C_\eta}(s)\mathrm{d}s + 2\Phi_{C_\theta}(t) + 2\tau_*^2 + \sigma^2\Big], \tag{25}$$

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{R_\theta}(t) = (\delta|\beta| + C_0)\Phi_{R_\theta}(t) + \int_0^t \Phi_{R_\eta}(t-s)\Phi_{R_\theta}(s)\mathrm{d}s \text{ with } \Phi_{R_\theta}(0) = 1, \tag{26}$$

$$\Phi_{R_\eta}(t) = |\beta|\Big(\int_0^t \Phi_{R_\theta}(t-s)\Phi_{R_\eta}(s)\mathrm{d}s + \delta|\beta|\Phi_{R_\theta}(t)\Big). \tag{27}$$

Lemma 3.1. *The system (23–27) has a unique continuous solution. Defining*

$$E(\lambda) = \Big\{\text{continuous functions } f : \mathbb{R}_+ \to \mathbb{R}_+ \text{ such that } \int_0^\infty e^{-\lambda s}f(s)\mathrm{d}s < \infty\Big\},$$

for any sufficiently large constant $\lambda > 0$, this solution satisfies $\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta}, \Phi_{R_\theta}, \Phi_{R_\eta} \in E(\lambda)$.

Proof. Let $\Phi_\eta = (\Phi_{C_\eta}, \Phi_{R_\eta})$, $\Phi_\theta = (\Phi_\alpha, \Phi_{C_\theta}, \Phi_{R_\theta})$, and $\Phi = (\Phi_\eta, \Phi_\theta)$. For any two continuous solutions Φ and $\tilde{\Phi}$, there exists some $M > 0$ such that all components of both solutions are uniformly bounded over $[0, T]$ by M. The above equations then imply

$$\|\Phi_\theta(t) - \tilde{\Phi}_\theta(t)\| \le \int_0^t C\|\Phi(s) - \tilde{\Phi}(s)\|\mathrm{d}s, \quad \|\Phi_\eta(t) - \tilde{\Phi}_\eta(t)\| \le C\Big(\int_0^t \|\Phi(s) - \tilde{\Phi}(s)\|\mathrm{d}s + \|\Phi_\theta(t) - \tilde{\Phi}_\theta(t)\|\Big)$$

for a constant $C > 0$ depending on M, T. Applying Gronwall's lemma to the second inequality shows

$$\sup_{s\in[0,t]} \|\Phi_\eta(s) - \tilde{\Phi}_\eta(s)\| \le C' \sup_{s\in[0,t]} \|\Phi_\theta(s) - \tilde{\Phi}_\theta(s)\|. \tag{28}$$

Then applying this in the first inequality gives

$$\|\Phi_\theta(t) - \tilde{\Phi}_\theta(t)\| \le \int_0^t C'' \sup_{r\in[0,s]} \|\Phi_\theta(r) - \tilde{\Phi}_\theta(r)\|\mathrm{d}s,$$

so Gronwall's lemma applied again shows $\Phi_\theta(t) = \tilde{\Phi}_\theta(t)$ for all $t \in [0, T]$. Then by (28), also $\Phi(t) = \tilde{\Phi}(t)$ for all $t \in [0, T]$, so any continuous solution to (23–27) is unique.

It remains to show existence of a continuous solution with all components in $E(\lambda)$. Consider (26–27) as a mapping from $\Phi_{R_\theta}, \Phi_{R_\eta}$ on the right side to $\tilde{\Phi}_{R_\theta}, \tilde{\Phi}_{R_\eta}$ on the left side, i.e.

$$\begin{aligned}\tilde{\Phi}_{R_\theta}(t) &= 1 + \int_0^t \Big((\delta|\beta| + C_0)\Phi_{R_\theta}(t') + \int_0^{t'} \Phi_{R_\eta}(t'-s)\Phi_{R_\theta}(s)\mathrm{d}s\Big)\mathrm{d}t', \\ \tilde{\Phi}_{R_\eta}(t) &= |\beta|\Big(\int_0^t \Phi_{R_\theta}(t-s)\Phi_{R_\eta}(s)\mathrm{d}s + \delta|\beta|\Phi_{R_\theta}(t)\Big).\end{aligned}$$

If $\Phi_{R_\theta}, \Phi_{R_\eta} \in E(\lambda)$, then writing $L_\theta(\lambda) = \int_0^\infty \Phi_{R_\theta}(s)e^{-\lambda s}\mathrm{d}s$ for the Laplace transform of Φ_{R_θ} and similarly writing $L_\eta, \tilde{L}_\theta, \tilde{L}_\eta$ for those of $\Phi_{R_\eta}, \tilde{\Phi}_{R_\theta}, \tilde{\Phi}_{R_\eta}$, taking Laplace transforms of the above gives

$$\begin{aligned}\lambda\tilde{L}_\theta(\lambda) - 1 &= (\delta|\beta| + C_0)L_\theta(\lambda) + L_\eta(\lambda)L_\theta(\lambda), \\ \tilde{L}_\eta(\lambda) &= |\beta|L_\theta(\lambda)L_\eta(\lambda) + \delta\beta^2 L_\theta(\lambda).\end{aligned} \tag{29}$$

This implies in particular that $\tilde{L}_\theta(\lambda), \tilde{L}_\eta(\lambda) < \infty$, i.e. $\tilde{\Phi}_{R_\theta}, \tilde{\Phi}_{R_\eta} \in E(\lambda)$. For $\iota > 0$, define further

$$E(\lambda, \iota) = \{(\Phi_{R_\theta}, \Phi_{R_\eta}) : L_\theta(\lambda) \le \iota,\, L_\eta(\lambda) \le (\delta\beta^2 + 1)\iota\}.$$

If $(\Phi_{R_\theta}, \Phi_{R_\eta}) \in E(\lambda, \iota)$, then for $\iota > 0$ sufficiently small and $\lambda > 0$ sufficiently large, this implies $\tilde{L}_\theta(\lambda) \le \lambda^{-1}(1 + (\delta|\beta| + C_0)\iota + (\delta\beta^2 + 1)\iota^2) \le \iota$ and $\tilde{L}_\eta(\lambda) \le |\beta|(\delta\beta^2 + 1)\iota^2 + \delta\beta^2\iota \le (\delta\beta^2 + 1)\iota$, so $(\tilde{\Phi}_{R_\theta}, \tilde{\Phi}_{R_\eta}) \in E(\lambda, \iota)$. For two pairs of inputs (L_θ^1, L_η^1) and (L_θ^2, L_η^2), note by (29) that the corresponding outputs satisfy

$$|\tilde{L}_\theta^1 - \tilde{L}_\theta^2| \le \frac{\delta|\beta| + C_0}{\lambda}|L_\theta^1 - L_\theta^2| + \frac{|L_\eta^1|}{\lambda}|L_\theta^1 - L_\theta^2| + \frac{|L_\theta^2|(\delta\beta^2 + 1)}{\lambda}\frac{|L_\eta^1 - L_\eta^2|}{\delta\beta^2 + 1}$$
$$\frac{|\tilde{L}_\eta^1 - \tilde{L}_\eta^2|}{\delta\beta^2 + 1} \le |\beta||L_\theta^1|\frac{|L_\eta^1 - L_\eta^2|}{\delta\beta^2 + 1} + \frac{|\beta||L_\eta^2|}{\delta\beta^2 + 1}|L_\theta^1 - L_\theta^2| + \frac{\delta\beta^2}{\delta\beta^2 + 1}|L_\theta^1 - L_\theta^2|.$$

Thus, defining a weighted L^1-norm on $E(\lambda) \times E(\lambda)$ given by $\|(\Phi_{R_\theta}, \Phi_{R_\eta})\| = |L_\theta(\lambda)| + (\delta\beta^2 + 1)^{-1}|L_\eta(\lambda)|$, one may check from the above that the mapping $(\Phi_{R_\theta}, \Phi_{R_\eta}) \mapsto (\tilde{\Phi}_{R_\theta}, \tilde{\Phi}_{R_\eta})$ is Lipschitz on $E(\lambda) \times E(\lambda)$ with respect to $\|\cdot\|$, with Lipschitz constant at most

$$\frac{\delta|\beta| + C_0}{\lambda} + \frac{2(\delta\beta^2 + 1)\iota}{\lambda} + 2|\beta|\iota + \frac{\delta\beta^2}{\delta\beta^2 + 1}.$$

This is less than 1 for $\iota > 0$ sufficiently small and $\lambda > 0$ sufficiently large, so $(\Phi_{R_\theta}, \Phi_{R_\eta}) \mapsto (\tilde{\Phi}_{R_\theta}, \tilde{\Phi}_{R_\eta})$ is a contraction with respect to $\|\cdot\|$. This norm is complete on $E(\lambda) \times E(\lambda)$, and $E(\lambda, \iota)$ is closed in $E(\lambda) \times E(\lambda)$, so by the Banach fixed-point theorem, there exists a unique fixed point $(\Phi_{R_\theta}, \Phi_{R_\eta}) \in E(\lambda, \iota) \subset E(\lambda) \times E(\lambda)$ which is a solution to (26–27).

Given this solution to (26–27), consider now (23–25) as a mapping from $(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta})$ on the right side to $(\tilde{\Phi}_\alpha, \tilde{\Phi}_{C_\theta}, \tilde{\Phi}_{C_\eta})$ on the left side. Now let $L_\alpha(\lambda), L_\theta(\lambda), L_\eta(\lambda)$ denote the Laplace transforms of $(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta})$, and define also the Laplace transforms $K_\eta(\lambda) = \int_0^t (s+1)^2 \Phi_{R_\eta}^2(s) e^{-\lambda s} \mathrm{d}s$ and $K_\theta(\lambda) = \int_0^t (s+1)^2 \Phi_{R_\theta}^2(s) e^{-\lambda s} \mathrm{d}s$. Choosing λ large enough so that $K_\eta(\lambda), K_\theta(\lambda) < \infty$, if $\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta} \in E(\lambda)$, then taking Laplace transforms of (23–25) gives

$$\lambda\tilde{L}_\alpha(\lambda) - \|\alpha^0\|^2 = \frac{4.1C_0}{\lambda} + 4.1C_0 L_\theta(\lambda) + 3C_0 L_\alpha(\lambda)$$
$$\lambda\tilde{L}_\theta(\lambda) - \mathbb{E}(\theta^0)^2 = C_1 L_\theta(\lambda) + 6K_\eta(\lambda)L_\theta(\lambda) + 18C_0^2 L_\alpha(\lambda) + \frac{6\tau_*^2}{\lambda}K_\eta(\lambda) + 6L_\eta(\lambda) + \frac{C_2}{\lambda}$$
$$\tilde{L}_\eta(\lambda) = 2\beta^2 K_\theta(\lambda)L_\eta(\lambda) + 4\delta\beta^2 L_\theta(\lambda) + \frac{C_3}{\lambda}$$

for some constants C_1, C_2, C_3 depending only on $\delta, \beta, C_0, \sigma^2, \tau_*^2$. For small $\iota > 0$, suppose further that $(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta}) \in E(\lambda, \iota)$ where

$$E(\lambda, \iota) = \{(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta}) : L_\alpha(\lambda) \le \iota,\, L_\theta(\lambda) \le \iota,\, L_\eta(\lambda) \le (4\delta\beta^2 + 1)\iota\}.$$

Then, using that $\lim_{\lambda\to\infty} K_\theta(\lambda) = 0$ and $\lim_{\lambda\to\infty} K_\eta(\lambda) = 0$, for sufficiently large $\lambda > 0$ and small $\iota > 0$, the above Laplace transform equations imply $(\tilde{\Phi}_\alpha, \tilde{\Phi}_{C_\theta}, \tilde{\Phi}_{C_\eta}) \in E(\lambda, \iota)$. Furthermore, defining the norm $\|(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta})\| = |L_\alpha(\lambda)| + |L_\theta(\lambda)| + (4\delta\beta^2 + 1)|L_\eta(\lambda)|$, it may be verified as above that the mapping $(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta}) \mapsto (\tilde{\Phi}_\alpha, \tilde{\Phi}_{C_\theta}, \tilde{\Phi}_{C_\eta})$ is Lipschitz in $\|\cdot\|$ on $E(\lambda, \iota)$, with Lipschitz constant at most

$$\frac{7.1C_0}{\lambda} + \frac{C_1 + 6K_\eta(\lambda) + 18C_0^2 + 6(4\delta\beta^2 + 1)}{\lambda} + 2\beta^2 K_\theta(\lambda) + \frac{4\delta\beta^2}{4\delta\beta^2 + 1}.$$

For sufficiently large $\lambda > 0$, this is again less than 1, so there exists a unique fixed point $(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta}) \in E(\lambda, \iota) \subset E(\lambda) \times E(\lambda) \times E(\lambda)$ which solves (23–25). □

Let $(\Phi_\alpha, \Phi_{C_\theta}, \Phi_{C_\eta}, \Phi_{R_\theta}, \Phi_{R_\eta})$ be the above solution to (23–27). For any $T > 0$ and finite set $D = \{d_1, \ldots, d_m\} \subset (0, T)$, we call $[0, d_1), [d_1, d_2), \ldots, [d_m, T]$ the maximal intervals of $[0, T] \setminus D$. Fixing $T > 0$ and denoting

$$(\alpha, C_\theta, C_\eta, R_\theta, R_\eta) \equiv \{\alpha^t, C_\theta(t,s), C_\theta(t,*), C_\theta(*,*), C_\eta(t,s), R_\theta(t,s), R_\eta(t,s)\}_{0 \le s \le t \le T},$$

we define the space $\mathcal{S} \equiv \mathcal{S}(T)$ in Theorem 2.4 as

$$\mathcal{S} = \{(\alpha, C_\theta, C_\eta, R_\theta, R_\eta) : (R_\eta, C_\eta, \alpha) \in \mathcal{S}_\eta,\ (R_\theta, C_\theta, \alpha) \in \mathcal{S}_\theta\}. \tag{30}$$

Here $\mathcal{S}_\eta \equiv \mathcal{S}_\eta(T)$ is the collection of (R_η, C_η, α) such that, for some (possibly empty) discontinuity set $D \subset (0, T)$ of at most finite cardinality:

- C_η is a positive-semidefinite covariance kernel on $[0, T]$ (identifying $C_\eta(s,t) = C_\eta(t,s)$) satisfying

$$C_\eta(t,t) \le \Phi_{C_\eta}(t) \text{ for all } 0 \le t \le T. \tag{31}$$

  Furthermore, $C_\eta(t,s)$ is uniformly continuous over $s,t \in I$ for each maximal interval I of $[0,T] \setminus D$, and satisfies

$$\begin{aligned}\big|C_\eta(t,t) - 2C_\eta(t,s) + C_\eta(s,s)\big| \le 3\beta^2 \Bigg[\Big(T^3 \sup_{r\in[0,T]} \Phi'_{R_\theta}(r)^2 + T \sup_{r\in[0,T]} \Phi_{R_\theta}(r)^2\Big) \sup_{r\in[0,T]} \Phi_{C_\eta}(r) \\ + \delta\Big(2T \sup_{r\in[0,T]} \Phi'_{C_\theta}(r) + 4\Big)\Bigg] \cdot |t-s| \text{ for all } s,t \in I.\end{aligned} \tag{32}$$

- $R_\eta(t,s)$ satisfies

$$|R_\eta(t,s)| \le \Phi_{R_\eta}(t-s) \text{ for all } 0 \le s \le t \le T. \tag{33}$$

  Furthermore, $R_\eta(t,s)$ is uniformly continuous over $s \in I'$ and $t \in I$ for any two (possibly equal) maximal intervals I, I' of $[0,T] \setminus D$.

- α^t satisfies

$$\|\alpha^t\|^2 \le \Phi_\alpha(t) \text{ for all } 0 \le t \le T. \tag{34}$$

  and is uniformly continuous on each maximal interval I of $[0,T] \setminus D$.

Similarly $\mathcal{S}_\theta \equiv \mathcal{S}_\theta(T)$ is the set of $(R_\theta, C_\theta, \alpha)$ such that

- C_θ is a positive-semidefinite covariance kernel on $\{*\} \cup [0,T]$ (identifying $C_\theta(s,t) = C_\theta(t,s)$ and $C_\theta(t,*) = C_\theta(*,t)$) satisfying

$$C_\theta(t,t) \le \Phi_{C_\theta}(t) \text{ for all } 0 \le t \le T. \tag{35}$$

  Furthermore, $C_\theta(t,s)$ is uniformly continuous over $s,t \in I$ for each maximal interval I of $[0,T] \setminus D$ and satisfies

$$\big|C_\theta(t,t) - 2C_\theta(t,s) + C_\theta(s,s)\big| \le \Big(2T \sup_{r\in[0,T]} \Phi'_{C_\theta}(r) + 4\Big)|t-s| \text{ for all } s,t \in I, \tag{36}$$

  and $C_\theta(t,*)$ is uniformly continuous over $t \in I$.

- $R_\theta(t,s)$ satisfies

$$|R_\theta(t,s)| \le \Phi_{R_\theta}(t-s) \text{ for all } 0 \le s \le t \le T. \tag{37}$$

  Furthermore, $R_\theta(t,s)$ is uniformly continuous over $s \in I'$ and $t \in I$ for any two (possibly equal) maximal intervals I, I' of $[0,T] \setminus D$, and satisfies

$$\big|R_\theta(t',s) - R_\theta(t,s)\big| \le \Big(\sup_{r\in[0,T]} \Phi'_{R_\theta}(r)\Big)|t'-t| \text{ for each fixed } s \in [0,T] \text{ and all } t,t' \in [s,T] \cap I. \tag{38}$$

- α^t satisfies (34) and is uniformly continuous on each maximal interval I of $[0,T] \setminus D$.

We define

$$\mathcal{S}^{\text{cont}}(T) \equiv \mathcal{S}^{\text{cont}} \subset \mathcal{S}, \qquad \mathcal{S}_\eta^{\text{cont}}(T) \equiv \mathcal{S}_\eta^{\text{cont}} \subset \mathcal{S}_\eta, \qquad \mathcal{S}_\theta^{\text{cont}}(T) \equiv \mathcal{S}_\theta^{\text{cont}} \subset \mathcal{S}_\theta$$

as the subsets of the above spaces where $D = \emptyset$, i.e. the above continuity conditions hold on all of $[0, T]$.

Remark 3.2. By (32), letting $\{u^t\}_{t\in[0,T]}$ be a mean-zero Gaussian process with covariance C_η, for any maximal interval I of $[0,T] \setminus D$, any $s, t \in I$, and some constant $C > 0$,

$$\mathbb{E}(u^t - u^s)^4 = 3[\mathbb{E}(u^t - u^s)^2]^2 \le C|t-s|^2.$$

Then Kolmogorov's continuity theorem ([57, Theorem 2.9]) implies that there exists a modification of $\{u^t\}_{t\in[0,T]}$ that is uniformly Hölder continuous on each such maximal interval I, and similarly for $\{w^t\}_{t\in[0,T]}$ with covariance C_θ satisfying (36). We will always take $\{u^t\}$ and $\{w^t\}$ to be the versions of these processes that satisfy this Hölder continuity.

Let us now establish existence and uniqueness of the solutions to (12–15) given $(\alpha, C_\theta, C_\eta, R_\theta, R_\eta) \in \mathcal{S}$.

Lemma 3.3. *Fix any $T > 0$, any $(R_\eta, C_\eta, \alpha) \in \mathcal{S}_\eta$, and any realizations of $\theta^0, \theta^*, \{b^t\}_{t\le T}$ and $\{u^t\}_{t\le T}$. Then there exist unique $\mathcal{F}_t^\theta$-adapted processes $\{\theta^t\}_{t\le T}$ and $\{\frac{\partial\theta^t}{\partial u^s}\}_{s\le t\le T}$ solving (12–13).*

Proof. Consider the drift function

$$v(t, \{\theta^s\}_{s\le t}) = -\delta\beta(\theta^t - \theta^*) + s(\theta^t, \alpha^t) + \int_0^t R_\eta(t,s)(\theta^s - \theta^*)\mathrm{d}s + u^t.$$

Conditioning on θ^0, θ^* and $\{u^t\}$ and writing 0 for the process $\theta^t \equiv 0$, we have (with probability 1 over θ^0, θ^* and $\{u^t\}$)

$$\sup_{t\in[0,T]} |v(t, 0)| \le \delta|\beta\theta^*| + \sup_{t\in[0,T]} |s(0, \alpha^t)| + \int_0^T \Phi_{R_\eta}(t)\mathrm{d}t \cdot |\theta^*| + \sup_{t\in[0,T]} |u^t| < \infty.$$

Furthermore, for all $t \in [0, T]$,

$$|v(t, \{\theta^s\}_{s\le t}) - v(t, \{\tilde\theta^s\}_{s\le t})| \le \Big(\delta|\beta| + \sup_{(\theta,\alpha)\in\mathbb{R}\times\mathbb{R}^K} |\partial_\theta s(\theta, \alpha)| + \int_0^T \Phi_{R_\eta}(s)\mathrm{d}s\Big) \sup_{s\in[0,t]} |\theta^s - \tilde\theta^s|,$$

showing under Assumption 2.2 that $\{\theta^s\}_{s\le t} \mapsto v(t, \{\theta^s\}_{s\le t})$ is Lipschitz in the norm of uniform convergence, uniformly over $t \in [0, T]$. Then existence and uniqueness of a solution $\{\theta^t\}_{t\le T}$ with $\theta^t|_{t=0} = \theta^0$ adapted to the filtration of $\{b^t\}_{t\ge0}$ is classical, see e.g. [58, Theorem 11.2]. This solution is a measurable function of θ^0, θ^*, and $\{u^t\}$, and hence is also $\mathcal{F}_t^\theta$-adapted.

Conditioning now on $\{\theta^t\}$, for any fixed $s \in [0, T]$, consider

$$v(t, \{x^{s'}\}_{s'\in[s,t]}) = -\Big(\delta\beta - \partial_\theta s(\theta^t, \alpha^t)\Big)x^t + \int_s^t R_\eta(t, s')x^{s'}\mathrm{d}s'.$$

This satisfies $v(t, 0) \equiv 0$ for all $t \in [s, T]$ and

$$|v(t, \{x^{s'}\}_{s'\in[s,t]}) - v(t, \{\tilde x^{s'}\}_{s'\in[s,t]})| \le \Big(\delta|\beta| + \sup_{(\theta,\alpha)\in\mathbb{R}\times\mathbb{R}^K} |\partial_\theta s(\theta, \alpha)| + \int_0^T \Phi_{R_\eta}(t)\mathrm{d}t\Big) \sup_{s'\in[s,t]} |x^{s'} - \tilde x^{s'}|,$$

so $\{x^{s'}\}_{s'\in[s,t]} \mapsto v(t, \{x^{s'}\}_{s'\in[s,t]})$ is also Lipschitz in the norm of uniform convergence, uniformly over $t \in [s, T]$. Then again for each $s \in [0, T]$, there exists a unique solution $\{\frac{\partial\theta^t}{\partial u^s}\}_{t\in[s,T]}$ with $\frac{\partial\theta^t}{\partial u^s}|_{t=s} = 1$, which is adapted to the filtration $\mathcal{F}_t \equiv \mathcal{F}(\{\theta^r\}_{r\in[s,t]})$ and hence also to $\mathcal{F}_t^\theta$, showing the lemma. ☐

Lemma 3.4. *Fix any $T > 0$, any $(R_\theta, C_\theta, \alpha) \in \mathcal{S}_\theta$, and any realizations of ε and $(w^*, \{w^t\}_{t\le T})$. Then there exist unique $\mathcal{F}_t^\eta$-adapted processes $\{\eta^t\}_{t\le T}$ and $\{\frac{\partial\eta^t}{\partial w^s}\}_{s\le t\le T}$ solving (14–15).*

Proof. Conditional on ε and $(w^*, \{w^t\})$, the equations (14–15) are linear Volterra integral equations for which the kernel $(s,t) \mapsto R_\theta(t,s)$ is continuous on each maximal interval I of $[0,T] \setminus D$. Then, for each maximal interval $I = [a,b)$, given the values of $\{\eta^t\}$ for $t \in [0,a]$, existence and uniqueness of $\{\eta^t\}_{t\in[a,b)}$ is classical and follows from e.g. [59, Theorem 2.1.2]. Applying this successively to each maximal interval I shows existence and uniqueness of $\{\eta^t\}$ over $t \in [0,T]$. A similar argument shows, for each fixed $s \in [0,T]$, the existence and uniqueness of $\{\frac{\partial \eta^t}{\partial w^s}\}$ over $t \in [s,T]$. Here $\frac{\partial \eta^t}{\partial w^s}$ is deterministic by its definition, while η^t is a measurable function of $\varepsilon, w^*, \{w^s\}_{s\le t}$ and hence is adapted to $\mathcal{F}_t^\eta$. □

Proof of Theorem 2.4(a). This follows from Lemmas 3.3 and 3.4. □

3.2 Contractive mapping

We fix $T > 0$. For any $(R_\eta, C_\eta, \alpha) \in \mathcal{S}_\eta$, define a map $\mathcal{T}_{\eta\to\theta} : (R_\eta, C_\eta, \alpha) \to (R_\theta, C_\theta, \tilde{\alpha})$ by

$$R_\theta(t,s) = \mathbb{E}\Big[\frac{\partial \theta^t}{\partial u^s}\Big], \quad C_\theta(t,s) = \mathbb{E}[\theta^t\theta^s], \quad C_\theta(t,*) = \mathbb{E}[\theta^t\theta^*], \quad C_\theta(*,*) = \mathbb{E}[(\theta^*)^2],$$
$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\alpha}^t = \mathcal{G}(\tilde{\alpha}^t, \mathsf{P}(\theta^t)) \text{ with } \tilde{\alpha}^t|_{t=0} = \alpha^0$$

where $\{\theta^t\}_{t\in[0,T]}$ and $\{\frac{\partial \theta^t}{\partial u^s}\}_{0\le s\le t\le T}$ are the unique solutions to (12–13) given (R_η, C_η, α) and $\theta^0, \theta^*, \{u^t\}$, guaranteed by Lemma 3.3, and $\mathsf{P}(\theta^t)$ is the law of θ^t. Similarly, for any $(R_\theta, C_\theta, \tilde{\alpha}) \in \mathcal{S}_\theta$, define a map $\mathcal{T}_{\theta\to\eta} : (R_\theta, C_\theta, \tilde{\alpha}) \to (R_\eta, C_\eta, \alpha)$ by

$$R_\eta(t,s) = \delta\beta\,\mathbb{E}\Big[\frac{\partial \eta^t}{\partial w^s}\Big], \quad C_\eta(t,s) = \delta\beta^2\mathbb{E}[(\eta^t + w^* - \varepsilon)(\eta^s + w^* - \varepsilon)], \quad \alpha^t = \tilde{\alpha}^t$$

where $\{\eta^t\}_{t\in[0,T]}$ and $\{\frac{\partial \eta^t}{\partial w^s}\}_{0\le s<t\le T}$ are the unique solutions to (14–15) given $(R_\theta, C_\theta, \tilde{\alpha})$ and $\varepsilon, w^*, \{w^t\}$, guaranteed by Lemma 3.4. Finally, define the composite maps

$$\mathcal{T}_{\eta\to\eta} = \mathcal{T}_{\theta\to\eta} \circ \mathcal{T}_{\eta\to\theta}, \qquad \mathcal{T}_{\theta\to\theta} = \mathcal{T}_{\eta\to\theta} \circ \mathcal{T}_{\theta\to\eta}. \tag{39}$$

The rest of this subsection is divided into two parts:

- (**Part 1**) We show in Lemma 3.5 (resp. Lemma 3.6) that $\mathcal{T}_{\theta\to\eta}$ maps $\mathcal{S}_\theta$ into $\mathcal{S}_\eta$ (resp. $\mathcal{T}_{\eta\to\theta}$ maps $\mathcal{S}_\eta$ into $\mathcal{S}_\theta$).
- (**Part 2**) We equip $\mathcal{S}_\eta$ and $\mathcal{S}_\theta$ with certain metrics and derive the moduli-of-continuity of the maps $\mathcal{T}_{\eta\to\theta}$ and $\mathcal{T}_{\theta\to\eta}$ in Lemmas 3.7 and 3.8, thereby concluding that $\mathcal{T}_{\eta\to\eta}$ and $\mathcal{T}_{\theta\to\theta}$ in (39) are contractions under these metrics.

Lemma 3.5. *$\mathcal{T}_{\theta\to\eta}$ maps $\mathcal{S}_\theta$ into $\mathcal{S}_\eta$, and $\mathcal{S}_\theta^{cont}$ into $\mathcal{S}_\eta^{cont}$.*

Proof. (**Condition for C_η**) Define $\xi^t = \eta^t + w^* - \varepsilon$, so that

$$\xi^t = -\beta \int_0^t R_\theta(t,s)\xi^s \mathrm{d}s - w^t + w^* - \varepsilon$$

and $C_\eta(t,s) = \delta\beta^2\mathbb{E}[\xi^t\xi^s]$. Then by Cauchy-Schwarz,

$$\begin{aligned}
C_\eta(t,t) &\le 2\delta\beta^2\,\mathbb{E}\Big[\beta^2\Big(\int_0^t R_\theta(t,s)\xi^s\mathrm{d}s\Big)^2 + (w^t - w^* + \varepsilon)^2\Big] \\
&\le 2\delta\beta^2\,\Big[\beta^2\int_0^t (t-s+1)^2 \cdot R_\theta(t,s)^2\mathbb{E}(\xi^s)^2\mathrm{d}s \cdot \int_0^t (t-s+1)^{-2}\mathrm{d}s + 2C_\theta(t,t) + 2\tau_*^2 + \sigma^2\Big] \qquad (40) \\
&\le 2\delta\beta^2\Big[\frac{1}{\delta}\int_0^t (t-s+1)^2 \cdot \Phi_{R_\theta}^2(t-s)C_\eta(s,s)\mathrm{d}s + 2\Phi_{C_\theta}(t) + 2\tau_*^2 + \sigma^2\Big].
\end{aligned}$$

Recalling the equation for $\Phi_{C_\eta}(\cdot)$ in (25),

$$\Phi_{C_\eta}(t) = 2\delta\beta^2\Big[\frac{1}{\delta}\int_0^t (t-s+1)^2 \cdot \Phi_{R_\theta}^2(t-s)\Phi_{C_\eta}(s)\mathrm{d}s + 2\Phi_{C_\theta}(t) + 2\tau_*^2 + \sigma^2\Big],$$

Gronwall's inequality implies that $C_\eta(t,t) \le \Phi_{C_\eta}(t)$, showing (31).

We now check (32) on each maximal interval I of $[0,T]\setminus D$ where $D \subset (0,T)$ is the discontinuity set of $\mathcal{S}_\theta$. Note that

$$\begin{aligned}
&C_\eta(t,t) - 2C_\eta(t,s) + C_\eta(s,s) = \delta\beta^2\mathbb{E}[(\xi^t - \xi^s)^2] \\
&= \delta\beta^2\mathbb{E}\Big[\Big(-\beta\int_0^t R_\theta(t,r)\xi^r\mathrm{d}r + \beta\int_0^s R_\theta(s,r)\xi^r\mathrm{d}r - w^t + w^s\Big)^2\Big] \\
&\le 3\delta\beta^2\Big[\beta^2\mathbb{E}\Big(\int_0^s \big(R_\theta(t,r) - R_\theta(s,r)\big)\xi^r\mathrm{d}r\Big)^2 + \beta^2\mathbb{E}\Big(\int_s^t R_\theta(t,r)\xi^r\mathrm{d}r\Big)^2 + \mathbb{E}(w^t - w^s)^2\Big].
\end{aligned}$$

Using $\delta\beta^2\mathbb{E}(\xi^t)^2 = C_\eta(t,t) \le \Phi_{C_\eta}(t)$ established above, together with the continuity conditions (38) for R_θ and (36) for C_θ, for any $s,t \in I$ it holds that

$$\begin{aligned}
\mathbb{E}\Big(\int_0^s \big(R_\theta(t,r) - R_\theta(s,r)\big)\xi^r\mathrm{d}r\Big)^2 &\le s\int_0^s (R_\theta(t,r) - R_\theta(s,r))^2\mathbb{E}(\xi^r)^2\mathrm{d}r \\
&\le \frac{T^2}{\delta\beta^2}\Big(\sup_{r\in[0,T]}\Phi'_{R_\theta}(r)\Big)^2 \cdot \sup_{r\in[0,T]}\Phi_{C_\eta}(r)\cdot|t-s|^2, \\
\mathbb{E}\Big(\int_s^t R_\theta(t,r)\xi^r\mathrm{d}r\Big)^2 &\le (t-s)\int_s^t R_\theta(t,r)^2\mathbb{E}(\xi^r)^2\mathrm{d}r \\
&\le \frac{1}{\delta\beta^2}\sup_{r\in[0,T]}\Phi_{R_\theta}(r)^2 \cdot \sup_{r\in[0,T]}\Phi_{C_\eta}(r)\cdot|t-s|^2, \\
\mathbb{E}(w^t - w^s)^2 &= C_\theta(t,t) - 2C_\theta(t,s) + C_\theta(s,s) \le (2T\sup_{r\in[0,T]}\Phi'_{C_\theta}(r) + 4)\cdot|t-s|.
\end{aligned}$$

Combining these bounds shows (32) over $s,t \in I$. Applying $|\mathbb{E}[\xi^s\xi^t - \xi^{s'}\xi^{t'}]|^2 \le 2\mathbb{E}(\xi^s - \xi^{s'})^2\mathbb{E}(\xi^{t'})^2 + 2\mathbb{E}(\xi^{s'})^2\mathbb{E}(\xi^t - \xi^{t'})^2$, this shows also that $C_\eta(t,s)$ is uniformly continuous over $s,t \in I$. If $(C_\theta, R_\theta, \tilde{\alpha}) \in \mathcal{S}_\theta^{\mathrm{cont}}$, then $D = \emptyset$ so this maximal interval is $I = [0,T]$.

(**Condition for** R_η) By definition, $R_\eta(t,s) = \beta[-\int_s^t R_\theta(t,s')R_\eta(s',s)\mathrm{d}s' + \delta\beta R_\theta(t,s)]$, hence

$$\begin{aligned}
|R_\eta(t,s)| &\le |\beta|\Big(\int_s^t |R_\theta(t,s')||R_\eta(s',s)|\mathrm{d}s' + \delta|\beta R_\theta(t,s)|\Big) \\
&\le |\beta|\Big(\int_0^{t-s} \Phi_{R_\theta}(t-s-s')|R_\eta(s+s',s)|\mathrm{d}s' + \delta|\beta|\Phi_{R_\theta}(t-s)\Big).
\end{aligned}$$

Recalling the equation for Φ_{R_η} in (27),

$$\Phi_{R_\eta}(t-s) = |\beta|\Big(\int_0^{t-s}\Phi_{R_\theta}(t-s-s')\Phi_{R_\eta}(s')\mathrm{d}s' + \delta|\beta|\Phi_{R_\theta}(t-s)\Big),$$

this implies for all $t \in [s,T]$ that $|R_\eta(t,s)| \le \Phi_{R_\eta}(t-s)$, verifying (33). To show uniform continuity on each pair of maximal intervals I, I' defining $\mathcal{S}_\theta$, observe first that for any $s,s' \in I'$ and $\tau \ge 0$ for which

$s+\tau, s'+\tau \in I$,

$$
\begin{aligned}
&|R_\eta(s'+\tau,s') - R_\eta(s+\tau,s)| \\
&\leq |\beta| \cdot \left| \int_s^{s+\tau} R_\theta(s+\tau,r)R_\eta(r,s)\mathrm{d}r - \int_{s'}^{s'+\tau} R_\theta(s'+\tau,r)R_\eta(r,s')\mathrm{d}r \right| + \delta\beta^2 |R_\theta(s'+\tau,s') - R_\theta(s+\tau,s)| \\
&= |\beta| \int_0^\tau \Big| R_\theta(s+\tau,s+r)R_\eta(s+r,s) - R_\theta(s'+\tau,s'+r)R_\eta(s'+r,s') \Big| \mathrm{d}r + \delta\beta^2 |R_\theta(s'+\tau,s') - R_\theta(s+\tau,s)| \\
&\leq |\beta| \int_0^\tau |R_\theta(s+\tau,s+r)| \cdot |R_\eta(s'+r,s') - R_\eta(s+r,s)| \mathrm{d}r \\
&\quad + |\beta| \int_0^\tau |R_\theta(s+\tau,s+r) - R_\theta(s'+\tau,s'+r)| \cdot |R_\eta(s'+r,s')| \mathrm{d}r + \delta\beta^2 |R_\theta(s'+\tau,s') - R_\theta(s+\tau,s)|.
\end{aligned}
$$

Denoting by $o_{|s-s'|}(1)$ an error that converges to 0 uniformly in τ as $|s-s'| \to 0$, observe that the last term above is $o_{|s-s'|}(1)$ by the uniform continuity of R_θ on $I' \times I$. For the second term, writing the range of integration as $[0,\tau] = A \cup B$ where $r \in A$ are the values for which $s+r, s'+r$ belong to a single maximal interval of $[0,T] \setminus D$ and $r \in B$ are the values for which $s+r, s'+r$ belong to two different maximal intervals, the integral over $r \in A$ is $o_{|s-s'|}(1)$ again by the continuity of R_θ, while the integral over $r \in B$ is also $o_{|s-s'|}(1)$ by the boundedness of R_θ, R_η and the bound $|B| \leq C|s-s'|$ for the total length of B. Putting this together,

$$
|R_\eta(s'+\tau,s') - R_\eta(s+\tau,s)| \leq C \int_0^\tau |R_\eta(s'+r,s') - R_\eta(s+r,s)| \mathrm{d}r + o_{|s-s'|}(1).
$$

Since $R_\eta(s',s') = R_\eta(s,s)$, the above and Gronwall's inequality imply that

$$
|R_\eta(s'+\tau,s') - R_\eta(s+\tau,s)| = o_{|s-s'|}(1) \tag{41}
$$

uniformly in τ. Now for any $s \in I'$ and $\tau' \geq \tau \geq 0$ for which $s+\tau, s+\tau' \in I$,

$$
\begin{aligned}
|R_\eta(s+\tau',s) - R_\eta(s+\tau,s)| \leq |\beta| &\int_s^{s+\tau} |R_\theta(s+\tau',r) - R_\theta(s+\tau,r)| |R_\eta(r,s)| \mathrm{d}r \\
&+ \int_{s+\tau}^{s+\tau'} |R_\theta(s+\tau',r)| \cdot |R_\eta(r,s)| \mathrm{d}r + \delta|\beta| |R_\theta(s+\tau',s) - R_\theta(s+\tau,s)|,
\end{aligned}
$$

so the continuity of R_θ and boundedness of R_θ, R_η again imply that

$$
|R_\eta(s+\tau',s) - R_\eta(s+\tau,s)| = o_{|\tau-\tau'|}(1) \tag{42}
$$

uniformly in s. The statements (41) and (42) show that $(s,\tau) \mapsto R_\eta(s+\tau,s)$ is uniformly continuous over $\{(s,\tau) : s \in I', \tau \geq 0, s+\tau \in I\}$, implying uniform continuity of $(s,t) \mapsto R_\eta(t,s)$ over $(s,t) \in I' \times I$. Again if $(C_\theta, R_\theta, \tilde{\alpha}) \in \mathcal{S}_\theta^{\mathrm{cont}}$, then this continuity holds over all of $I = [0,T]$.

(**Condition for** α) By definition, the mapping $\tilde{\alpha} \mapsto \alpha$ under $\mathcal{T}_{\theta\to\eta}$ is the identity, so the required conditions for α hold by those assumed for $\tilde{\alpha}$. □

Lemma 3.6. *$\mathcal{T}_{\eta\to\theta}$ maps $\mathcal{S}_\eta$ into $\mathcal{S}_\theta^{cont}$.*

Proof. (**Condition for** C_θ) To verify (35), denote

$$
v^t = -\delta\beta(\theta^t - \theta^*) + s(\theta^t, \alpha^t) + \int_0^t R_\eta(t,s)(\theta^s - \theta^*)\mathrm{d}s + u^t.
$$

Applying Ito's formula to $(\theta^t)^2$ yields

$$
\frac{\mathrm{d}C_\theta(t,t)}{\mathrm{d}t} = \frac{\mathrm{d}\mathbb{E}(\theta^t)^2}{\mathrm{d}t} = \mathbb{E}[2\theta^t v^t] + 2 \leq 1.1 \cdot \mathbb{E}(\theta^t)^2 + \mathbb{E}(v^t)^2 + 2. \tag{43}
$$

(The bound holds with 1 in place of 1.1, and we enlarge this to 1.1 to accommodate a later discretized version of this computation.) Using an argument similar to (40), and letting $C_0 > 0$ be the constant defining (23–27) which upper bounds $C > 0$ in (8) of Assumption 2.2, we may bound $\mathbb{E}(v^t)^2$ as

$$\mathbb{E}(v^t)^2 \leq 6\Big[\delta^2\beta^2\mathbb{E}(\theta^t)^2 + \delta^2\beta^2\tau_*^2 + \mathbb{E}[s(\theta^t,\alpha^t)^2] + \mathbb{E}\Big(\int_0^t R_\eta(t,s)\theta^s \mathrm{d}s\Big)^2 + \mathbb{E}\Big(\int_0^t R_\eta(t,s)\theta^* \mathrm{d}s\Big)^2 + \mathbb{E}(u^t)^2\Big]$$
$$\leq (6\delta^2\beta^2 + 18C_0^2)\mathbb{E}(\theta^t)^2 + 6\int_0^t (t-s+1)^2\Phi_{R_\eta}^2(t-s)\mathbb{E}(\theta^s)^2\mathrm{d}s$$
$$+ 6\Big(\delta^2\beta^2\tau_*^2 + 3C_0^2 + 3C_0^2\Phi_\alpha(t) + \int_0^t (t-s+1)^2\Phi_{R_\eta}^2(t-s)\mathrm{d}s \cdot \tau_*^2 + \Phi_{C_\eta}(t)\Big). \quad (44)$$

Applying this to (43) and comparing with the equation for Φ_{C_θ} from (24),

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{C_\theta}(t) = (6\delta^2\beta^2 + 18C_0^2 + 1.1)\Phi_{C_\theta}(t) + 6\int_0^t (t-s+1)^2\Phi_{R_\eta}^2(t-s)\Phi_{C_\theta}(s)\mathrm{d}s \quad (45)$$
$$+ 6\Big(\delta^2\beta^2\tau_*^2 + 3C_0^2 + 3C_0^2\Phi_\alpha(t) + \int_0^t (t-s+1)^2\Phi_{R_\eta}^2(t-s)\mathrm{d}s \cdot \tau_*^2 + \Phi_{C_\eta}(t)\Big) + 2,$$

we see that since $C_\theta(0,0) = \Phi_{C_\theta}(0)$, we have $C_\theta(t,t) \leq \Phi_{C_\theta}(t)$.

Next we prove (36) for all $0 \leq s \leq t \leq T$. We have $\theta^t - \theta^s = \int_s^t v^r\,\mathrm{d}r + \sqrt{2}(b^t - b^s)$. Then it holds that

$$C_\theta(t,t) - 2C_\theta(t,s) + C_\theta(s,s) = \mathbb{E}[(\theta^t-\theta^s)^2] \leq 2|t-s|\int_s^t \mathbb{E}(v^r)^2\mathrm{d}r + 4\mathbb{E}(b^t-b^s)^2$$
$$\leq 2|t-s|^2 \sup_{r\in[0,T]} |\Phi'_{C_\theta}(r)| + 4|t-s| \leq \Big(2T\sup_{r\in[0,T]}\Phi'_{C_\theta}(r) + 4\Big)|t-s|,$$

where the second inequality compares (44) to the definition of $\Phi'_{C_\theta}(t)$ in (45). This verifies (36). As in the preceding argument for C_η, this condition (36) and Cauchy-Schwarz implies that $C_\theta(t,s)$ is uniformly continuous over all $0 \leq s \leq t \leq T$, and also that $C_\theta(t,*)$ is uniformly continuous over all $t \in [0,T]$.

(**Condition for** R_θ) Let $\bar{R}_\theta(t,s) = \mathbb{E}|\frac{\partial\theta^t}{\partial u^s}|$ so that $|R_\theta(t,s)| \leq \bar{R}_\theta(t,s)$ by definition. Note that

$$\frac{\mathrm{d}}{\mathrm{d}t}\Big|\frac{\partial\theta^t}{\partial u^s}\Big| \leq \Big|\frac{\mathrm{d}}{\mathrm{d}t}\frac{\partial\theta^t}{\partial u^s}\Big| \leq \big(\delta|\beta| + |\partial_\theta s(\theta^t,\alpha^t)|\big)\Big|\frac{\partial\theta^t}{\partial u^s}\Big| + \int_s^t |R_\eta(t,s')|\Big|\frac{\partial\theta^{s'}}{\partial u^s}\Big|\mathrm{d}s'$$
$$\leq (\delta|\beta| + C_0)\Big|\frac{\partial\theta^t}{\partial u^s}\Big| + \int_s^t \Phi_{R_\eta}(t-s')\Big|\frac{\partial\theta^{s'}}{\partial u^s}\Big|\mathrm{d}s', \quad (46)$$

where $C_0 > 0$ is the constant defining (23–27) which upper bounds $C > 0$ in (8) of Assumption 2.2. Taking expectation on both sides yields

$$\frac{\mathrm{d}}{\mathrm{d}t}\bar{R}_\theta(t,s) \leq (\delta|\beta| + C_0)\bar{R}_\theta(t,s) + \int_0^{t-s} \Phi_{R_\eta}(t-s-s')\bar{R}_\theta(s+s',s)\mathrm{d}s'.$$

Recall the equation for Φ_{R_θ} in (26),

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_{R_\theta}(t-s) = (\delta|\beta| + C_0)\Phi_{R_\theta}(t-s) + \int_0^{t-s} \Phi_{R_\eta}(t-s-s')\Phi_{R_\theta}(s')\mathrm{d}s'.$$

Since $\bar{R}_\theta(s,s) = 1 = \Phi_{R_\theta}(0)$, this implies for all $t \in [s,T]$ that $|R_\theta(t,s)| \leq \bar{R}_\theta(t,s) \leq \Phi_{R_\theta}(t-s)$, verifying (37) for all $0 \leq s \leq t \leq T$.

To show (38) for all $0 \leq s \leq t \leq t' \leq T$, observe that we have

$$|R_\theta(t',s) - R_\theta(t,s)| = \Big|\int_t^{t'} \mathbb{E}\Big[-\Big(\delta\beta - \partial_\theta s(\theta^r,\alpha^r)\Big)\frac{\partial\theta^r}{\partial u^s}\Big]\mathrm{d}r + \int_t^{t'}\Big(\int_s^r R_\eta(r,r')R_\theta(r',s)\mathrm{d}r'\Big)\mathrm{d}r\Big|$$
$$\leq \int_t^{t'}\Big((\delta|\beta| + C_0)\bar{R}_\theta(r,s)\mathrm{d}r + \int_s^r \Phi_{R_\eta}(r-r')\bar{R}_\theta(r',s)\mathrm{d}r'\Big)\mathrm{d}r \leq |t'-t| \cdot \sup_{r\in[0,T]}\Phi'_{R_\theta}(r),$$

verifying (38). In particular, this shows continuity of $\tau \mapsto R_\theta(s+\tau,s)$ uniformly over $s \in [0,T]$ and $\tau \in [0,T-s]$. For continuity in s, observe that

$$\begin{aligned}
&\frac{\mathrm{d}}{\mathrm{d}\tau}\Big|\frac{\partial\theta^{s+\tau}}{\partial u^s} - \frac{\partial\theta^{s'+\tau}}{\partial u^{s'}}\Big| \\
&\leq (\delta|\beta| + C_0)\Big|\frac{\partial\theta^{s+\tau}}{\partial u^s} - \frac{\partial\theta^{s'+\tau}}{\partial u^{s'}}\Big| + \int_0^\tau \Big|R_\eta(s+\tau,s+r)\frac{\partial\theta^{s+r}}{\partial u^s} - R_\eta(s'+\tau,s'+r)\frac{\partial\theta^{s'+\tau}}{\partial u^{s'}}\Big|\mathrm{d}r.
\end{aligned}$$

We may again divide the range of integration of the second term as $[0,\tau] = A \cup B$ where $s+r, s'+r$ belong to the same maximal interval defining $\mathcal{S}_\eta$ for $r \in A$, and to two different maximal intervals for $r \in B$. Then taking expectations on both sides above and applying boundedness of $\bar{R}_\theta, R_\eta$, continuity of R_η to bound the integral over $r \in A$, and $|B| \leq C|s-s'|$ to bound the integral over $r \in B$, this shows

$$\frac{\mathrm{d}}{\mathrm{d}\tau}\mathbb{E}\Big|\frac{\partial\theta^{s+\tau}}{\partial u^s} - \frac{\partial\theta^{s'+\tau}}{\partial u^{s'}}\Big| \leq C\left(\mathbb{E}\Big|\frac{\partial\theta^{s+\tau}}{\partial u^s} - \frac{\partial\theta^{s'+\tau}}{\partial u^{s'}}\Big| + \int_0^\tau \mathbb{E}\Big|\frac{\partial\theta^{s+r}}{\partial u^s} - \frac{\partial\theta^{s'+\tau}}{\partial u^{s'}}\Big|\mathrm{d}r\right) + o_{|s-s'|}(1)$$

where $o_{|s-s'|}(1)$ converges to 0 uniformly in τ as $|s-s'| \to 0$. Then, since $\mathbb{E}|\frac{\partial\theta^s}{\partial u^s} - \frac{\partial\theta^{s'}}{\partial u^{s'}}| = 0$, a Gronwall argument implies $\mathbb{E}\Big|\frac{\partial\theta^{s+\tau}}{\partial u^s} - \frac{\partial\theta^{s'+\tau}}{\partial u^{s'}}\Big| = o_{|s-s'|}(1)$, so also $s \mapsto R_\theta(s+\tau,s)$ is continuous uniformly over $s \in [0,T]$ and $\tau \in [0,T-s]$. Thus $(s,t) \mapsto R_\theta(t,s)$ is uniformly continuous over all $0 \leq s \leq t \leq T$.

(**Condition for** $\tilde{\alpha}$) By definition, we have

$$\frac{\mathrm{d}}{\mathrm{d}t}\tilde{\alpha}^t = \mathcal{G}(\tilde{\alpha}^t, \mathsf{P}(\theta^t))$$

with $\tilde{\alpha}^0 = \alpha^0$. The condition (9) and boundedness of C_θ shown above imply that $\alpha \mapsto \mathcal{G}(\alpha, \mathsf{P}(\theta^t))$ is Lipschitz uniformly over $t \in [0,T]$, so there exists a unique solution $\{\tilde{\alpha}^t\}_{t\in[0,T]}$ of this equation, which is uniformly continuous on $[0,T]$. Letting $C_0 > 0$ be the constant defining (23–27) which upper bounds (9) of Assumption 2.3, and applying the above bound $\mathbb{E}(\theta^t)^2 = C_\theta(t,t) \leq \Phi_{C_\theta}(t)$, this solution satisfies

$$\frac{\mathrm{d}}{\mathrm{d}t}\|\tilde{\alpha}^t\|^2 \leq 2\|\tilde{\alpha}^t\| \cdot \|\mathcal{G}(\tilde{\alpha}^t, \mathsf{P}(\theta^t))\| \leq 2C_0(1 + \sqrt{\Phi_{C_\theta}(t)} + \|\tilde{\alpha}^t\|)\|\tilde{\alpha}^t\| \leq 4.1C_0(1+\Phi_{C_\theta}(t)) + 3C_0\|\tilde{\alpha}^t\|^2$$

(where we again relax a constant 4 to 4.1). Recalling the equation for Φ_α in (23),

$$\frac{\mathrm{d}}{\mathrm{d}t}\Phi_\alpha(t) = 4.1C_0(1+\Phi_{C_\theta}(t)) + 3C_0\Phi_\alpha(t),$$

since $C_0 > 0$ and $\|\tilde{\alpha}^0\|^2 = \Phi_\alpha(0)$, this shows $\|\tilde{\alpha}^t\|^2 \leq \Phi_\alpha(t)$. □

Next we equip the spaces $\mathcal{S}_\eta$ and $\mathcal{S}_\theta$ with metrics. Fixing a large constant $\lambda > 0$, define

$$\begin{aligned}
d(\alpha_1,\alpha_2) &= \sup_{t\in[0,T]} e^{-\lambda t}\|\alpha_1^t - \alpha_2^t\| \\
d(C_\theta^1, C_\theta^2) &= \inf_{(w_1^*,\{w_1^t\})\sim C_\theta^1, (w_2^*,\{w_2^t\})\sim C_\theta^2} \left[\sqrt{\mathbb{E}(w_1^* - w_2^*)^2} + \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}(w_1^t - w_2^t)^2}\right] \\
d(C_\eta^1, C_\eta^2) &= \inf_{\{u_1^t\}\sim C_\eta^1, \{u_2^t\}\sim C_\eta^2} \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}(u_1^t - u_2^t)^2} \\
d(R_\theta^1, R_\theta^2) &= \sup_{0\leq s\leq t\leq T} e^{-\lambda t}\big|R_\theta^1(t,s) - R_\theta^2(t,s)\big| \\
d(R_\eta^1, R_\eta^2) &= \sup_{0\leq s\leq t\leq T} e^{-\lambda t}\big|R_\eta^1(t,s) - R_\eta^2(t,s)\big|.
\end{aligned} \tag{47}$$

In the definitions of $d(C_\theta^1, C_\theta^2)$ and $d(C_\eta^1, C_\eta^2)$ above, the infima are taken over all couplings of mean-zero Gaussian processes with covariances (C_θ^1, C_θ^2) and (C_η^1, C_η^2). Writing $X^i = (R_\eta^i, C_\eta^i, \alpha_i) \in \mathcal{S}_\eta$ and $Y^i = (R_\theta^i, C_\theta^i, \tilde{\alpha}_i) \in \mathcal{S}_\theta$ for $i = 1,2$, let

$$d(X^1, X^2) = d(R_\eta^1, R_\eta^2) + d(C_\eta^1, C_\eta^2) + d(\alpha_1, \alpha_2), \tag{48}$$

$$d(Y^1, Y^2) = d(R_\theta^1, R_\theta^2) + d(C_\theta^1, C_\theta^2) + d(\tilde{\alpha}_1, \tilde{\alpha}_2). \tag{49}$$

Lemma 3.7 (Modulus of $\mathcal{T}_{\eta\to\theta}$)**.** *Let $X^i = (R^i_\eta, C^i_\eta, \alpha_i) \in \mathcal{S}_\eta$ and $Y^i = \mathcal{T}_{\eta\to\theta}(X^i) = (R^i_\theta, C^i_\theta, \tilde{\alpha}_i) \in \mathcal{S}_\theta$ for $i = 1, 2$. Then for any $\varepsilon > 0$, there exists a constant $\lambda = \lambda(\varepsilon) > 0$ sufficiently large defining the metrics (47) such that*

$$d(Y^1, Y^2) \le \varepsilon \cdot d(X^1, X^2).$$

Proof. We write $C, C' > 0$ for constants that may depend on T, but not on λ, and changing from instance to instance.

Bound of $d(C^1_\theta, C^2_\theta)$. Let $\{u^t_1\}_{t\in[0,T]}$ and $\{u^t_2\}_{t\in[0,T]}$ be an optimal coupling in the definition of $d(C^1_\eta, C^2_\eta)$, i.e.,

$$\sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}[(u^t_1 - u^t_2)^2]} = d(C^1_\eta, C^2_\eta). \tag{50}$$

Let $\{\theta^t_i\}$ be the solution to (12) driven by $\{u^t_i, \alpha^t_i, R^i_\eta\}$ for $i = 1, 2$, with a common Brownian motion $\{b^t\}$ and initialization θ^0, i.e.

$$\theta^t_i = \theta^0 + \int_0^t \Big(-\delta\beta(\theta^s_i - \theta^*) + s(\theta^s_i, \alpha^s_i) + \int_0^s R^i_\eta(s, s')(\theta^{s'}_i - \theta^*)\mathrm{d}s' + u^s_i\Big)\mathrm{d}s + \sqrt{2}b^t. \tag{51}$$

By definition, we have $\mathbb{E}[\theta^t_1\theta^s_1] = \mathbb{E}[\theta^t_2\theta^s_2] = C_\theta(t, s)$. Moreover,

$$\mathbb{E}(\theta^t_1 - \theta^t_2)^2 \le 5[(I) + (II) + (III) + (IV) + (V)]$$

where we set

$$\begin{aligned}
(I) &= \mathbb{E}\Big(\int_0^t \delta\beta|\theta^s_1 - \theta^s_2|\mathrm{d}s\Big)^2 \\
(II) &= \mathbb{E}\Big(\int_0^t |s(\theta^s_1, \alpha^s_1) - s(\theta^s_2, \alpha^s_2)|\mathrm{d}s\Big)^2 \\
(III) &= \mathbb{E}\Big(\int_0^t |\theta^{s'}_1 - \theta^{s'}_2|\Big(\int_{s'}^t |R^1_\eta(s, s')|\mathrm{d}s\Big)\mathrm{d}s'\Big)^2 \\
(IV) &= \mathbb{E}\Big(\int_0^t |\theta^{s'}_2 - \theta^*|\Big(\int_{s'}^t |R^1_\eta(s, s') - R^2_\eta(s, s')|\mathrm{d}s\Big)\mathrm{d}s'\Big)^2 \\
(V) &= \mathbb{E}\Big(\int_0^t |u^s_1 - u^s_2|\mathrm{d}s\Big)^2.
\end{aligned}$$

Term (I) satisfies

$$\begin{aligned}
(I) \le C\int_0^t \mathbb{E}(\theta^s_1 - \theta^s_2)^2\mathrm{d}s &= C\int_0^t e^{2\lambda s}e^{-2\lambda s}\mathbb{E}(\theta^s_1 - \theta^s_2)^2\mathrm{d}s \\
&\le C\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta^t_1 - \theta^t_2)^2\int_0^t e^{2\lambda s}\mathrm{d}s \le \frac{C'}{\lambda}e^{2\lambda t}\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta^t_1 - \theta^t_2)^2.
\end{aligned}$$

To bound (II), applying the Lipschitz properties of $s(\cdot)$ in Assumption 2.2 and a similar argument,

$$\begin{aligned}
(II) \le C\int_0^t \Big(\mathbb{E}(\theta^s_1 - \theta^s_2)^2 + \|\alpha^s_1 - \alpha^s_2\|^2\Big)\mathrm{d}s &\le \frac{C'}{\lambda}e^{2\lambda t}\sup_{t\in[0,T]} e^{-2\lambda t}\Big(\mathbb{E}(\theta^t_1 - \theta^t_2)^2 + \|\alpha^t_1 - \alpha^t_2\|^2\Big) \\
&\le \frac{C'}{\lambda}e^{2\lambda t}\Big(\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta^t_1 - \theta^t_2)^2 + d(\alpha_1, \alpha_2)^2\Big).
\end{aligned}$$

For (III), using the condition $|R^1_\eta(t, s)| \le \Phi_{R_\eta}(t - s) \le C$, we have

$$(III) \le C\int_0^t \mathbb{E}(\theta^s_1 - \theta^s_2)^2\mathrm{d}s \le \frac{C'}{\lambda}e^{2\lambda t}\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta^t_1 - \theta^t_2)^2.$$

For (IV), using $\mathbb{E}(\theta_2^s-\theta^*)^2 \le 2C_\theta(s,s)+2\tau_*^2 \le 2\Phi_{C_\theta}(s)+2\tau_*^2 \le C$, we have

$$(IV) \le C\int_0^t\int_{s'}^t \big(R_\eta^1(s,s')-R_\eta^2(s,s')\big)^2 \mathrm{d}s\,\mathrm{d}s' \le C\int_0^t\int_{s'}^t e^{2\lambda s}\mathrm{d}s\,\mathrm{d}s' \cdot \sup_{0\le s\le t\le T} e^{-2\lambda t}\big(R_\eta^1(t,s)-R_\eta^2(t,s)\big)^2$$
$$\le \frac{C'}{\lambda}e^{2\lambda t} d(R_\eta^1,R_\eta^2)^2.$$

Lastly for (V), using (50), we have

$$(V) \le C\int_0^t \mathbb{E}(u_1^s-u_2^s)^2\mathrm{d}s \le \frac{C'}{\lambda}e^{2\lambda t}\cdot d(C_\eta^1,C_\eta^2)^2.$$

Combining these bounds, for a constant $C>0$ independent of λ,

$$\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta_1^t-\theta_2^t)^2 \le \frac{C}{\lambda}\Big(\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta_1^t-\theta_2^t)^2 + d(X^1,X^2)^2\Big).$$

Thus for any $\varepsilon>0$, choosing $\lambda=\lambda(\varepsilon)$ large enough and rearranging gives

$$\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta_1^t-\theta_2^t)^2 \le \varepsilon^2 d(X^1,X^2)^2.$$

Finally, let $(w^*,\{w_1^t\},\{w_2^t\})$ be a centered Gaussian process with second moments matching $(\theta^*,\{\theta_1^t\},\{\theta_2^t\})$. Then $(w^*,\{w_1^t\})$ and $(w^*,\{w_2^t\})$ realizes a coupling defining the metric $d(C_\theta^1,C_\theta^2)$ in (47), so

$$d(C_\theta^1,C_\theta^2) \le \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}(w_1^t-w_2^t)^2} = \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}(\theta_1^t-\theta_2^t)^2} \le \varepsilon\cdot d(X^1,X^2). \tag{52}$$

Bound of $d(R_\theta^1,R_\theta^2)$. Defining the processes $\frac{\partial\theta_i^t}{\partial u^s}$ for $i=1,2$ from the above coupling of $\{\theta_1^t\}$ and $\{\theta_2^t\}$, by definition we have

$$\frac{\partial\theta_i^t}{\partial u^s} = 1-\int_s^t \Big(\delta\beta-\partial_\theta s(\theta_i^{s'},\alpha_i^{s'})\Big)\frac{\partial\theta_i^{s'}}{\partial u^s}\mathrm{d}s' + \int_s^t\Big(\int_s^{s'} R_\eta^i(s',s'')\frac{\partial\theta_i^{s''}}{\partial u^s}\mathrm{d}s''\Big)\mathrm{d}s',$$

Then

$$\mathbb{E}\Big|\frac{\partial\theta_1^t}{\partial u^s}-\frac{\partial\theta_2^t}{\partial u^s}\Big| \le 4[(I)+(II)+(III)+(IV)]$$

where

$$(I) = \int_s^t \mathbb{E}\Big[\Big|\partial_\theta s(\theta_1^{s'},\alpha_1^{s'})-\partial_\theta s(\theta_2^{s'},\alpha_2^{s'})\Big|\Big|\frac{\partial\theta_1^{s'}}{\partial u^s}\Big|\Big]\mathrm{d}s',$$
$$(II) = \int_s^t \mathbb{E}\Big[\Big(\delta|\beta|+|\partial_\theta s(\theta_2^{s'},\alpha_2^{s'})|\Big)\Big|\frac{\partial\theta_1^{s'}}{\partial u^s}-\frac{\partial\theta_2^{s'}}{\partial u^s}\Big|\Big]\mathrm{d}s',$$
$$(III) = \int_s^t\int_s^{s'} \mathbb{E}\Big[\Big|R_\eta^1(s',s'')-R_\eta^2(s',s'')\Big|\Big|\frac{\partial\theta_1^{s''}}{\partial u^s}\Big|\Big]\mathrm{d}s''\mathrm{d}s',$$
$$(IV) = \int_s^t\int_s^{s'} \mathbb{E}\Big[|R_\eta^2(s',s'')|\Big|\frac{\partial\theta_1^{s''}}{\partial u^s}-\frac{\partial\theta_2^{s''}}{\partial u^s}\Big|\Big]\mathrm{d}s''\mathrm{d}s'.$$

For (I), note that (46) implies $|\frac{\partial\theta_1^{s'}}{\partial u^s}|\le C$ for a constant $C>0$ with probability 1. Then, using the Lipschitz continuity of $\partial_\theta s(\cdot)$ in Assumption 2.2, we have

$$(I) \le C\int_s^t \Big(\mathbb{E}|\theta_1^{s'}-\theta_2^{s'}|+\|\alpha_1^{s'}-\alpha_2^{s'}\|\Big)\mathrm{d}s' \le C\int_s^t e^{\lambda s'}\mathrm{d}s' \sup_{s'\in[0,T]} e^{-\lambda s'}\Big(\mathbb{E}|\theta_1^{s'}-\theta_2^{s'}|+\|\alpha_1^{s'}-\alpha_2^{s'}\|\Big)$$
$$\le \frac{C'}{\lambda}e^{\lambda t}\sup_{s'\in[0,T]} e^{-\lambda s'}\Big(\sqrt{\mathbb{E}\big(\theta_1^{s'}-\theta_2^{s'}\big)^2}+\|\alpha_1^{s'}-\alpha_2^{s'}\|\Big)$$
$$\le \frac{C'}{\lambda}e^{\lambda t}\Big(\varepsilon\cdot d(X^1,X^2)+d(\alpha_1,\alpha_2)\Big),$$

the last step using (52) already shown. For (II), applying the boundedness of $\partial_\theta s(\cdot)$ in Assumption 2.2, we have

$$(II) \leq C \int_s^t e^{\lambda s'} e^{-\lambda s'} \mathbb{E}\Big[\Big|\frac{\partial \theta_1^{s'}}{\partial u^s} - \frac{\partial \theta_2^{s'}}{\partial u^s}\Big|\Big] \mathrm{d}s' \leq \frac{C'}{\lambda} e^{\lambda t} \cdot \sup_{0 \leq s \leq t \leq T} e^{-\lambda t} \mathbb{E}\Big|\frac{\partial \theta_1^t}{\partial u^s} - \frac{\partial \theta_2^t}{\partial u^s}\Big|.$$

For (III), applying again (46) to bound $|\frac{\partial \theta_1^{s'}}{\partial u^s}| \leq C$,

$$(III) \leq \frac{C}{\lambda} e^{\lambda t} \cdot d(R_\eta^1, R_\eta^2).$$

For (IV), applying $|R_\eta^2(t,s)| \leq \Phi_{R_\eta}(t-s) \leq C$,

$$(IV) \leq \frac{C}{\lambda} e^{\lambda t} \sup_{0 \leq s \leq t \leq T} e^{-\lambda t} \mathbb{E}\Big|\frac{\partial \theta_1^t}{\partial u^s} - \frac{\partial \theta_2^t}{\partial u^s}\Big|.$$

Combining these bounds,

$$\sup_{0 \leq s \leq t \leq T} e^{-\lambda t} \mathbb{E}\Big|\frac{\partial \theta_1^t}{\partial u^s} - \frac{\partial \theta_2^t}{\partial u^s}\Big| \leq \frac{C}{\lambda}\Big(\sup_{0 \leq s \leq t \leq T} e^{-\lambda t} \mathbb{E}\Big|\frac{\partial \theta_1^t}{\partial u^s} - \frac{\partial \theta_2^t}{\partial u^s}\Big| + d(X^1, X^2)\Big),$$

so rearranging and choosing $\lambda = \lambda(\varepsilon)$ large enough gives

$$d(R_\theta^1, R_\theta^2) \leq \sup_{0 \leq s \leq t \leq T} e^{-\lambda t} \mathbb{E}\Big|\frac{\partial \theta_1^t}{\partial u^s} - \frac{\partial \theta_2^t}{\partial u^s}\Big| \leq \varepsilon \cdot d(X^1, X^2). \tag{53}$$

Bound of $d(\tilde{\alpha}_1, \tilde{\alpha}_2)$. By definition,

$$\tilde{\alpha}_i^t = \alpha^0 + \int_0^t \mathcal{G}(\tilde{\alpha}_i^s, \mathsf{P}(\theta_i^s)) \mathrm{d}s$$

for $i = 1, 2$. Letting $\{\theta_1^t\}$ and $\{\theta_2^t\}$ be coupled as above and applying Assumption 2.3,

$$\|\tilde{\alpha}_1^t - \tilde{\alpha}_2^t\| \leq C \int_0^t \Big(\|\tilde{\alpha}_1^s - \tilde{\alpha}_2^s\| + W_2(\mathsf{P}(\theta_1^s), \mathsf{P}(\theta_2^s))\Big) \mathrm{d}s \leq C \int_0^t \Big(\|\tilde{\alpha}_1^s - \tilde{\alpha}_2^s\| + \sqrt{\mathbb{E}(\theta_1^s - \theta_2^s)^2}\Big) \mathrm{d}s.$$

Then

$$\|\tilde{\alpha}_1^t - \tilde{\alpha}_2^t\| \leq C \int_0^t e^{\lambda s} \mathrm{d}s \sup_{s \in [0,T]} e^{-\lambda s} \Big(\|\tilde{\alpha}_1^s - \tilde{\alpha}_2^s\| + \sqrt{\mathbb{E}(\theta_1^s - \theta_2^s)^2}\Big) \leq \frac{C'}{\lambda} e^{\lambda t} \Big(d(\tilde{\alpha}_1, \tilde{\alpha}_2) + \varepsilon \cdot d(X^1, X^2)\Big).$$

Choosing $\lambda = \lambda(\varepsilon)$ large enough and rearranging shows

$$d(\tilde{\alpha}_1, \tilde{\alpha}_2) = \sup_{t \in [0,T]} e^{-\lambda t} \|\tilde{\alpha}_1^t - \tilde{\alpha}_2^t\| \leq \varepsilon \cdot d(X^1, X^2). \tag{54}$$

The lemma follows from (52), (53), and (54). □

Lemma 3.8 (Modulus of $\mathcal{T}_{\theta \to \eta}$)**.** *Let $Y^i = (R_\theta^i, C_\theta^i, \tilde{\alpha}_i) \in \mathcal{S}_\theta$ and $X^i = \mathcal{T}_{\theta \to \eta}(Y^i) = (C_\eta^i, R_\eta^i, \alpha_i) \in \mathcal{S}_\eta$ for $i = 1, 2$. Then there exists a constant $C > 0$ such that for any sufficiently large $\lambda > 0$ defining the metrics (47),*

$$d(X^1, X^2) \leq C \cdot d(Y^1, Y^2).$$

Proof. The proof is similar to that of Lemma 3.7 so we will omit some details. Again let $C, C', C'' > 0$ denote constants depending on T but not on λ.

Bound of $d(C_\eta^1, C_\eta^2)$. Let $(w_1^*, \{w_1^t\})$ and $(w_2^*, \{w_2^t\})$ be an optimal coupling for which

$$\sqrt{\mathbb{E}[(w_1^* - w_2^*)^2]} + \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}[(w_1^t - w_2^t)^2]} = d(C_\theta^1, C_\theta^2).$$

For $i = 1, 2$, let

$$\eta_i^t = -\beta \int_0^t R_\theta^i(t,s)\big(\eta_i^s + w_i^* - \varepsilon\big)\mathrm{d}s - w_i^t$$

be the corresponding coupled solutions to (14). We write $\xi_i^t = \eta_i^t + w_i^* - \varepsilon$, so that

$$\xi_i^t = -\beta \int_0^t R_\theta^i(t,s)\xi_i^s \mathrm{d}s - w_i^t + w_i^* - \varepsilon$$

and $C_\eta^i(t,s) = \delta\beta^2 \mathbb{E}[\xi_i^t \xi_i^s]$. Then

$$\begin{aligned}
\mathbb{E}(\xi_1^t - \xi_2^t)^2 &\leq C\Big[\int_0^t (R_\theta^1(t,s) - R_\theta^2(t,s))^2 \mathbb{E}(\xi_1^s)^2 \mathrm{d}s + \int_0^t R_\theta^2(t,s)^2 \mathbb{E}(\xi_1^s - \xi_2^s)^2 \mathrm{d}s + \mathbb{E}(w_1^t - w_1^* - w_2^t + w_2^*)^2\Big] \\
&\leq C'\Big[\int_0^t e^{2\lambda s}\cdot e^{-2\lambda s}\Big((R_\theta^1(t,s) - R_\theta^2(t,s))^2 + \mathbb{E}(\xi_1^s - \xi_2^s)^2\Big)\mathrm{d}s + \mathbb{E}(w_1^t - w_1^* - w_2^t + w_2^*)^2\Big] \\
&\leq \frac{C''}{\lambda} e^{2\lambda t}\Big(\sup_{s\in[0,T]} e^{-2\lambda s}\mathbb{E}(\xi_1^s - \xi_2^s)^2 + d(R_\theta^1, R_\theta^2)^2\Big) + C'' e^{2\lambda t} d(C_\theta^1, C_\theta^2)^2.
\end{aligned}$$

Choosing $\lambda > 2C''$ and rearranging yields, for a constant $C > 0$,

$$\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\xi_1^t - \xi_2^t)^2 \leq C \cdot d(Y^1, Y^2)^2.$$

Then letting $(\{u_1^t\}, \{u_2^t\})$ be a centered Gaussian process with second moments $\mathbb{E}[u_1^t u_2^s] = \delta\beta^2\mathbb{E}[\xi_1^t \xi_2^s]$, this realizes a coupling defining $d(C_\eta^1, C_\eta^2)$, so

$$d(C_\eta^1, C_\eta^2) \leq \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}[(u_1^t - u_2^t)^2]} = \sqrt{\delta\beta^2}\cdot \sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}(\xi_1^t - \xi_2^t)^2} \leq C' \cdot d(Y^1, Y^2).$$

Bound of $d(R_\eta^1, R_\eta^2)$. Defining the (deterministic) process $\frac{\partial \eta_i^t}{\partial w^s}$ driven by R_θ^i for $i = 1, 2$, we have

$$R_\eta^i(t,s) = -\beta \int_s^t R_\theta^i(t,s') R_\eta^i(s',s)\mathrm{d}s' + \delta\beta^2 R_\theta^i(t,s),$$

hence

$$\begin{aligned}
|R_\eta^1(t,s) - R_\eta^2(t,s)| \leq |\beta| &\int_s^t |R_\theta^1(t,s') - R_\theta^2(t,s')||R_\eta^1(s',s)|\mathrm{d}s' + |\beta|\int_s^t |R_\theta^2(t,s')||R_\eta^1(s',s) - R_\eta^2(s',s)|\mathrm{d}s' \\
&+ \delta\beta^2 |R_\theta^1(t,s) - R_\theta^2(t,s)| \\
&\leq C\int_s^t e^{\lambda s'} e^{-\lambda s'}|R_\eta^1(s',s) - R_\eta^2(s',s)|\mathrm{d}s' + Ce^{\lambda t} d(R_\theta^1, R_\theta^2) \\
&\leq \frac{C'}{\lambda} e^{\lambda t}\Big(\sup_{0\leq s\leq t\leq T} e^{-\lambda t}|R_\eta^1(t,s) - R_\eta^2(t,s)|\Big) + Ce^{\lambda t} d(R_\theta^1, R_\theta^2).
\end{aligned}$$

Choosing $\lambda > 2C'$ and rearranging yields

$$d(R_\eta^1, R_\eta^2) = \sup_{0\leq s\leq t\leq T} e^{-\lambda t}|R_\eta^1(t,s) - R_\eta^2(t,s)| \leq C \cdot d(R_\theta^1, R_\theta^2) \leq C \cdot d(Y^1, Y^2).$$

We note that $\alpha_i = \tilde{\alpha}_i$ for $i = 1, 2$ by definition, so also $d(\alpha_1, \alpha_2) = d(\tilde{\alpha}_1, \tilde{\alpha}_2) \leq d(Y^1, Y^2)$. Combining these bounds shows the lemma. □

Proof of Theorem 2.4(b). Combining Lemmas 3.7 and 3.8, for sufficiently large $\lambda > 0$, the composition map $\mathcal{T}_{\eta\to\eta}$ is a contraction on $\mathcal{S}_\eta^{\mathrm{cont}}$ with respect to the metric $d(X^1, X^2)$ and similarly $\mathcal{T}_{\theta\to\theta}$ is a contraction on $\mathcal{S}_\theta^{\mathrm{cont}}$. We note that for any sequence $\{C_\theta^k\}$ of correlation functions in $\mathcal{S}^{\mathrm{cont}}$, as $k \to \infty$,

$$d(C_\theta^k, C_\theta) \to 0 \text{ implies } \sup_{s,t\in[0,T]} |C_\theta^k(s,t) - C_\theta(s,t)| \to 0$$

by definition of the metric and Cauchy-Schwarz, while

$$\sup_{s,t\in[0,T]} |C_\theta^k(s,t) - C_\theta(s,t)| \to 0 \text{ implies } d(C_\theta^k, C_\theta) \to 0$$

by e.g. the construction of a coupling in [52, Lemma D.1]. The same holds for C_η, so each metric in (47) induces a topology equivalent to that of uniform convergence over continuous functions on the appropriate space $[0,T]$, $\{s,t : 0 \le s \le t \le T\}$, or $\{*\} \cup \{s,t : 0 \le s \le t \le T\}$. Furthermore, each condition defining $\mathcal{S}_\eta^{\mathrm{cont}}, \mathcal{S}_\theta^{\mathrm{cont}}$ is closed with respect to this topology. Thus $d(X_\eta^1, X_\eta^2)$ and $d(Y_\theta^1, Y_\theta^2)$ are complete metrics on $\mathcal{S}_\eta^{\mathrm{cont}}, \mathcal{S}_\theta^{\mathrm{cont}}$, so the Banach fixed-point theorem guarantees $\mathcal{T}_{\eta\to\eta}$ and $\mathcal{T}_{\theta\to\theta}$ have unique fixed points $X = (R_\eta, C_\eta, \alpha) \in \mathcal{S}_\eta^{\mathrm{cont}}$ and $Y = (R_\theta, C_\theta, \alpha) \in \mathcal{S}_\theta^{\mathrm{cont}}$, for which also $\mathcal{T}_{\eta\to\theta}(X) = Y$. These fixed points remain unique in $\mathcal{S}_\eta$ and $\mathcal{S}_\theta$, because Lemmas 3.5 and 3.6 imply that the images of $\mathcal{T}_{\eta\to\eta}, \mathcal{T}_{\theta\to\theta}$ on $\mathcal{S}_\eta, \mathcal{S}_\theta$ are contained in $\mathcal{S}_\eta^{\mathrm{cont}}, \mathcal{S}_\theta^{\mathrm{cont}}$. Then the tuple $(\alpha, C_\theta, C_\eta, R_\theta, R_\eta) \in \mathcal{S}^{\mathrm{cont}}$ is the unique fixed point in $\mathcal{S}$ solving the dynamical fixed point equations (17–18). □

4 The dynamical mean-field approximation

In this section we prove Theorem 2.5. We assume throughout Assumptions 2.1, 2.2, and 2.3. The proof consists of three steps:

- (**Step 1**) We prove in Section 4.1 a discrete DMFT limit for a discretized version of the dynamics.
- (**Step 2**) We show in Section 4.2 that, as the discretization step size goes to zero, the discrete DMFT equations converge in an appropriate sense to (12–18).
- (**Step 3**) We show in Section 4.3 that, as the discretization step size goes to zero, the discretized dynamics converges in an appropriate sense to (4–5).

This argument follows closely the approach of [42], although we will use in Steps 2 and 3 a different and somewhat simpler piecewise-constant embedding of the discretized DMFT process and discretized Langevin dynamics into continuous time.

4.1 Step 1: DMFT approximation of discrete dynamics

Fix a step size $\gamma > 0$. We first define a discretized version of the process (4–5), which we denote by $\{\boldsymbol{\theta}_\gamma^t\}$ and $\{\widehat{\alpha}_\gamma^t\}$ for $t \in \mathbb{Z}_+ = \{0, 1, 2, \ldots\}$:

$$\boldsymbol{\theta}_\gamma^{t+1} = \boldsymbol{\theta}_\gamma^t + \gamma\Big(-\beta\mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}_\gamma^t - \mathbf{y}) + s(\boldsymbol{\theta}_\gamma^t, \widehat{\alpha}_\gamma^t)\Big) + \sqrt{2}(\mathbf{b}_\gamma^{t+1} - \mathbf{b}_\gamma^t) \tag{55}$$

$$\widehat{\alpha}_\gamma^{t+1} = \widehat{\alpha}_\gamma^t + \gamma \cdot \mathcal{G}\Big(\widehat{\alpha}_\gamma^t, \frac{1}{d}\sum_{j=1}^d \delta_{\theta_{\gamma,j}^t}\Big) \tag{56}$$

with initialization $(\boldsymbol{\theta}_\gamma^0, \widehat{\alpha}_\gamma^0) = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$, where $\{\mathbf{b}_\gamma^t\}$ is a discrete Gaussian process with $\mathbf{b}_\gamma^0 = 0$ and independent increments $\mathbf{b}_\gamma^{t+1} - \mathbf{b}_\gamma^t \sim \mathcal{N}(0, \gamma\mathbf{I})$. Here and throughout the sequel, we write as shorthand $s(\boldsymbol{\theta}, \widehat{\alpha}) = (s(\theta_j, \widehat{\alpha}))_{j=1}^d$. We set

$$\boldsymbol{\eta}_\gamma^t = \mathbf{X}\boldsymbol{\theta}_\gamma^t, \qquad \boldsymbol{\eta}^* = \mathbf{X}\boldsymbol{\theta}^*.$$

We correspondingly define a discretized version of the DMFT system (12–18): Given discrete-time correlation and response matrices $\{C_\eta^\gamma(s,r)\}_{r\le s\le t}$, $\{R_\eta^\gamma(s,r)\}_{r<s\le t}$ and a deterministic process $\{\alpha_\gamma^s\}_{s\le t}$ up to

time t, define (in the probability space of $(\theta^*, \theta^0) \sim \mathsf{P}(\theta^*, \theta^0)$)

$$\theta_\gamma^{t+1} = \theta_\gamma^t + \gamma\Big(-\delta\beta(\theta_\gamma^t - \theta^*) + s(\theta_\gamma^t, \alpha_\gamma^t) + \sum_{s=0}^{t-1} R_\eta^\gamma(t,s)(\theta_\gamma^s - \theta^*) + u_\gamma^t\Big) + \sqrt{2}(b_\gamma^{t+1} - b_\gamma^t)$$
$$\text{with } \theta_0^\gamma = \theta^0, \tag{57}$$

$$\frac{\partial \theta_\gamma^{t+1}}{\partial u_\gamma^s} = \begin{cases} \gamma & \text{for } s = t, \\ \frac{\partial \theta_\gamma^t}{\partial u_\gamma^s} + \gamma\Big[\Big(-\delta\beta + \partial_\theta s(\theta_\gamma^t, \alpha_\gamma^t)\Big)\frac{\partial \theta_\gamma^t}{\partial u_\gamma^s} + \sum_{r=s+1}^{t-1} R_\eta^\gamma(t,r)\frac{\partial \theta_\gamma^r}{\partial u_\gamma^s}\Big] & \text{for } s < t. \end{cases} \tag{58}$$

Here, $\{u_\gamma^s\}_{0\le s\le t}$ and $\{b_\gamma^s\}_{0\le s\le t}$ are mean-zero Gaussian vectors independent of each other and of (θ^*, θ^0), where $\{u_\gamma^s\}_{0\le s\le t}$ has covariance

$$\mathbb{E}[u_\gamma^s u_\gamma^r] = C_\eta^\gamma(s,r), \tag{59}$$

and $\{b_\gamma^s\}_{0\le s\le t}$ has independent increments $b_\gamma^{s+1} - b_\gamma^s \sim \mathcal{N}(0,\gamma)$ with $b_\gamma^0 = 0$. We note that $\frac{\partial \theta_\gamma^{t+1}}{\partial u_\gamma^s}$ is the usual partial derivative of θ_γ^{t+1} in u_γ^s, whose form (58) is derived from (57) via the chain rule. These processes then define $\{C_\theta^\gamma(s,r)\}_{r\le s\le t+1}$, $\{C_\theta^\gamma(s,*)\}_{s\le t+1}$, $\{R_\theta^\gamma(s,r)\}_{r<s\le t+1}$, and $\{\alpha_\gamma^s\}_{s\le t+1}$ up to time $t+1$ via

$$\begin{aligned} C_\theta^\gamma(s,r) = \mathbb{E}[\theta_\gamma^s \theta_\gamma^r], \quad C_\theta^\gamma(s,*) = \mathbb{E}[\theta_\gamma^s \theta^*], \quad C_\theta^\gamma(*,*) = \mathbb{E}[(\theta^*)^2], \\ R_\theta^\gamma(s,r) = \mathbb{E}\Big[\frac{\partial \theta_\gamma^s}{\partial u_\gamma^r}\Big], \quad \alpha_\gamma^{t+1} = \alpha_\gamma^t + \gamma \cdot \mathcal{G}(\alpha_\gamma^t, \mathsf{P}(\theta_\gamma^t)) \end{aligned} \tag{60}$$

where $\mathsf{P}(\theta_\gamma^t)$ is the law of θ_γ^t.

Conversely, given $\{C_\theta^\gamma(s,r)\}_{r\le s\le t}$, $\{C_\theta^\gamma(s,*)\}_{s\le t}$, and $\{R_\theta^\gamma(s,r)\}_{r<s\le t}$ up to time t, define (in the probability space of $\varepsilon \sim \mathsf{P}(\varepsilon)$)

$$\eta_\gamma^t = -\beta \sum_{s=0}^{t-1} R_\theta^\gamma(t,s)(\eta_\gamma^s + w_\gamma^* - \varepsilon) - w_\gamma^t, \tag{61}$$

$$\frac{\partial \eta_\gamma^t}{\partial w_\gamma^s} = \beta\Big[-\sum_{r=s+1}^{t-1} R_\theta^\gamma(t,r)\frac{\partial \eta_\gamma^r}{\partial w_\gamma^s} + R_\theta^\gamma(t,s)\Big] \quad \text{for } s < t. \tag{62}$$

Here, $(w_\gamma^*, \{w_\gamma^s\}_{0\le s\le t})$ is a mean-zero Gaussian vector with covariance

$$\mathbb{E}[w_\gamma^s w_\gamma^r] = C_\theta^\gamma(s,r), \qquad \mathbb{E}[w_\gamma^s w_\gamma^*] = C_\theta^\gamma(s,*), \qquad \mathbb{E}[(w_\gamma^*)^2] = C_\theta^\gamma(*,*), \tag{63}$$

and again $\frac{\partial \eta_\gamma^t}{\partial w_\gamma^s}$ is the usual partial derivative computed from the chain rule. These define $\{C_\eta^\gamma(s,r)\}_{r\le s\le t}$, $\{R_\eta^\gamma(s,r)\}_{r<s\le t}$ up to time t via

$$C_\eta^\gamma(s,r) = \delta\beta^2 \mathbb{E}[(\eta_\gamma^s + w_\gamma^* - \varepsilon)(\eta_\gamma^r + w_\gamma^* - \varepsilon)], \quad R_\eta^\gamma(s,r) = \delta\beta\left(\frac{\partial \eta_\gamma^s}{\partial w_\gamma^r}\right), \tag{64}$$

where we note that $\frac{\partial \eta_\gamma^s}{\partial w_\gamma^r}$ is deterministic. These definitions should be understood in the iterative sense

$$\begin{aligned} &\{\theta_\gamma^s\}_{s\le t}, \{u_\gamma^s\}_{s<t}, \{\tfrac{\partial \theta_\gamma^s}{\partial u_\gamma^r}\}_{r<s\le t} \Rightarrow \{C_\theta^\gamma(s,r), C_\theta^\gamma(s,*)\}_{r\le s\le t}, \{R_\theta^\gamma(s,r)\}_{r<s\le t}, \{\alpha^s\}_{s\le t} \Rightarrow \\ &w_\gamma^*, \{\eta_\gamma^s, w_\gamma^s\}_{s\le t}, \{\tfrac{\partial \eta_\gamma^s}{\partial w_\gamma^r}\}_{r<s\le t} \Rightarrow \{C_\eta^\gamma(s,r)\}_{r\le s\le t}, \{R_\eta^\gamma(s,r)\}_{r<s\le t} \Rightarrow \\ &\{\theta_\gamma^s\}_{s\le t+1}, \{u_\gamma^s\}_{s<t+1}, \{\tfrac{\partial \theta_\gamma^s}{\partial \theta_\gamma^r}\}_{r<s\le t+1} \Rightarrow \dots \end{aligned} \tag{65}$$

with initialization $\theta_\gamma^0 = \theta^0$.

The goal of this section is to show the following discrete analogue of Theorem 2.5.

Lemma 4.1. *For any fixed integer $T \geq 0$, almost surely as $n, d \to \infty$,*

$$\frac{1}{d}\sum_{j=1}^{d} \delta_{\left(\theta_j^*, \theta_{\gamma,j}^0, \dots, \theta_{\gamma,j}^T\right)} \overset{W_2}{\to} \mathsf{P}\left(\theta^*, \theta_\gamma^0, \dots, \theta_\gamma^T\right) \tag{66}$$

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\left(\eta_i^*, \varepsilon_i, \eta_{\gamma,i}^0, \dots, \eta_{\gamma,i}^0\right)} \overset{W_2}{\to} \mathsf{P}\left(-w_\gamma^*, \varepsilon, \eta_\gamma^0, \dots, \eta_\gamma^T\right) \tag{67}$$

$$(\widehat{\alpha}_\gamma^0, \dots, \widehat{\alpha}_\gamma^T) \to (\alpha_\gamma^0, \dots, \alpha_\gamma^T). \tag{68}$$

For convenience of the proof, we define also an auxiliary response function

$$R_\eta^\gamma(t, *) = \delta\beta\left(\frac{\partial \eta_\gamma^t}{\partial w_\gamma^*}\right) \quad \text{where} \quad \frac{\partial \eta_\gamma^t}{\partial w_\gamma^*} = -\beta\sum_{s=0}^{t-1} R_\theta^\gamma(t, s)\Big(\frac{\partial \eta_\gamma^s}{\partial w_\gamma^*} + 1\Big), \tag{69}$$

initialized from $\frac{\partial \eta_\gamma^0}{\partial w_\gamma^*} = 0$. Here $\frac{\partial \eta_\gamma^t}{\partial w_\gamma^*}$ is the usual partial derivative of η_γ^t with respect to w_γ^*, which is also deterministic. We have the following basic fact relating the response functions (62) and (69).

Lemma 4.2. *For any $t \geq 1$, we have $\frac{\partial \eta_\gamma^t}{\partial w_\gamma^*} = -\sum_{s=0}^{t-1} \frac{\partial \eta_\gamma^t}{\partial w_\gamma^s}$, and consequently $R_\eta^\gamma(t, *) = -\sum_{s=0}^{t-1} R_\eta^\gamma(t, s)$.*

Proof. Let us shorthand $r_\eta(t, s) = \frac{\partial \eta_\gamma^t}{\partial w_\gamma^s}$ for $s < t$ and $r_\eta(t, *) = \frac{\partial \eta_\gamma^t}{\partial w_\gamma^*}$. We prove $r_\eta(t, *) = -\sum_{s=0}^{t-1} r_\eta(t, s)$ by induction, with the base case $t = 1$ verified by the initial conditions $r_\eta(1, *) = -\gamma\beta$ and $r_\eta(1, 0) = \gamma\beta$. Suppose the claim holds for some t, then

$$\begin{aligned}
r_\eta(t+1, *) &= -\beta\sum_{s=0}^{t} R_\theta^\gamma(t+1, s)(r_\eta(s, *) + 1) \\
&= -\beta\sum_{s=0}^{t} R_\theta^\gamma(t+1, s)\Big(-\sum_{r=0}^{s-1} r_\eta(s, r) + 1\Big) \\
&= \beta\Big[\sum_{r=0}^{t-1}\sum_{s=r+1}^{t} R_\theta^\gamma(t+1, s) r_\eta(s, r) - \sum_{r=0}^{t} R_\theta^\gamma(t+1, r)\Big] \\
&= \beta\sum_{r=0}^{t-1}\Big(\sum_{s=r+1}^{t} R_\theta^\gamma(t+1, s) r_\eta(s, r) - R_\theta^\gamma(t+1, r)\Big) - \beta R_\theta^\gamma(t+1, t) \\
&= -\sum_{r=0}^{t-1} r_\eta(t+1, r) - r_\eta(t+1, t) = -\sum_{r=0}^{t} r_\eta(t+1, r),
\end{aligned}$$

as desired. ◻

Proof of Lemma 4.1. **Step 1: Convergence of auxiliary dynamics.** Consider the following non-adaptive auxiliary dynamics

$$\tilde{\boldsymbol{\theta}}_\gamma^{t+1} = \tilde{\boldsymbol{\theta}}_\gamma^t - \gamma\bigg(\beta\mathbf{X}^\top(\mathbf{X}\tilde{\boldsymbol{\theta}}_\gamma^t - \mathbf{X}\boldsymbol{\theta}^* - \boldsymbol{\varepsilon}) - s(\tilde{\boldsymbol{\theta}}_\gamma^t, \alpha_\gamma^t)\bigg) + \sqrt{2}(\mathbf{b}_\gamma^{t+1} - \mathbf{b}_\gamma^t). \tag{70}$$

This differs from $\{\boldsymbol{\theta}_\gamma^t\}$ in that we replace $\{\widehat{\alpha}_\gamma^t\}$ by the deterministic process $\{\alpha_\gamma^t\}$ of the discrete DMFT system. Let $\tilde{\boldsymbol{\eta}}_\gamma^t = \mathbf{X}\tilde{\boldsymbol{\theta}}_\gamma^t$. We will first show

$$\frac{1}{d}\sum_{j=1}^{d} \delta_{\left(\theta_j^*, \tilde{\theta}_{\gamma,j}^0, \dots, \tilde{\theta}_{\gamma,j}^T\right)} \overset{W_2}{\to} \mathsf{P}\left(\theta^*, \theta_\gamma^0, \dots, \theta_\gamma^T\right), \quad \frac{1}{n}\sum_{i=1}^{n} \delta_{\left(\eta_i^*, \varepsilon_i, \tilde{\eta}_{\gamma,i}^0, \dots, \tilde{\eta}_{\gamma,i}^T\right)} \overset{W_2}{\to} \mathsf{P}\left(-w_\gamma^*, \varepsilon, \eta_\gamma^0, \dots, \eta_\gamma^T\right). \tag{71}$$

The proof is based on a reduction to an AMP algorithm: Let $\boldsymbol{\varepsilon} \in \mathbb{R}^n$ be as in the above dynamics, define

$$\mathbf{V} = (\boldsymbol{\theta}^*, \boldsymbol{\theta}^0, \mathbf{b}^1 - \mathbf{b}^0, \dots, \mathbf{b}^T - \mathbf{b}^{T-1}) \in \mathbb{R}^{d \times (T+2)}$$

and let

$$\varepsilon \sim \mathsf{P}(\varepsilon), \qquad V = (\theta^*, \theta^0, \rho^1, \dots, \rho^T) \sim \mathsf{P}(\theta^*, \theta^0) \otimes \mathcal{N}(0, \gamma I_T).$$

Assumption 2.1 ensures that $|\varepsilon|$ and $\|v\|_2$ have finite moment generating functions in a neighborhood of 0, and for each fixed $p \geq 1$, almost surely as $n, d \to \infty$,

$$\frac{1}{n}\sum_{i=1}^n \delta_{\varepsilon_i} \overset{W_p}{\to} \varepsilon, \quad \frac{1}{d}\sum_{j=1}^d \delta_{V_j} \overset{W_p}{\to} V \tag{72}$$

where V_j is the j^{th} row of $\mathbf{V}$. Fixing some $k \geq 1$, consider the AMP iterations

$$\begin{aligned} \mathbf{W}^i &= \mathbf{X} g_i(\mathbf{U}^1, \dots, \mathbf{U}^i; \mathbf{V}) - \sum_{j=0}^{i-1} f_j(\mathbf{W}^0, \dots, \mathbf{W}^j; \boldsymbol{\varepsilon}) \zeta_{ij} \in \mathbb{R}^{n \times k}, \\ \mathbf{U}^{i+1} &= \mathbf{X}^\top f_i(\mathbf{W}^0, \dots, \mathbf{W}^i; \boldsymbol{\varepsilon}) - \sum_{j=0}^{i} g_j(\mathbf{U}^1, \dots, \mathbf{U}^j; \mathbf{V}) \xi_{ij} \in \mathbb{R}^{d \times k}, \end{aligned} \tag{73}$$

initialized at $\mathbf{W}^0 = \mathbf{X} g_0(\mathbf{V})$, where the nonlinearities

$$f_i = \big(f_{i,1}, \dots, f_{i,k}\big) : \mathbb{R}^{k(i+1)} \times \mathbb{R} \to \mathbb{R}^k, \qquad g_i = \big(g_{i,1}, \dots, g_{i,k}\big) : \mathbb{R}^{ki} \times \mathbb{R}^{T+2} \to \mathbb{R}^k$$

are Lipschitz-continuous and applied row-wise, the Onsager coefficients are recursively defined as

$$\begin{aligned} \xi_{ij} &= \Big(\delta\, \mathbb{E}\Big[\mathrm{d}_{W^j} f_i(W^0, \dots, W^i; \varepsilon)\Big]\Big)^\top \in \mathbb{R}^{k \times k}, \quad 0 \leq j \leq i, \\ \zeta_{ij} &= \Big(\mathbb{E}\Big[\mathrm{d}_{U^{j+1}} g_i(U^1, \dots, U^i; V)\Big]\Big)^\top \in \mathbb{R}^{k \times k}, \quad 0 \leq j \leq i-1, \end{aligned}$$

and $\{W^j\}_{j \geq 0}$ and $\{U^j\}_{j \geq 1}$ are mean-zero Gaussian processes in $\mathbb{R}^k$ independent of ε, V with covariance structure

$$\begin{aligned} \mathbb{E}[W^i W^{j\top}] &= \mathbb{E}\Big[g_i(U^1, \dots, U^i; V) g_j(U^1, \dots, U^j; V)^\top\Big] \in \mathbb{R}^{k \times k}, \quad i, j \geq 0, \\ \mathbb{E}[U^{i+1} U^{j+1\top}] &= \mathbb{E}\Big[\delta f_i(W^0, \dots, W^i; \varepsilon) f_j(W^0, \dots, W^j; \varepsilon)^\top\Big] \in \mathbb{R}^{k \times k}, \quad i, j \geq 0. \end{aligned} \tag{74}$$

This is a standard form of an AMP algorithm, see e.g. [45,60]. The iterations for $(\mathbf{W}^0, \dots, \mathbf{W}^{T-1}) \in \mathbb{R}^{n \times kT}$ and $(\mathbf{U}^1, \dots, \mathbf{U}^T) \in \mathbb{R}^{d \times kT}$ admit a mapping to the form of [60, Eqs. (2.14) and (D.1–D.2)] with kT vector iterates. Then by the AMP state evolution (c.f. [60, Theorem 2.21 and Remark 2.2]), under the conditions of Assumption 2.1, almost surely as $n, d \to \infty$,

$$\begin{aligned} \frac{1}{d}\sum_{j=1}^d \delta_{U_j^1, \dots, U_j^m, V_j} &\overset{W_2}{\to} \mathsf{P}(U^1, \dots, U^m, V), \\ \frac{1}{n}\sum_{i=1}^n \delta_{W_i^0, \dots, W_i^m, \varepsilon_i} &\overset{W_2}{\to} \mathsf{P}(W^0, \dots, W^m, \varepsilon). \end{aligned} \tag{75}$$

We will now use the above state evolution to prove the desired conclusion (71). In the AMP algorithm (73), let $k = 2$. We show the existence of Lipschitz nonlinearities $g_i = (g_{i,1}, g_{i,2}) : \mathbb{R}^{2i} \times \mathbb{R}^{T+2} \to \mathbb{R}^2$ and $f_i = (f_{i,1}, f_{i,2}) : \mathbb{R}^{2(i+1)} \times \mathbb{R} \to \mathbb{R}^2$ such that

$$(\tilde{\boldsymbol{\theta}}_\gamma^j, \boldsymbol{\theta}^*) = g_j(\mathbf{U}^1, \dots, \mathbf{U}^j; \mathbf{V}), \tag{76}$$

$$\Big(-(\beta/\delta)(\mathbf{X}\tilde{\boldsymbol{\theta}}_\gamma^j - \mathbf{X}\boldsymbol{\theta}^* - \boldsymbol{\varepsilon}), 0\Big) = f_j(\mathbf{W}^0, \dots, \mathbf{W}^j; \boldsymbol{\varepsilon}). \tag{77}$$

The base case is $g_0(\mathbf{V}) = (\boldsymbol{\theta}^0, \boldsymbol{\theta}^*)$ and $f_0(\mathbf{W}^0; \boldsymbol{\varepsilon}) = (-(\beta/\delta)(\mathbf{W}_1^0 - \mathbf{W}_2^0 - \boldsymbol{\varepsilon}), 0)$ where $\mathbf{W}^0 = (\mathbf{W}_1^0, \mathbf{W}_2^0) = (\mathbf{X}\boldsymbol{\theta}^0, \mathbf{X}\boldsymbol{\theta}^*)$. Supposing inductively that (76–77) hold for some Lipschitz functions $g_0, f_0, \ldots, g_j, f_j$, we note that this implies $(\xi_{j\ell})_{12} = (\xi_{j\ell})_{22} = 0$ for all $\ell \leq j$. Then writing $\mathbf{U}^j = (\mathbf{U}_1^j, \mathbf{U}_2^j) \in \mathbb{R}^{d\times 2}$, we have

$$\begin{aligned}
\mathbf{U}_1^{j+1} &= \mathbf{X}^\top f_{j,1}(\mathbf{W}^0, \ldots, \mathbf{W}^j; \boldsymbol{\varepsilon}) - \sum_{\ell=0}^{j} \Big(g_{\ell,1}(\mathbf{U}^1, \ldots, \mathbf{U}^\ell; \mathbf{V})(\xi_{j\ell})_{11} + g_{\ell,2}(\mathbf{U}^1, \ldots, \mathbf{U}^\ell; \mathbf{V})(\xi_{j\ell})_{21} \Big) \\
&= -\frac{\beta}{\delta} \mathbf{X}^\top (\mathbf{X}\tilde{\boldsymbol{\theta}}_\gamma^j - \mathbf{X}\boldsymbol{\theta}^* - \boldsymbol{\varepsilon}) - \sum_{\ell=0}^{j} g_{\ell,1}(\mathbf{U}^1, \ldots, \mathbf{U}^\ell; \mathbf{V})(\xi_{j\ell})_{11} - \sum_{\ell=0}^{j} (\xi_{j\ell})_{21} \cdot \boldsymbol{\theta}^*.
\end{aligned}$$

So

$$\begin{aligned}
\tilde{\boldsymbol{\theta}}_\gamma^{j+1} &= \tilde{\boldsymbol{\theta}}_\gamma^j - \gamma \Big(\beta \mathbf{X}^\top (\mathbf{X}\tilde{\boldsymbol{\theta}}_\gamma^j - \mathbf{X}\boldsymbol{\theta}^* - \boldsymbol{\varepsilon}) - s(\tilde{\boldsymbol{\theta}}_\gamma^j, \alpha_\gamma^j) \Big) + \sqrt{2}(\mathbf{b}^{j+1} - \mathbf{b}^j) \\
&= \tilde{\boldsymbol{\theta}}_\gamma^j + \gamma\delta \Big(\mathbf{U}_1^{j+1} + \sum_{\ell=0}^{j} g_{\ell,1}(\mathbf{U}^1, \ldots, \mathbf{U}^\ell; \mathbf{V})(\xi_{j\ell})_{11} + \sum_{\ell=0}^{j} (\xi_{j\ell})_{21} \cdot \boldsymbol{\theta}^* \Big) + \gamma s(\tilde{\boldsymbol{\theta}}_\gamma^t, \alpha_\gamma^j) + \sqrt{2}(\mathbf{b}^{j+1} - \mathbf{b}^j),
\end{aligned}$$

and to satisfy (76) we may define $g_{j+1}(\cdot)$ as $g_{j+1,2}(U^1, \ldots, U^{j+1}; V) = \theta^*$ and

$$\begin{aligned}
g_{j+1,1}(U^1, \ldots, U^{j+1}; V) = g_{j,1}(U^1, \ldots, U^j; V) &+ \gamma\delta \Big(U_1^{j+1} + \sum_{\ell=0}^{j} g_{\ell,1}(U^1, \ldots, U^\ell; V)(\xi_{j\ell})_{11} + \sum_{\ell=0}^{j} (\xi_{j\ell})_{21} \cdot \theta^* \Big) \\
&+ \gamma s\big(g_{j,1}(U^1, \ldots, U^j; V), \alpha_\gamma^j\big) + \sqrt{2}\rho_j, \qquad (78)
\end{aligned}$$

where we recall $V = (\theta^*, \theta^0, \rho_1, \ldots, \rho_T)$. We note that $\theta \mapsto s(\theta, \alpha_\gamma^j)$ is Lipschitz by Assumption 2.2, so this function $g_{j+1}(\cdot)$ is also Lipschitz by the induction hypothesis. Next, the condition $g_{j+1,2}(\cdot) = \theta^*$ implies $(\zeta_{j+1,\ell})_{12} = (\zeta_{j+1,\ell})_{22} = 0$ for $\ell \leq j$, and allows us to compute $\mathbf{W}^{j+1} = (\mathbf{W}_1^{j+1}, \mathbf{W}_2^{j+1})$ as $\mathbf{W}_2^{j+1} = \mathbf{X}\boldsymbol{\theta}^*$ and

$$\begin{aligned}
\mathbf{W}_1^{j+1} &= \mathbf{X} g_{j+1,1}(\mathbf{U}^1, \ldots, \mathbf{U}^{j+1}; \mathbf{V}) - \sum_{\ell=0}^{j} \Big(f_{\ell,1}(\mathbf{W}^0, \ldots, \mathbf{W}^\ell; \boldsymbol{\varepsilon})(\zeta_{j+1,\ell})_{11} + f_{\ell,2}(\mathbf{W}^0, \ldots, \mathbf{W}^\ell; \boldsymbol{\varepsilon})(\zeta_{j+1,\ell})_{21} \Big) \\
&= \mathbf{X}\tilde{\boldsymbol{\theta}}_\gamma^{j+1} - \sum_{\ell=0}^{j} f_{\ell,1}(\mathbf{W}^0, \ldots, \mathbf{W}^\ell; \boldsymbol{\varepsilon})(\zeta_{j+1,\ell})_{11}.
\end{aligned}$$

Hence with

$$\mathbf{X}\tilde{\boldsymbol{\theta}}_\gamma^{j+1} - \mathbf{X}\boldsymbol{\theta}^* - \boldsymbol{\varepsilon} = \mathbf{W}_1^{j+1} - \mathbf{W}_2^{j+1} + \sum_{\ell=0}^{j} f_{\ell,1}(\mathbf{W}^0, \ldots, \mathbf{W}^\ell; \boldsymbol{\varepsilon})(\zeta_{j+1,\ell})_{11} - \boldsymbol{\varepsilon},$$

to satisfy (77) we can define $f_{j+1,2} = 0$ and

$$f_{j+1,1}(W^0, \ldots, W^{j+1}; \varepsilon) = -\frac{\beta}{\delta} \Big(W_1^{j+1} - W_2^{j+1} + \sum_{\ell=0}^{j} f_{\ell,1}(W^0, \ldots, W^\ell; \varepsilon)(\zeta_{j+1,\ell})_{11} - \varepsilon \Big). \qquad (79)$$

This is also Lipschitz by the induction hypothesis, completing the induction. So using (76–77), the state evolution (75), and the fact that $X_n \overset{W_2}{\to} X$ implies $f(X_n) \overset{W_2}{\to} f(X)$ for Lipschitz f, we conclude that

$$\frac{1}{d} \sum_{j=1}^{d} \delta_{\left(\theta_j^*, \tilde{\theta}_{\gamma,j}^0, \ldots, \tilde{\theta}_{\gamma,j}^T\right)} \overset{W_2}{\to} \mathsf{P}\big(\theta^*, \theta^0, \ldots, \theta^T\big), \quad \frac{1}{n} \sum_{i=1}^{n} \delta_{\left(\eta_i^*, \varepsilon_i, \tilde{\eta}_{\gamma,i}^0, \ldots, \tilde{\eta}_{\gamma,i}^T\right)} \overset{W_2}{\to} \mathsf{P}\big(W^*, \varepsilon, \eta^0, \ldots, \eta^T\big). \qquad (80)$$

Here, the laws on the right side are defined by setting $W^* = W_2^i$ for each $i \geq 1$, and

$$\theta^i = g_{i,1}(U^1, \ldots, U^i; V), \quad \eta^i = -\frac{\delta}{\beta} f_{i,1}(W^0, \ldots, W^i; \varepsilon) + W^* + \varepsilon,$$

where $\{U^i\} = \{(U_1^i, U_2^i)\}$ and $\{W^i\} = \{(W_1^i, W_2^i)\}$ are the Gaussian processes from AMP state evolution, independent of ε, V with covariance kernels given by (74).

Let us now show that

$$\mathsf{P}(\theta^*, \theta^0, \ldots, \theta^T) = \mathsf{P}\big(\theta^*, \theta_\gamma^0, \ldots, \theta_\gamma^T\big), \qquad \mathsf{P}(W^*, \varepsilon, \eta^0, \ldots, \eta^T) = \mathsf{P}\big(-w_\gamma^*, \varepsilon, \eta_\gamma^0, \ldots, \eta_\gamma^T\big), \tag{81}$$

where the laws on the right sides are the variables of the discrete DMFT equations. This will conclude the proof of (71). To do so, let us define from the AMP state evolution variables (80) the quantities

$$\begin{gathered} u_\gamma^i = \delta U_1^{i+1}, \quad w_\gamma^i = -W_1^i, \quad w_\gamma^* = -W^*, \quad \theta_\gamma^i = \theta^i, \quad \eta_\gamma^i = \eta^i, \\ \frac{\partial \theta_\gamma^i}{\partial u_\gamma^j} = \frac{1}{\delta} \frac{\partial g_{i,1}}{\partial U_1^{j+1}}(U^1, \ldots, U^i; V), \quad \frac{\partial \eta_\gamma^i}{\partial w_\gamma^j} = \frac{\delta}{\beta} \frac{\partial f_{i,1}}{\partial W_1^j}(W^0, \ldots, W^i; \varepsilon). \end{gathered} \tag{82}$$

Then it suffices to check that these quantities satisfy the discrete DMFT equations (57–64), by uniqueness of the iterative construction (65) of the solution to these discrete DMFT equations. We first note that by (74), $\{u_\gamma^j\}$ and $(w_\gamma^*, \{w_\gamma^j\})$ thus defined are centered Gaussian processes with covariance

$$\begin{aligned} \mathbb{E}[u_\gamma^i u_\gamma^j] &= \delta^3 \mathbb{E}[f_{i,1}(W^0, \ldots, W^i; \varepsilon) f_{j,1}(W^0, \ldots, W^j; \varepsilon)] = \delta \beta^2 \mathbb{E}[(\eta^i - W^* - \varepsilon)(\eta^j - W^* - \varepsilon)], \\ \mathbb{E}[w_\gamma^i w_\gamma^j] &= \mathbb{E}[g_{i,1}(U^1, \ldots, U^i; V) g_{j,1}(U^1, \ldots, U^j; V)] = \mathbb{E}[\theta^i \theta^j], \\ \mathbb{E}[w_\gamma^i w_\gamma^*] &= \mathbb{E}[g_{i,1}(U^1, \ldots, U^i; V) g_{0,2}(V)] = \mathbb{E}[\theta^i \theta^*], \\ \mathbb{E}[(w_\gamma^*)^2] &= \mathbb{E}[g_{0,2}(V)^2] = \mathbb{E}[(\theta^*)^2], \end{aligned}$$

which verifies (59) and (63) in light of (82).

We next check the recursions (58) and (62) for the response: Recall that the AMP Onsager corrections are

$$\begin{gathered} (\zeta_{j,s})_{11} = \mathbb{E}\Big[\frac{\partial g_{j,1}}{\partial U_1^{s+1}}(U^1, \ldots, U^j; V)\Big], \\ (\xi_{j,s})_{11} = \mathbb{E}\Big[\delta \frac{\partial f_{j,1}}{\partial W_1^s}(W^0, \ldots, W^j; \varepsilon)\Big], \quad (\xi_{j,s})_{21} = \mathbb{E}\Big[\delta \frac{\partial f_{j,1}}{\partial W_2^s}(W^0, \ldots, W^j; \varepsilon)\Big]. \end{gathered} \tag{83}$$

By definition of $g_{j,1}$ in (78), we have $\frac{\partial g_{j,1}}{\partial U_1^{s+1}} = 0$ for $s \geq j$, $\frac{\partial g_{j,1}}{\partial U_1^{s+1}} = \gamma\delta$ if $s = j-1$, and if $s \leq j-2$,

$$\begin{aligned} \frac{\partial g_{j,1}}{\partial U_1^{s+1}} &= \frac{\partial g_{j-1,1}}{\partial U_1^{s+1}} + \gamma\delta \sum_{\ell=s+1}^{j-1} \frac{\partial g_{\ell,1}}{\partial U_1^{s+1}} (\xi_{j-1,\ell})_{11} + \gamma \partial_\theta s(g_{j-1,1}, \alpha_\gamma^{t_{j-1}}) \frac{\partial g_{j-1,1}}{\partial U_1^{s+1}} \\ &= \Big(1 + \gamma\delta(\xi_{j-1,j-1})_{11} + \gamma \partial_\theta s(g_{j-1,1}, \alpha_\gamma^{t_{j-1}})\Big) \frac{\partial g_{j-1,1}}{\partial U_1^{s+1}} + \gamma\delta \sum_{\ell=s+1}^{j-2} (\xi_{j-1,\ell})_{11} \frac{\partial g_{\ell,1}}{\partial U_1^{s+1}} \end{aligned} \tag{84}$$

where both sides are evaluated at $(U^1, \ldots, U^j; V)$. Similarly, by definition of $f_{j,1}(\cdot)$ in (79), we have $\frac{\partial f_{j,1}}{\partial W_1^s} = \frac{\partial f_{j,1}}{\partial W_2^s} = 0$ if $s > j$, $\frac{\partial f_{j,1}}{\partial W_1^s} = -\frac{\partial f_{j,1}}{\partial W_2^s} = -\beta/\delta$ if $s = j$, and if $s < j$,

$$\frac{\partial f_{j,1}}{\partial W_1^s} = -\frac{\beta}{\delta} \sum_{\ell=s}^{j-1} \frac{\partial f_{\ell,1}}{\partial W_1^s} (\zeta_{j,\ell})_{11} = -\frac{\beta}{\delta}\Big(\sum_{\ell=s+1}^{j-1} \frac{\partial f_{\ell,1}}{\partial W_1^s} (\zeta_{j,\ell})_{11} - \frac{\beta}{\delta}(\zeta_{j,s})_{11}\Big), \tag{85}$$

$$\frac{\partial f_{j,1}}{\partial W_2^s} = -\frac{\beta}{\delta} \sum_{\ell=s}^{j-1} \frac{\partial f_{\ell,1}}{\partial W_2^s} (\zeta_{j,\ell})_{11} = -\frac{\beta}{\delta}\Big(\sum_{\ell=s+1}^{j-1} \frac{\partial f_{\ell,1}}{\partial W_2^s} (\zeta_{j,\ell})_{11} + \frac{\beta}{\delta}(\zeta_{j,s})_{11}\Big). \tag{86}$$

Here, these recursions imply that $\{\frac{\partial f_{j,1}}{\partial W_1^s}\}$ and $\{\frac{\partial f_{j,1}}{\partial W_2^s}\}$ are deterministic. Under the definitions (60), (64), and (82), we have

$$R_\theta^\gamma(j,s) = \frac{1}{\delta} \mathbb{E}\Big[\frac{\partial g_{j,1}}{\partial U_1^{s+1}}(U^1, \ldots, U^j; V)\Big] = \frac{1}{\delta}(\zeta_{j,s})_{11}, \quad R_\eta^\gamma(j,s) = \delta^2 \Big[\frac{\partial f_{j,1}}{\partial W_1^s}\Big] = \delta(\xi_{j,s})_{11}. \tag{87}$$

Then by (84) and (85), $\{\frac{\partial g_{j,1}}{\partial U_1^{s+1}}\}$ and $\{\frac{\partial f_{j,1}}{\partial W_1^s}\}$ satisfy the recursions

$$\frac{\partial g_{j,1}}{\partial U_1^{s+1}} = \Big(1 - \gamma\delta\beta + \gamma\partial_\theta s(g_{j-1,1};\alpha_\gamma^{j-1})\Big)\frac{\partial g_{j-1,1}}{\partial U_1^{s+1}} + \gamma\sum_{\ell=s+1}^{j-2} R_\eta^\gamma(j-1,\ell)\frac{\partial g_{\ell,1}}{\partial U_1^{s+1}},$$
$$\frac{\partial f_{j,1}}{\partial W_1^s} = -\beta\Big(\sum_{\ell=s+1}^{j-1} R_\theta^\gamma(j,\ell)\frac{\partial f_{\ell,1}}{\partial W_1^s} - \frac{\beta}{\delta}R_\theta^\gamma(j,s)\Big),$$

which verify (58) and (62) in view of (82) and above boundary conditions $\frac{\partial g_{j,1}}{\partial U_1^j} = \gamma\delta$ and $\frac{\partial f_{j,1}}{\partial W_1^j} = -\beta/\delta$.

Finally we check the primary recursions (57) and (61). By (82), definition of $g_{j+1,1}(\cdot)$ in (78), and $(\xi_{j,j})_{11} = -(\xi_{j,j})_{21} = -\beta$, we have

$$\begin{aligned}\theta^{j+1} &= g_{j+1,1}(U^1,\ldots,U^{j+1};V)\\ &= \theta^j + \gamma\delta\Big(U_1^{j+1} + \sum_{\ell=0}^{j}\theta^\ell(\xi_{j,\ell})_{11} + \sum_{\ell=0}^{j}(\xi_{j,\ell})_{21}\cdot\theta^*\Big) + \gamma s\big(\theta^j,\alpha_\gamma^j\big) + \sqrt{2}\rho_j\\ &= \theta^j - \gamma\delta\beta(\theta^j-\theta^*) + \gamma s\big(\theta^j,\alpha_\gamma^j\big) + \gamma\delta\sum_{\ell=0}^{j-1}\theta^\ell(\xi_{j,\ell})_{11} + \gamma\delta\sum_{\ell=0}^{j-1}\theta^*(\xi_{j,\ell})_{21} + \gamma\delta U_1^{j+1} + \sqrt{2}\rho_j. \end{aligned} \tag{88}$$

Similarly, by definition of $f_{j,1}(\cdot)$ in (79) and $W_2^i \equiv W^*$,

$$\eta^j = -\frac{\delta}{\beta}f_{j,1}(W^0,\ldots,W^j;\varepsilon) + W^* + \varepsilon = -\frac{\beta}{\delta}\sum_{\ell=0}^{j-1}(\zeta_{j,\ell})_{11}(\eta^\ell - W^* - \varepsilon) + W_1^j. \tag{89}$$

Applying (83), note that in (88) we have $A_j := \sum_{\ell=0}^{j-1}(\xi_{j,\ell})_{21} = \delta\sum_{\ell=0}^{j-1}\frac{\partial f_{j,1}}{\partial W_2^s}$. Then by the recursion (86) and first identification of (87), we have

$$A_j = -\beta\delta\sum_{\ell=0}^{j-1}\sum_{s=\ell}^{j-1}\frac{\partial f_{s,1}}{\partial W_2^\ell}R_\theta^\gamma(j,s) = -\beta\sum_{s=0}^{j-1}R_\theta^\gamma(j,s)\delta\sum_{\ell=0}^{s}\frac{\partial f_{s,1}}{\partial W_2^\ell} = -\beta\sum_{s=0}^{j-1}R_\theta^\gamma(j,s)(A_s+\beta).$$

This coincides with the recursion for $\{\beta\frac{\partial\eta_\gamma^j}{\partial w_\gamma^*}\}$ in (69). Hence $A_j = \beta\frac{\partial\eta_\gamma^j}{\partial w_\gamma^*} = \frac{1}{\delta}R_\eta^\gamma(j,*) = -\frac{1}{\delta}\sum_{s=0}^{j-1}R_\eta^\gamma(j,s)$, where the last step applies Lemma 4.2. Applying this form of A_j and (87), we may write the equations (88–89) as

$$\theta^{j+1} = \theta^j - \gamma\delta\beta(\theta^j-\theta^*) + \gamma s\big(\theta^j,\alpha_\gamma^j\big) + \gamma\sum_{\ell=0}^{j-1}R_\eta^\gamma(j,\ell)(\theta^\ell-\theta^*) + \gamma\delta U_1^{j+1} + \sqrt{2}\rho_j,$$
$$\eta^j = -\beta\sum_{\ell=0}^{j-1}R_\theta^\gamma(j,\ell)(\eta^\ell - W^* - \varepsilon) + W_1^j,$$

which verifies (57) and (61) in view of (82). This verifies that the definitions (82) indeed satisfy (57–64), concluding the proof of (71).

Step 2: Comparison with auxiliary dynamics. Let us now prove (66–68) for the original dynamics with an adaptive drift parameter $\widehat{\alpha}_\gamma^t$. We will prove via induction that, almost surely as $n,d\to\infty$, for each $t = 0,\ldots,T$,

$$\frac{1}{d}\|\boldsymbol{\theta}_\gamma^t - \tilde{\boldsymbol{\theta}}_\gamma^t\|^2 \to 0, \qquad \widehat{\alpha}_\gamma^t \to \alpha_\gamma^t. \tag{90}$$

Since

$$W_2\Big(\frac{1}{d}\sum_{j=1}^{d}\delta_{\left(\theta_j^*,\theta_{\gamma,j}^0,\ldots,\theta_{\gamma,j}^T\right)},\frac{1}{d}\sum_{j=1}^{d}\delta_{\left(\theta_j^*,\tilde{\theta}_{\gamma,j}^0,\ldots,\tilde{\theta}_{\gamma,j}^T\right)}\Big)^2\leq\sum_{t=0}^{T}\frac{1}{d}\|\boldsymbol{\theta}_\gamma^t-\tilde{\boldsymbol{\theta}}_\gamma^t\|^2 \tag{91}$$

$$W_2\Big(\frac{1}{n}\sum_{i=1}^{n}\delta_{\left(\eta_i^*,\varepsilon_i,\eta_{\gamma,i}^0,\ldots,\eta_{\gamma,i}^T\right)},\frac{1}{n}\sum_{i=1}^{n}\delta_{\left(\eta_i^*,\varepsilon_i,\tilde{\eta}_{\gamma,i}^0,\ldots,\tilde{\eta}_{\gamma,i}^T\right)}\Big)^2\leq\sum_{t=0}^{T}\frac{1}{n}\|\boldsymbol{\eta}_\gamma^t-\tilde{\boldsymbol{\eta}}_\gamma^t\|^2\leq\sum_{t=0}^{T}\frac{1}{n}\|\mathbf{X}\|_{\mathrm{op}}^2\|\boldsymbol{\theta}_\gamma^t-\tilde{\boldsymbol{\theta}}_\gamma^t\|^2$$

and $\|\mathbf{X}\|_{\mathrm{op}}$ is almost surely bounded for all large n,d, the above inductive claim together with (71) implies (66–68).

The base case of $t=0$ in (90) holds exactly. Suppose (90) holds up to time t. For $t+1$, we see that

$$\frac{1}{d}\|\boldsymbol{\theta}_\gamma^{t+1}-\tilde{\boldsymbol{\theta}}_\gamma^{t+1}\|^2\leq C\Big((1+\|\mathbf{X}\|_{\mathrm{op}}^4)\frac{1}{d}\|\boldsymbol{\theta}_\gamma^t-\tilde{\boldsymbol{\theta}}_\gamma^t\|^2+\frac{1}{d}\|s(\boldsymbol{\theta}_\gamma^t,\widehat{\alpha}_\gamma^t)-s(\tilde{\boldsymbol{\theta}}_\gamma^t,\alpha_\gamma^t)\|^2\Big).$$

Applying boundedness of $\|\mathbf{X}\|_{\mathrm{op}}$ and Lipschitz continuity of $s(\cdot)$ in Assumption 2.2, we have by the induction hypothesis that $\frac{1}{d}\|\boldsymbol{\theta}_\gamma^{t+1}-\tilde{\boldsymbol{\theta}}_\gamma^{t+1}\|^2\to 0$ almost surely. Next, we have by the Lipschitz continuity of $\mathcal{G}(\cdot)$ in Assumption 2.3 that

$$\begin{aligned}\|\widehat{\alpha}_\gamma^{t+1}-\alpha_\gamma^{t+1}\|&\leq\|\widehat{\alpha}_\gamma^t-\alpha_\gamma^t\|+\gamma\Big\|\mathcal{G}\Big(\widehat{\alpha}_\gamma^t,\frac{1}{d}\sum_{\ell=1}^{d}\delta_{\theta_{\gamma,\ell}^t}\Big)-\mathcal{G}(\alpha_\gamma^t,\mathsf{P}(\theta_\gamma^t))\Big\|\\&\leq C\|\widehat{\alpha}_\gamma^t-\alpha_\gamma^t\|+CW_2\Big(\frac{1}{d}\sum_{\ell=1}^{d}\delta_{\theta_{\gamma,\ell}^t},\mathsf{P}(\theta_\gamma^t)\Big),\end{aligned}$$

which converges almost surely to 0 by the induction hypothesis and the above implication (91). This establishes the induction for (90) and hence completes the proof. □

4.2 Step 2: Discretization error of DMFT equation

We now define a piecewise constant embedding of the components of the discrete DMFT system (57–64) into continuous time, and show that this converges to the solution of the continuous DMFT system established in Theorem 2.4, in the limit $\gamma\to 0$.

For all times $t\in\mathbb{R}_+$, define

$$\lfloor t\rfloor=\max\{i\gamma:i\gamma\leq t,i\in\mathbb{Z}_+\}\in\gamma\mathbb{Z}_+,\quad\lceil t\rceil=\lfloor t\rfloor+\gamma\in\gamma\mathbb{Z}_+,\quad[t]=\lfloor t\rfloor/\gamma\in\mathbb{Z}_+. \tag{92}$$

Fixing $T>0$, let $\mathcal{D}_\eta^\gamma$ be the space of functions $(\bar{R}_\eta^\gamma,\bar{C}_\eta^\gamma,\bar{\alpha}_\gamma)\equiv\{(\bar{R}_\eta^\gamma(t,s),\bar{C}_\eta^\gamma(t,s),\bar{\alpha}_\gamma^t\}_{0\leq s\leq t\leq T}$ that are piecewise constant and right-continuous in (s,t) with jumps at $\gamma\mathbb{Z}_+$, i.e. $\bar{R}_\eta^\gamma(t,s)=\bar{R}_\eta^\gamma(\lfloor t\rfloor,\lfloor s\rfloor)$ for all $0\leq s\leq t\leq T$ and similarly for $\bar{C}_\eta^\gamma,\bar{\alpha}_\gamma$. Analogously, let $\mathcal{D}_\theta^\gamma$ be the space of functions $(\bar{R}_\theta^\gamma,\bar{C}_\theta^\gamma,\bar{\bar{\alpha}}_\gamma)\equiv\{\bar{R}_\theta^\gamma(t,s),\bar{C}_\theta^\gamma(t,s),\bar{\bar{\alpha}}_\gamma^t\}_{0\leq s\leq t\leq T}$ that are piecewise constant and right-continuous with jumps at $\gamma\mathbb{Z}_+$.

We define a map $\mathcal{T}_{\eta\to\theta}^\gamma:\mathcal{D}_\eta^\gamma\to\mathcal{D}_\theta^\gamma$ as follows: Given $X^\gamma=(\bar{R}_\eta^\gamma,\bar{C}_\eta^\gamma,\bar{\alpha}_\gamma)\in\mathcal{D}_\eta^\gamma$, let $\{\bar{u}_\gamma^t\}_{t\in[0,T]}$ be a mean-zero Gaussian process with covariance $\bar{C}_\eta^\gamma$, let $\{b^t\}_{t\geq 0}$ be a standard Brownian motion, and define processes $\{\bar{\theta}_\gamma^t\}_{t\in[0,T]}$ and $\{\frac{\partial\bar{\theta}_\gamma^t}{\partial\bar{u}_\gamma^s}\}_{0\leq s\leq t\leq T}$ by

$$\bar{\theta}_\gamma^t=\theta^0+\int_0^{\lfloor t\rfloor}\Big[-\delta\beta(\bar{\theta}_\gamma^s-\theta^*)+s(\bar{\theta}_\gamma^s,\bar{\alpha}_\gamma^s)+\int_0^{\lfloor s\rfloor}\bar{R}_\eta^\gamma(s,r)(\bar{\theta}_\gamma^r-\theta^*)\mathrm{d}r+\bar{u}_\gamma^s\Big]\mathrm{d}s+\sqrt{2}b^{\lfloor t\rfloor}, \tag{93}$$

$$\frac{\partial\bar{\theta}_\gamma^t}{\partial\bar{u}_\gamma^s}=1+\mathbf{1}\{\lceil s\rceil\leq\lfloor t\rfloor\}\int_{\lceil s\rceil}^{\lfloor t\rfloor}\Big[\Big(-\delta\beta+\partial_\theta s(\bar{\theta}_\gamma^r,\bar{\alpha}_\gamma^r)\Big)\frac{\partial\bar{\theta}_\gamma^r}{\partial\bar{u}_\gamma^s}+\int_{\lceil s\rceil}^{\lfloor r\rfloor}\bar{R}_\eta^\gamma(r,r')\frac{\partial\bar{\theta}_\gamma^{r'}}{\partial\bar{u}_\gamma^s}\mathrm{d}r'\Big]\mathrm{d}r. \tag{94}$$

Define from these processes

$$\begin{gathered}\bar{C}_\theta^\gamma(t,s)=\mathbb{E}\Big[\bar{\theta}_\gamma^t\bar{\theta}_\gamma^s\Big],\quad\bar{C}_\theta^\gamma(t,*)=\mathbb{E}\Big[\bar{\theta}_\gamma^t\theta^*\Big],\quad\bar{C}_\theta^\gamma(*,*)=\mathbb{E}[(\theta^*)^2],\\\bar{R}_\theta^\gamma(t,s)=\mathbb{E}\Big[\frac{\partial\bar{\theta}_\gamma^t}{\partial\bar{u}_\gamma^s}\Big],\quad\bar{\bar{\alpha}}_\gamma^t=\alpha^0+\int_0^{\lfloor t\rfloor}\mathcal{G}(\bar{\bar{\alpha}}_\gamma^s,\mathsf{P}(\bar{\theta}_\gamma^s))\mathrm{d}s\end{gathered} \tag{95}$$

and set $\mathcal{T}^\gamma_{\eta\to\theta}(X^\gamma) = (\bar{R}^\gamma_\theta, \bar{C}^\gamma_\theta, \bar{\bar{\alpha}}_\gamma)$. We also define a map $\mathcal{T}^\gamma_{\theta\to\eta} : \mathcal{D}^\gamma_\theta \to \mathcal{D}^\gamma_\eta$ as follows: Given $Y^\gamma = (\bar{R}^\gamma_\theta, \bar{C}^\gamma_\theta, \bar{\bar{\alpha}}_\gamma) \in \mathcal{D}^\gamma_\eta$, let $(\bar{w}^*_\gamma, \{\bar{w}^t_\gamma\}_{t\in[0,T]})$ be a mean-zero Gaussian process with covariance $\bar{C}^\gamma_\theta$, and define

$$\bar{\eta}^t_\gamma = -\beta \int_0^{\lfloor t\rfloor} \bar{R}^\gamma_\theta(t,s)(\bar{\eta}^s_\gamma + \bar{w}^*_\gamma - \varepsilon)\mathrm{d}s - \bar{w}^t_\gamma, \tag{96}$$

$$\frac{\partial \bar{\eta}^t_\gamma}{\partial \bar{w}^s_\gamma} = \beta \bar{R}_\theta(t,s) - \mathbf{1}\{\lceil s\rceil \le \lfloor t\rfloor\}\beta \int_{\lceil s\rceil}^{\lfloor t\rfloor} \bar{R}^\gamma_\theta(t,r)\frac{\partial \bar{\eta}^r_\gamma}{\partial \bar{w}^s_\gamma}\mathrm{d}r. \tag{97}$$

Define from these processes

$$\bar{R}^\gamma_\eta(t,s) = \delta\beta\,\mathbb{E}\Big[\frac{\partial \bar{\eta}^t_\gamma}{\partial \bar{w}^s_\gamma}\Big], \quad \bar{C}^\gamma_\eta(t,s) = \delta\beta^2\mathbb{E}\Big[(\bar{\eta}^t_\gamma + \bar{w}^*_\gamma - \varepsilon)(\bar{\eta}^s_\gamma + \bar{w}^*_\gamma - \varepsilon)\Big], \quad \bar{\alpha}^t_\gamma = \bar{\bar{\alpha}}^t_\gamma \tag{98}$$

and set $\mathcal{T}^\gamma_{\theta\to\eta}(Y^\gamma) = (\bar{R}^\gamma_\eta, \bar{C}^\gamma_\eta, \bar{\alpha}_\gamma)$. These maps may be understood as discrete approximations of $\mathcal{T}_{\eta\to\theta}$ and $\mathcal{T}_{\theta\to\eta}$ constructed in Section 3.2, with domains restricted to the spaces $\mathcal{D}^\gamma_\theta$ and $\mathcal{D}^\gamma_\eta$ of piecewise constant inputs.

Recall the spaces $\mathcal{S}_\theta, \mathcal{S}_\eta, \mathcal{S}$ of Section 3.1. The following lemma shows that the above maps $\mathcal{T}^\gamma_{\eta\to\theta}, \mathcal{T}^\gamma_{\theta\to\eta}$ are well-defined, that the unique fixed point of these maps is a piecewise constant embedding of the discrete-time DMFT system in Section 4.1, and that furthermore this fixed point belongs to $\mathcal{S}$.

Lemma 4.3. *(a) Given any $X^\gamma \in \mathcal{D}^\gamma_\eta$ and realization of θ^*, θ^0, $\{b^t\}_{t\in[0,T]}$, and $\{\bar{u}^t_\gamma\}_{t\in[0,T]}$, the processes (93–94) have a unique solution, and this solution is piecewise constant and right-continuous with jumps at $\gamma\mathbb{Z}_+$. Consequently, $\mathcal{T}^\gamma_{\eta\to\theta}$ is a well-defined map from $\mathcal{D}^\gamma_\eta$ to $\mathcal{D}^\gamma_\theta$.*

(b) Given any $Y^\gamma \in \mathcal{D}^\gamma_\theta$ and realization of ε and $(\bar{w}^*_\gamma, \{\bar{w}^t_\gamma\})$, the processes (96–97) have a unique solution, and this solution is piecewise constant and right-continuous with jumps at $\gamma\mathbb{Z}_+$. Consequently, $\mathcal{T}^\gamma_{\theta\to\eta}$ is a well-defined map from $\mathcal{D}^\gamma_\theta$ to $\mathcal{D}^\gamma_\eta$.

(c) The map $\mathcal{T}^\gamma_{\eta\to\eta} = \mathcal{T}^\gamma_{\eta\to\theta} \circ \mathcal{T}^\gamma_{\theta\to\eta}$ has a unique fixed point in $\mathcal{D}^\gamma_\eta$, and the map $\mathcal{T}^\gamma_{\theta\to\theta} = \mathcal{T}^\gamma_{\theta\to\eta} \circ \mathcal{T}^\gamma_{\eta\to\theta}$ has a unique fixed point in $\mathcal{D}^\gamma_\theta$. These fixed points are given precisely by

$$\begin{gathered}\bar{\alpha}^t_\gamma = \bar{\bar{\alpha}}^t_\gamma = \alpha^{[t]}_\gamma, \quad \bar{C}^\gamma_\theta(t,s) = C^\gamma_\theta([t],[s]), \quad \bar{C}^\gamma_\theta(t,*) = C^\gamma_\theta([t],*), \quad \bar{C}^\gamma_\eta(t,s) = C^\gamma_\eta([t],[s]),\\ \bar{R}^\gamma_\theta(t,s) = \begin{cases} 1 & \text{if } [s]=[t] \\ \frac{1}{\gamma}R^\gamma_\theta([t],[s]) & \text{if } [s]<[t], \end{cases} \quad \bar{R}^\gamma_\eta(t,s) = \begin{cases} \delta\beta^2 & \text{if } [s]=[t] \\ \frac{1}{\gamma}R^\gamma_\eta([t],[s]) & \text{if } [s]<[t] \end{cases}\end{gathered} \tag{99}$$

for all $0 \le s \le t \le T$, where $(\alpha_\gamma, C^\gamma_\theta, C^\gamma_\eta, R^\gamma_\theta, R^\gamma_\eta)$ are the components of the discrete DMFT system defined iteratively in time via (60) and (64).

(d) For any $\gamma > 0$ sufficiently small, we have $\mathcal{T}^\gamma_{\eta\to\theta}(\mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta) \subseteq \mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$, $\mathcal{T}^\gamma_{\theta\to\eta}(\mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta) \subseteq \mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$, and the fixed point (99) belongs to $\mathcal{S}$.

Proof. For (a), if $X^\gamma = (\bar{R}^\gamma_\eta, \bar{C}^\gamma_\eta, \bar{\alpha}_\gamma) \in \mathcal{D}^\gamma_\eta$, then $\{\bar{u}^t_\gamma\}$ is also piecewise constant and right-continuous by these properties of $\bar{C}^\gamma_\eta$. Then an easy induction on k shows that (93) has a unique solution over $t \in [0, k\gamma)$ for each integer $k \ge 1$, which is given by $\bar{\theta}^t_\gamma = \bar{\theta}^{\lfloor t\rfloor}_\gamma$. By definition, (94) is given by $\frac{\partial \bar{\theta}^t_\gamma}{\partial \bar{u}^s_\gamma} = 1$ for all $s \ge 0$ and $t \in [s, \lceil s\rceil)$. Then for each $s \ge 0$, an induction on k shows also that (94) has a unique solution on $[s, \lceil s\rceil + k\gamma)$ for each integer $k \ge 1$, which is given by $\frac{\partial \bar{\theta}^t_\gamma}{\partial \bar{u}^s_\gamma} = \frac{\partial \bar{\theta}^{\lfloor t\rfloor}_\gamma}{\partial \bar{u}^s_\gamma}$, and furthermore this solution depends on s only via $\lfloor s\rfloor$, i.e. $\frac{\partial \bar{\theta}^t_\gamma}{\partial \bar{u}^s_\gamma} = \frac{\partial \bar{\theta}^{\lfloor t\rfloor}_\gamma}{\partial \bar{u}^{\lfloor s\rfloor}_\gamma}$. Thus the solutions of (93–94) are piecewise constant and right-continuous, implying the same properties for $\bar{R}^\gamma_\theta, \bar{C}^\gamma_\theta, \bar{\bar{\alpha}}$ defined by (95). This shows (a).

Part (b) follows from analogous inductive arguments, using that if $Y^\gamma = (\bar{R}^\gamma_\theta, \bar{C}^\gamma_\theta, \bar{\bar{\alpha}}_\gamma) \in \mathcal{D}^\gamma_\theta$, then $\{\bar{w}^t_\gamma\}$ is also piecewise constant and right-continuous, and hence so are $\{\bar{\eta}^t_\gamma\}$ and $\{\frac{\bar{\eta}^t_\gamma}{\bar{w}^s_\gamma}\}$.

Part (c) also follows by induction: Since any fixed point is piecewise constant, it suffices to consider the values at $\gamma\mathbb{Z}_+$. By (93), $\bar{\theta}^0_\gamma = \theta^0$. Then by (95),

$$\bar{R}^\gamma_\theta(0,0) = 1, \quad \bar{C}^\gamma_\theta(0,0) = \mathbb{E}[(\theta^0)^2] = C^\gamma_\theta(0,0), \quad \bar{C}^\gamma_\theta(0,*) = \mathbb{E}[\theta^0\theta^*] = C^\gamma_\theta(0,*), \quad \bar{\bar{\alpha}}^0_\gamma = 0.$$

Then by (96), $\bar{\eta}_\gamma^0 = -\bar{w}_\gamma^0$, so $(\bar{\eta}_\gamma^0, \bar{w}_\gamma^*)$ is equal in joint law to the discrete DMFT variables $(\eta_\gamma^0, w_\gamma^*)$. Then by (98), any fixed point must satisfy

$$\bar{R}_\eta^\gamma(0,0) = \delta\beta^2, \quad \bar{C}_\eta^\gamma(0,0) = \delta\beta^2\mathbb{E}[(\eta_\gamma^0 + w_\gamma^* - \varepsilon)^2] = C_\eta(0,0), \quad \bar{\alpha}_\gamma^0 = 0.$$

Suppose inductively that there is a unique fixed point over times $s \le t$ in $\{0, \gamma, \ldots, k\gamma\}$, and consider now $t = (k+1)\gamma$. The equations (93–94) and the piecewise constant nature of all processes imply

$$\begin{aligned}
\bar{\theta}_\gamma^{(k+1)\gamma} &= \bar{\theta}_\gamma^{k\gamma} + \int_{k\gamma}^{(k+1)\gamma} \Big[-\delta\beta(\bar{\theta}_\gamma^s - \theta^*) + s(\bar{\theta}_\gamma^s, \bar{\alpha}_\gamma^s) + \int_0^{\lfloor s \rfloor} \bar{R}_\eta^\gamma(s,r)(\bar{\theta}_\gamma^r - \theta^*)\mathrm{d}r + \bar{u}_\gamma^s\Big]\mathrm{d}s + \sqrt{2}(b^{(k+1)\gamma} - b^{k\gamma}) \\
&= \bar{\theta}_\gamma^{k\gamma} + \gamma\Big(-\delta\beta(\bar{\theta}_\gamma^{k\gamma} - \theta^*) + s(\bar{\theta}_\gamma^{k\gamma}, \bar{\alpha}_\gamma^{k\gamma}) + \gamma\sum_{\ell=0}^{k-1} \bar{R}_\eta^\gamma(k\gamma, \ell\gamma)(\bar{\theta}_\gamma^{\ell\gamma} - \theta^*) + \bar{u}_\gamma^{k\gamma}\Big) + \sqrt{2}(b^{(k+1)\gamma} - b^{k\gamma}),
\end{aligned}$$

$\frac{\partial\bar{\theta}_\gamma^{(k+1)\gamma}}{\partial\bar{u}_\gamma^{(k+1)\gamma}} = 1$, and for any $j \le k$,

$$\begin{aligned}
\frac{\partial\bar{\theta}_\gamma^{(k+1)\gamma}}{\partial\bar{u}_\gamma^{j\gamma}} &= \frac{\partial\bar{\theta}_\gamma^{k\gamma}}{\partial\bar{u}_\gamma^{j\gamma}} + \int_{k\gamma}^{(k+1)\gamma} \Big[\Big(-\delta\beta + \partial_\theta s(\bar{\theta}_\gamma^r, \bar{\alpha}_\gamma^r)\Big)\frac{\partial\bar{\theta}_\gamma^r}{\partial\bar{u}_\gamma^{j\gamma}} + \int_{(j+1)\gamma}^{k\gamma} \bar{R}_\eta^\gamma(r,r')\frac{\partial\bar{\theta}_\gamma^{r'}}{\partial\bar{u}_\gamma^{j\gamma}}\mathrm{d}r'\Big]\mathrm{d}r \\
&= \frac{\partial\bar{\theta}_\gamma^{k\gamma}}{\partial\bar{u}_\gamma^{j\gamma}} + \gamma\Big[\Big(-\delta\beta + \partial_\theta s(\bar{\theta}_\gamma^{k\gamma}, \bar{\alpha}_\gamma^{k\gamma})\Big)\frac{\partial\bar{\theta}_\gamma^{k\gamma}}{\partial\bar{u}_\gamma^{j\gamma}} + \gamma\sum_{\ell=j+1}^{k-1} \bar{R}_\eta^\gamma(k\gamma, \ell\gamma)\frac{\partial\bar{\theta}_\gamma^{\ell\gamma}}{\partial\bar{u}_\gamma^{j\gamma}}\Big]
\end{aligned}$$

Comparing these equations with (57–58) and applying $\gamma\bar{R}_\eta^\gamma(k\gamma, \ell\gamma) = R_\eta^\gamma(k,\ell)$, $\bar{\alpha}_\gamma^{k\gamma} = \alpha_\gamma^k$, and the equality in law $(\bar{u}_\gamma^0, \ldots, \bar{u}_\gamma^{k\gamma}) \overset{L}{=} (u_\gamma^0, \ldots, u_\gamma^k)$ by the induction hypothesis, this shows the equality in law

$$\{\theta^*, \bar{\theta}_\gamma^{i\gamma}, \tfrac{\partial\bar{\theta}_\gamma^{i\gamma}}{\partial\bar{u}_\gamma^{j\gamma}}\}_{i<j\le k+1} \overset{L}{=} \{\theta^*, \theta_\gamma^i, \gamma^{-1}\tfrac{\partial\theta_\gamma^i}{\partial u_\gamma^j}\}_{i<j\le k+1}.$$

Then (99) holds for the components $\bar{R}_\theta^\gamma, \bar{C}_\theta^\gamma, \bar{\bar{\alpha}}_\gamma$ of any fixed point up to times $s \le t$ in $\{0, \gamma, \ldots, (k+1)\gamma\}$.

Now the equations (96–97) imply

$$\begin{aligned}
\bar{\eta}_\gamma^{(k+1)\gamma} &= -\beta\int_0^{(k+1)\gamma} \bar{R}_\theta^\gamma((k+1)\gamma, s)(\bar{\eta}_\gamma^s + \bar{w}_\gamma^* - \varepsilon)\mathrm{d}s - \bar{w}_\gamma^{(k+1)\gamma} \\
&= -\beta\gamma\sum_{j=0}^{k} \bar{R}_\theta^\gamma((k+1)\gamma, j\gamma)\big(\bar{\eta}_\gamma^{j\gamma} + \bar{w}_\gamma^* - \varepsilon\big) - \bar{w}_\gamma^{(k+1)\gamma},
\end{aligned}$$

$\frac{\partial\bar{\eta}_\gamma^{(k+1)\gamma}}{\partial\bar{w}_\gamma^{(k+1)\gamma}} = \delta\beta^2$, and for all $j \le k$,

$$\begin{aligned}
\frac{\partial\bar{\eta}_\gamma^{(k+1)\gamma}}{\partial\bar{w}_\gamma^{j\gamma}} &= \beta\bar{R}_\theta((k+1)\gamma, j\gamma) - \beta\int_{(j+1)\gamma}^{(k+1)\gamma} \bar{R}_\theta^\gamma((k+1)\gamma, r)\frac{\partial\bar{\eta}_\gamma^r}{\partial\bar{w}_\gamma^{j\gamma}}\mathrm{d}r \\
&= \beta\bar{R}_\theta((k+1)\gamma, j\gamma) - \beta\gamma\sum_{\ell=j+1}^{k} \bar{R}_\theta^\gamma((k+1)\gamma, \ell\gamma)\frac{\partial\bar{\eta}_\gamma^{\ell\gamma}}{\partial\bar{w}_\gamma^{j\gamma}}.
\end{aligned}$$

Comparing these equations with (61–62) and applying again the induction hypothesis, this shows the equality in law

$$\{\bar{\eta}_\gamma^{i\gamma}, \tfrac{\partial\bar{\eta}_\gamma^{i\gamma}}{\partial\bar{w}_\gamma^{j\gamma}}\}_{i<j\le k+1} \overset{L}{=} \{\eta_\gamma^i, \gamma^{-1}\tfrac{\partial\eta_\gamma^i}{\partial w_\gamma^j}\}_{i<j\le k+1}.$$

Then (99) also holds for the components $\bar{R}_\eta^\gamma, \bar{C}_\eta^\gamma, \bar{\alpha}_\gamma$ of any fixed point up to times $s \le t$ in $\{0, \gamma, \ldots, (k+1)\gamma\}$, completing the induction. Thus any fixed points of $\mathcal{T}_{\eta\to\eta}^\gamma$ and $\mathcal{T}_{\theta\to\theta}^\gamma$ must satisfy (99) for all $0 \le s \le t \le T$, implying also that such fixed points are unique in $\mathcal{D}_\eta^\gamma$ and $\mathcal{D}_\theta^\gamma$ by uniqueness of the iterative construction (65) of the solution to the discrete DMFT equations. This shows (c).

For (d), we check that $\mathcal{T}^\gamma_{\eta\to\eta}$ and $\mathcal{T}^\gamma_{\theta\to\theta}$ define contractive mappings on $\mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$ and $\mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$. Note that if $Y^\gamma = (\bar{R}^\gamma_\theta, \bar{C}^\gamma_\theta, \bar{\bar{\alpha}}_\gamma) \in \mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$ and $X^\gamma = \mathcal{T}^\gamma_{\theta\to\eta}(Y^\gamma) = (\bar{R}^\gamma_\eta, \bar{C}^\gamma_\eta, \bar{\alpha}_\gamma)$, then setting $\bar{\xi}^t_\gamma = \bar{\eta}^t_\gamma + \bar{w}^*_\gamma - \varepsilon$, the same arguments as in Lemma 3.5 show

$$\mathbb{E}(\bar{\xi}^t_\gamma)^2 \le 2\Big[\beta^2 \int_0^{\lfloor t\rfloor} (t-s+1)^2 \cdot \Phi^2_{R_\theta}(t-s)\mathbb{E}(\bar{\xi}^s_\gamma)^2 \mathrm{d}s + 2\Phi_{C_\theta}(t) + 2\tau_*^2 + \sigma^2\Big],$$
$$|\bar{R}^\gamma_\eta(t,s)| \le |\beta|\Big(\mathbf{1}\{\lceil s\rceil \le \lfloor t\rfloor\} \int_{\lceil s\rceil}^{\lfloor t\rfloor} \Phi_{R_\theta}(t-s')|\bar{R}^\gamma_\eta(s',s)|\mathrm{d}s' + \delta|\beta|\Phi_{R_\theta}(t-s)\Big).$$

Upper-bounding these integrals $\int_0^{\lfloor t\rfloor}$ and $\int_{\lceil s\rceil}^{\lfloor t\rfloor}$ by $\int_0^t$ and $\int_s^t$, we obtain as in Lemma 3.5 that $\bar{C}^\gamma_\eta(t,t) \le \Phi_{C_\gamma}(t)$ and $\bar{R}^\gamma_\eta(t,s) \le \Phi_{R_\gamma}(t-s)$. All continuity conditions defining $\mathcal{S}_\eta$ are automatically satisfied since the components of X^γ are piecewise constant outside the knots $D = [0,T] \cap \gamma\mathbb{Z}_+$. Thus $X^\gamma \in \mathcal{S}_\eta$, i.e. $\mathcal{T}^\gamma_{\theta\to\eta}$ maps $\mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$ into $\mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$.

Conversely, suppose $X^\gamma = (\bar{R}^\gamma_\eta, \bar{C}^\gamma_\eta, \bar{\alpha}) \in \mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$, and let $Y^\gamma = \mathcal{T}^\gamma_{\theta\to\eta}(\bar{R}^\gamma_\theta, \bar{C}^\gamma_\theta, \bar{\bar{\alpha}})$. A small extension of the argument in Lemma 3.6 shows $Y^\gamma \in \mathcal{S}_\theta$. Let us explain this extension for $\bar{C}^\gamma_\theta$: Defining

$$\bar{v}^t_\gamma = -\delta\beta(\bar{\theta}^t_\gamma - \theta^*) + s(\bar{\theta}^t_\gamma, \bar{\alpha}^t_\gamma) + \int_0^{\lfloor t\rfloor} \bar{R}^\gamma_\eta(t,s)(\bar{\theta}^s_\gamma - \theta^*)\mathrm{d}s + \bar{u}^t_\gamma,$$

we have $\bar{\theta}^{t+\gamma}_\gamma = \bar{\theta}^t_\gamma + \gamma \cdot \bar{v}^t_\gamma + \sqrt{2}(b^{t+\gamma} - b^t)$ for $t \in \gamma\mathbb{Z}_+$. Then, analogous to the calculation using Ito's formula in Lemma 3.6, for sufficiently small $\gamma > 0$,

$$\begin{aligned}\bar{C}^\gamma_\theta(t+\gamma, t+\gamma) - \bar{C}^\gamma_\theta(t,t) &= 2\gamma\mathbb{E}\bar{\theta}^t_\gamma \bar{v}^t_\gamma + \gamma^2\mathbb{E}(\bar{v}^t_\gamma)^2 + 2\gamma \\ &\le \frac{\gamma}{1-\gamma}\mathbb{E}(\bar{\theta}^t_\gamma)^2 + [\gamma(1-\gamma)+\gamma^2]\mathbb{E}(\bar{v}^t_\gamma)^2 + 2\gamma \\ &\le \gamma\Big(1.1\mathbb{E}(\bar{\theta}^t_\gamma)^2 + \mathbb{E}(\bar{v}^t_\gamma)^2 + 2\Big) = \int_t^{t+\gamma}\Big(1.1\mathbb{E}(\bar{\theta}^s_\gamma)^2 + \mathbb{E}(\bar{v}^s_\gamma)^2 + 2\Big)\mathrm{d}s,\end{aligned}$$

the last equality holding because $\bar{v}_\gamma$ and $\bar{\theta}_\gamma$ are piecewise constant. Summing this inequality shows

$$\bar{C}^\gamma_\theta(t,t) - \bar{C}^\gamma_\theta(0,0) \le \int_0^t \Big(1.1\mathbb{E}(\bar{\theta}^s_\gamma)^2 + \mathbb{E}(\bar{v}^s_\gamma)^2 + 2\Big)\mathrm{d}s$$

for all $t \in [0,T]$. Then, bounding $\mathbb{E}(\bar{v}^s_\gamma)^2$ as in (44) of Lemma 3.6, this shows that $\bar{C}^\gamma_\theta(t,t) \le \Phi_{C_\theta}(t)$. Similar extensions of the arguments in Lemma 3.6 show that $\bar{R}^\gamma_\theta(t,s) \le \Phi_{R_\theta}(t-s)$ and $\|\bar{\alpha}^t_\gamma\|^2 \le \Phi_\alpha(t)$, so $Y^\gamma \in \mathcal{S}_\theta$ as claimed. Then $\mathcal{T}^\gamma_{\eta\to\theta}$ maps $\mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$ into $\mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$.

The same argument as in Lemmas 3.7 and 3.8 bound the moduli-of-continuity of $\mathcal{T}^\gamma_{\eta\to\theta}$ and $\mathcal{T}^\gamma_{\theta\to\eta}$ in the metrics $d(\cdot)$ of Section 3.2, implying that $\mathcal{T}^\gamma_{\eta\to\eta}$ and $\mathcal{T}^\gamma_{\theta\to\theta}$ are contractive for sufficiently large $\lambda > 0$ defining $d(\cdot)$. These metrics induce the topologies of uniform convergence on the spaces $\mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$ and $\mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$, which are equivalent to closed subsets of finite-dimensional vector spaces and hence also complete. Then $\mathcal{T}^\gamma_{\eta\to\eta}$ and $\mathcal{T}^\gamma_{\theta\to\theta}$ have unique fixed points in $\mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$ and $\mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$ by the Banach fixed-point theorem. These must coincide with the fixed point (99), by the uniqueness statement (without restriction to $\mathcal{S}_\theta$ and $\mathcal{S}_\eta$) shown in part (c). Thus this fixed point (99) belongs to $\mathcal{S}$. □

Lemma 4.4. *There exists a constant $C > 0$ (depending on T but not on λ, γ) such that for all large enough $\lambda > 0$ defining the metrics (47) and all sufficiently small $\gamma > 0$,*

(a) For any $X^\gamma \in \mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$,

$$d\big(\mathcal{T}_{\eta\to\theta}(X^\gamma), \mathcal{T}^\gamma_{\eta\to\theta}(X^\gamma)\big) \le C\sqrt{\gamma}.$$

(b) For any $Y^\gamma \in \mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$,

$$d\big(\mathcal{T}_{\theta\to\eta}(Y^\gamma), \mathcal{T}^\gamma_{\theta\to\eta}(Y^\gamma)\big) \le C\sqrt{\gamma}.$$

Proof. We show part (a). Consider any $X^\gamma = (\bar{R}_\eta^\gamma, \bar{C}_\eta^\gamma, \bar{\alpha}_\gamma)$, and denote $\mathcal{T}_{\eta\to\theta}(X^\gamma) = (R_\theta, C_\theta, \tilde{\alpha})$ and $\mathcal{T}_{\eta\to\theta}^\gamma(X^\gamma) = (\bar{R}_\theta^\gamma, \bar{C}_\theta^\gamma, \bar{\tilde{\alpha}}_\gamma)$. Throughout, $C, C' > 0$ denote constants depending on T but not on λ, γ and changing from instance to instance.

Bound of $d(C_\theta, \bar{C}_\theta^\gamma)$**.** Given X^γ, let us couple the evolutions (12) and (93) by a common realization of $\{\bar{u}_\gamma^t\}$ with covariance $\bar{C}_\eta^\gamma$ and a common Brownian motion. Then by definition, we have

$$\theta^t = \theta^0 + \int_0^t \Big[-\delta\beta(\theta^s - \theta^*) + s(\theta^s, \bar{\alpha}_\gamma^s) + \int_0^s \bar{R}_\eta^\gamma(s,s')(\theta^{s'} - \theta^*)\mathrm{d}s' + \bar{u}_\gamma^s\Big]\mathrm{d}s + \sqrt{2}\, b^t,$$
$$\bar{\theta}_\gamma^t = \theta^0 + \int_0^{\lfloor t\rfloor} \Big[-\delta\beta(\bar{\theta}_\gamma^s - \theta^*) + s(\bar{\theta}_\gamma^s, \bar{\alpha}_\gamma^s) + \int_0^{\lfloor s\rfloor} \bar{R}_\eta^\gamma(s,s')(\bar{\theta}_\gamma^{s'} - \theta^*)\mathrm{d}s' + \bar{u}_\gamma^s\Big]\mathrm{d}s + \sqrt{2}\, b^{\lfloor t\rfloor}.$$

Then $\mathbb{E}(\theta^t - \bar{\theta}_\gamma^t)^2 \le 6[(I) + (II) + (III) + (IV) + (V) + (VI)]$ where

$$\begin{aligned}
(I) &= \mathbb{E}\Big(\int_0^{\lfloor t\rfloor} \delta\beta(\theta^s - \bar{\theta}_\gamma^s)\mathrm{d}s\Big)^2,\\
(II) &= \mathbb{E}\Big(\int_0^{\lfloor t\rfloor} (s(\theta^s, \bar{\alpha}_\gamma^s) - s(\bar{\theta}_\gamma^s, \bar{\alpha}_\gamma^s)\mathrm{d}s\Big)^2,\\
(III) &= \mathbb{E}\Big(\int_0^{\lfloor t\rfloor}\int_0^{\lfloor s\rfloor} \bar{R}_\eta^\gamma(s,s')(\theta^{s'} - \bar{\theta}_\gamma^{s'})\mathrm{d}s'\,\mathrm{d}s\Big)^2,\\
(IV) &= \mathbb{E}\Big(\int_0^{\lfloor t\rfloor}\int_{\lfloor s\rfloor}^{s} \bar{R}_\eta^\gamma(s,s')(\theta^{s'} - \theta^*)\mathrm{d}s'\,\mathrm{d}s\Big)^2,\\
(V) &= \mathbb{E}\Big(\int_{\lfloor t\rfloor}^{t} \Big[-\delta\beta(\theta^s - \theta^*) + s(\theta^s, \bar{\alpha}_\gamma^s) + \int_0^s \bar{R}_\eta^\gamma(s,s')(\theta^{s'} - \theta^*)\mathrm{d}s' + \bar{u}_\gamma^s\Big]\mathrm{d}s\Big)^2,\\
(VI) &= \mathbb{E}(\sqrt{2}\, b^t - \sqrt{2}\, b^{\lfloor t\rfloor})^2.
\end{aligned}$$

By the same arguments as in the proof of Lemma 3.7, using the Lipschitz continuity of $s(\cdot)$ in Assumption 2.2, we may show

$$(I) + (II) + (III) \le \frac{C}{\lambda} e^{2\lambda t} \sup_{s\in[0,T]} e^{-2\lambda s}\mathbb{E}(\theta^s - \bar{\theta}_\gamma^s)^2.$$

Applying $s - \lfloor s\rfloor \le \gamma$, $t - \lfloor t\rfloor \le \gamma$, and the bounds for $\bar{R}_\eta^\gamma, \bar{C}_\eta^\gamma$ implied by $X^\gamma \in \mathcal{S}_\eta$, we have

$$(IV) + (V) + (VI) \le C\gamma.$$

Then

$$\sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta^t - \bar{\theta}_\gamma^t)^2 \le \frac{C}{\lambda} \sup_{t\in[0,T]} e^{-2\lambda t}\mathbb{E}(\theta^t - \bar{\theta}_\gamma^t)^2 + C\gamma,$$

and choosing large enough $\lambda > 0$ yields

$$\sup_{t\in[0,T]} e^{-\lambda t}\sqrt{\mathbb{E}(\theta^t - \bar{\theta}_\gamma^t)^2} \le C'\sqrt{\gamma}. \tag{100}$$

This implies as in the proof of Lemma 3.7 that $d(C_\theta, \bar{C}_\theta^\gamma) \le C'\sqrt{\gamma}$.

Bound of $d(R_\theta, \bar{R}_\theta^\gamma)$**.** Denote by $r_\theta(t,s) = \frac{\partial\theta^t}{\partial u^s}$ and $\bar{r}_\theta^\gamma(t,s) = \frac{\partial\bar{\theta}_\gamma^t}{\partial\bar{u}_\gamma^s}$ the processes (13) and (94) defined from the above coupling of $\{\theta^t\}$ and $\{\bar{\theta}_\gamma^t\}$. Then by definition, we have

$$r_\theta(t,s) = 1 + \int_s^t \Big[\Big(-\delta\beta + \partial_\theta s(\theta^{s'}, \bar{\alpha}_\gamma^{s'})\Big)r_\theta(s',s) + \int_s^{s'} \bar{R}_\eta^\gamma(s',s'')r_\theta(s'',s)\mathrm{d}s''\Big]\mathrm{d}s',$$
$$\bar{r}_\theta^\gamma(t,s) = 1 + \mathbf{1}\{\lceil s\rceil \le \lfloor t\rfloor\}\int_{\lceil s\rceil}^{\lfloor t\rfloor} \Big[\Big(-\delta\beta + \partial_\theta s(\bar{\theta}_\gamma^{s'}, \bar{\alpha}_\gamma^{s'})\Big)\bar{r}_\theta^\gamma(s',s) + \int_{\lceil s\rceil}^{\lfloor s'\rfloor} \bar{R}_\eta^\gamma(s',s'')\bar{r}_\theta^\gamma(s'',s)\mathrm{d}s''\Big]\mathrm{d}s'.$$

Hence $\mathbb{E}|r_\theta(t,s)-\bar{r}^\gamma_\theta(t,s)| \le (I)+(II)+(III)+(IV)$ where

$$(I)=\mathbb{E}\bigg[\mathbf{1}\{\lceil s\rceil \le \lfloor t\rfloor\}\int_{\lceil s\rceil}^{\lfloor t\rfloor}\Big|\Big(-\delta\beta+\partial_\theta s(\theta^{s'},\bar{\alpha}_\gamma^{s'})\Big)r_\theta(s',s)-\Big(-\delta\beta+\partial_\theta s(\bar{\theta}_\gamma^{s'},\bar{\alpha}_\gamma^{s'})\Big)\bar{r}^\gamma_\theta(s',s)\Big|\mathrm{d}s'\bigg],$$

$$(II)=\mathbb{E}\bigg[\mathbf{1}\{\lceil s\rceil \le \lfloor t\rfloor\}\int_{\lceil s\rceil}^{\lfloor t\rfloor}\int_{\lceil s\rceil}^{\lfloor s'\rfloor}\Big|\bar{R}^\gamma_\eta(s',s'')(r_\theta(s'',s)-\bar{r}^\gamma_\theta(s'',s))\Big|\mathrm{d}s''\,\mathrm{d}s'\bigg],$$

$$(III)=\mathbb{E}\bigg[\mathbf{1}\{\lceil s\rceil \le \lfloor t\rfloor\}\int_{\lceil s\rceil}^{\lfloor t\rfloor}\int_{(s,\lceil s\rceil)\cup(\lfloor s'\rfloor,s')}\Big|\bar{R}^\gamma_\eta(s',s'')r_\theta(s'',s)\Big|\mathrm{d}s''\,\mathrm{d}s'\bigg],$$

$$(IV)=\int_{(s,\lceil s\rceil)\cup(\lfloor t\rfloor,t)}\bigg|\Big(-\delta\beta+\partial_\theta s(\theta^{s'},\bar{\alpha}_\gamma^{s'})\Big)r_\theta(s',s)+\int_s^{s'}\bar{R}^\gamma_\eta(s',s'')r_\theta(s'',s)\mathrm{d}s''\bigg|\mathrm{d}s'.$$

By the same arguments as in the proof of Lemma 3.7, using the above bound (100) and Lipschitz continuity of $\partial_\theta s(\cdot)$ in Assumption 2.2, we may show

$$(I)+(II)\le \frac{C}{\lambda}e^{\lambda t}\Big(\sup_{0\le s\le t\le T}e^{-\lambda t}\mathbb{E}|r_\theta(t,s)-\bar{r}^\gamma_\theta(t,s)|\Big)+C\sqrt{\gamma}.$$

Applying $\lceil s\rceil - s\le\gamma$, $s'-\lfloor s'\rfloor\le\gamma$, and $t-\lfloor t\rfloor\le\gamma$, we have

$$(III)+(IV)\le C\gamma.$$

Then

$$\sup_{0\le s\le t\le T}e^{-\lambda t}\mathbb{E}|r_\theta(t,s)-\bar{r}^\gamma_\theta(t,s)|\le\frac{C}{\lambda}\sup_{0\le s\le t\le T}e^{-\lambda t}\mathbb{E}|r_\theta(t,s)-\bar{r}^\gamma_\theta(t,s)|+C\sqrt{\gamma},$$

and choosing large enough $\lambda>0$ and rearranging gives

$$d(R_\theta,\bar{R}^\gamma_\theta)\le\sup_{0\le s\le t\le T}e^{-\lambda t}\mathbb{E}|r_\theta(t,s)-\bar{r}^\gamma_\theta(t,s)|\le C\sqrt{\gamma}.$$

Bound of $d(\tilde{\alpha},\bar{\tilde{\alpha}}_\gamma)$. By definition,

$$\tilde{\alpha}^t=\alpha^0+\int_0^t\mathcal{G}(\tilde{\alpha}^s,\mathsf{P}(\theta^s))\mathrm{d}s,\qquad \bar{\tilde{\alpha}}^t_\gamma=\alpha^0+\int_0^{\lfloor t\rfloor}\mathcal{G}(\bar{\tilde{\alpha}}^s_\gamma,\mathsf{P}(\bar{\theta}^s_\gamma))\mathrm{d}s,$$

so $\|\tilde{\alpha}^t-\bar{\tilde{\alpha}}^t_\gamma\|\le(I)+(II)$ where

$$(I)=\int_0^{\lfloor t\rfloor}\Big\|\mathcal{G}(\tilde{\alpha}^s,\mathsf{P}(\theta^s))-\mathcal{G}(\bar{\tilde{\alpha}}^s_\gamma,\mathsf{P}(\bar{\theta}^s_\gamma))\Big\|\mathrm{d}s,\qquad (II)=\int_{\lfloor t\rfloor}^t\Big\|\mathcal{G}(\tilde{\alpha}^s,\mathsf{P}(\theta^s))\Big\|\mathrm{d}s.$$

By the same arguments as in the proof of Lemma 3.7, using the above bound (100) and the Lipschitz continuity of $\mathcal{G}(\cdot)$ in Assumption 2.3, we have

$$(I)\le\frac{C}{\lambda}e^{\lambda t}\sup_{s\in[0,T]}e^{-\lambda s}\|\tilde{\alpha}^s-\bar{\tilde{\alpha}}^s_\gamma\|\,\mathrm{d}s+C\sqrt{\gamma}$$

Using $t-\lfloor t\rfloor\le\gamma$, we have $(II)\le C\gamma$. So choosing $\lambda>0$ large enough and rearranging shows

$$d(\tilde{\alpha},\bar{\tilde{\alpha}}_\gamma)=\sup_{t\in[0,T]}e^{-\lambda t}\|\tilde{\alpha}^t-\bar{\tilde{\alpha}}^t_\gamma\|\le C\sqrt{\gamma}.$$

This concludes the proof of (a). The proof of (b) is analogous, and we omit this for brevity. ☐

Lemma 4.5. *Let $\{\theta^t\}_{t\in[0,T]}$, $\{\eta^t\}_{t\in[0,T]}$, and $\{\alpha^t\}_{t\in[0,T]}$ be the components of the solution to the DMFT system in Theorem 2.4, and let $\{\bar{\theta}^t_\gamma\}_{t\in[0,T]}$, $\{\bar{\eta}^t_\gamma\}_{t\in[0,T]}$, $\{\bar{\alpha}^t_\gamma\}_{t\in[0,T]}$ be defined from the components of the fixed point (99) via (93) and (96). Then for any fixed $m\ge0$ and $t_1,\ldots,t_m\in[0,T]$, as $\gamma\to0$,*

$$\mathsf{P}(\theta^*,\bar{\theta}^{t_1}_\gamma,\ldots,\bar{\theta}^{t_m}_\gamma)\xrightarrow{W_2}\mathsf{P}(\theta^*,\theta^{t_1},\ldots,\theta^{t_m}) \tag{101}$$

$$\mathsf{P}(\bar{w}^*_\gamma,\varepsilon,\bar{\eta}^{t_1}_\gamma,\ldots,\bar{\eta}^{t_m}_\gamma)\xrightarrow{W_2}\mathsf{P}(w^*,\varepsilon,\eta^{t_1},\ldots,\eta^{t_m}) \tag{102}$$

$$\{\bar{\alpha}^t_\gamma\}_{t\in[0,T]}\to\{\alpha^t\}_{t\in[0,T]} \tag{103}$$

where (103) holds in the sense of uniform convergence on $C([0,T],\mathbb{R}^K)$.

Proof. Let $X^\gamma = (\bar{R}^\gamma_\eta, \bar{C}^\gamma_\eta, \bar{\alpha}_\gamma)$ and $Y^\gamma = (\bar{R}^\gamma_\theta, \bar{C}^\gamma_\theta, \bar{\alpha}_\gamma)$ be the components of the fixed point (99), and let $X = (R_\eta, C_\eta, \alpha)$ and $Y = (R_\theta, C_\theta, \alpha)$ be those of the unique solution to the continuous DMFT system prescribed by Theorem 2.4. Let $d(\cdot)$ denote the metrics introduced in (48) and (49), for a sufficiently large choice of $\lambda > 0$. By Lemma 4.3, $X^\gamma \in \mathcal{D}^\gamma_\eta \cap \mathcal{S}_\eta$ for all sufficiently small $\gamma > 0$, so $\mathcal{T}_{\eta\to\eta}(X^\gamma)$ is well-defined. Then, applying the fixed point conditions for X^γ and X,

$$d(X, X^\gamma) = d\big(\mathcal{T}_{\eta\to\eta}(X), \mathcal{T}^\gamma_{\eta\to\eta}(X^\gamma)\big) \leq d\big(\mathcal{T}_{\eta\to\eta}(X), \mathcal{T}_{\eta\to\eta}(X^\gamma)\big) + d\big(\mathcal{T}_{\eta\to\eta}(X^\gamma), \mathcal{T}^\gamma_{\eta\to\eta}(X^\gamma)\big).$$

By Lemmas 3.7 and 3.8, $\mathcal{T}_{\eta\to\eta}$ is a contraction on $\mathcal{S}_\eta$ for large enough $\lambda > 0$, for which the first term satisfies $d\big(\mathcal{T}_{\eta\to\eta}(X), \mathcal{T}_{\eta\to\eta}(X^\gamma)\big) \leq \frac{1}{2} d(X, X^\gamma)$. Thus, rearranging shows

$$d(X, X^\gamma) \leq 2d\big(\mathcal{T}_{\eta\to\eta}(X^\gamma), \mathcal{T}^\gamma_{\eta\to\eta}(X^\gamma)\big) = 2d\big(\mathcal{T}_{\theta\to\eta} \circ \mathcal{T}_{\eta\to\theta}(X^\gamma), \mathcal{T}^\gamma_{\theta\to\eta} \circ \mathcal{T}^\gamma_{\eta\to\theta}(X^\gamma)\big).$$

Letting $Y' = \mathcal{T}_{\eta\to\theta}(X^\gamma) \in \mathcal{S}_\theta$ and $Y^\gamma = \mathcal{T}^\gamma_{\eta\to\theta}(X^\gamma) \in \mathcal{D}^\gamma_\theta \cap \mathcal{S}_\theta$, this shows

$$d(X, X^\gamma) \leq 2d\big(\mathcal{T}_{\theta\to\eta}(Y'), \mathcal{T}^\gamma_{\theta\to\eta}(Y^\gamma)\big) \leq 2d\big(\mathcal{T}_{\theta\to\eta}(Y'), \mathcal{T}_{\theta\to\eta}(Y^\gamma)\big) + 2d\big(\mathcal{T}_{\theta\to\eta}(Y^\gamma), \mathcal{T}^\gamma_{\theta\to\eta}(Y^\gamma)\big).$$

By Lemma 4.4, $d(Y', Y^\gamma) = d(\mathcal{T}_{\eta\to\theta}(X^\gamma), \mathcal{T}^\gamma_{\eta\to\theta}(X^\gamma)) \leq C\sqrt{\gamma}$. Then by Lemma 3.8, the first term is bounded as $d(\mathcal{T}_{\theta\to\eta}(Y'), \mathcal{T}_{\theta\to\eta}(Y^\gamma)) \leq Cd(Y', Y^\gamma) \leq C'\sqrt{\gamma}$. By Lemma 4.4, the second term is also bounded as $d(\mathcal{T}_{\theta\to\eta}(Y^\gamma), \mathcal{T}^\gamma_{\theta\to\eta}(Y^\gamma)) \leq C\sqrt{\gamma}$. So combining these statements and taking $\gamma \to 0$ shows

$$\lim_{\gamma\to 0} d(X, X^\gamma) = 0, \qquad \lim_{\gamma\to 0} d(Y, Y^\gamma) = 0. \tag{104}$$

By definition of the metrics $d(\cdot)$, the convergence (104) implies the uniform convergence statement (103). It also implies

$$\lim_{\gamma\to 0} \sup_{0\leq s\leq t\leq T} |C_\eta(t,s) - \bar{C}^\gamma_\eta(t,s)| = 0.$$

We recall that Theorem 2.4 shows $(R_\eta, C_\eta, \alpha) \in \mathcal{S}^{\text{cont}}_\eta$, for which the continuity property (32) holds for all $0 \leq s \leq t \leq T$. Then there exists a coupling of $\{u^t\}_{t\in[0,T]}$ and $\{\bar{u}^t_\gamma\}_{t\in[0,T]}$ with covariance kernels $C_\eta(t,s)$ and $\bar{C}^\gamma_\eta(t,s)$ for which

$$\lim_{\gamma\to 0} \sup_{t\in[0,T]} \mathbb{E}(u^t - \bar{u}^t_\gamma)^2 = 0,$$

see e.g. [52, Lemma D.1]. Defining $\{\theta^t\}$ and $\{\bar{\theta}^t_\gamma\}$ by this coupling of $\{u^t\}$ and $\{\bar{u}^t_\gamma\}$ and a common Brownian motion, the same arguments as leading to (100) shows

$$\lim_{\gamma\to 0} \sup_{t\in[0,T]} e^{-\lambda t}\mathbb{E}(\theta^t - \bar{\theta}^t_\gamma)^2 = 0,$$

hence in particular $\lim_{\gamma\to 0} \mathbb{E}(\theta^t - \bar{\theta}^t_\gamma)^2 = 0$ for each fixed $t \in [0,T]$ under this coupling, which implies (101). A similar argument shows (102). ☐

4.3 Step 3: Discretization of Langevin dynamics

We now consider a piecewise constant embedding $\{\bar{\boldsymbol{\theta}}^t_\gamma, \bar{\widehat{\alpha}}^t_\gamma\}_{t\in[0,T]}$ of the discretized Langevin process (55–56), defined as

$$\bar{\boldsymbol{\theta}}^t_\gamma = \boldsymbol{\theta}^{[t]}_\gamma, \qquad \bar{\widehat{\alpha}}^t_\gamma = \widehat{\alpha}^{[t]}_\gamma, \qquad \bar{\boldsymbol{\eta}}^t_\gamma = \mathbf{X}\bar{\boldsymbol{\theta}}^t_\gamma$$

where $[t] \in \mathbb{Z}_+$ is as previously defined in (92). A simple induction shows that this is equivalently the solution to a modification of the dynamics (4–5),

$$\begin{aligned} \bar{\boldsymbol{\theta}}^t_\gamma &= \boldsymbol{\theta}^0 + \int_0^{\lfloor t\rfloor} \Big[-\beta\,\mathbf{X}^\top(\mathbf{X}\bar{\boldsymbol{\theta}}^s_\gamma - \mathbf{y}) + \big(s(\bar{\theta}^s_{\gamma,j}, \bar{\widehat{\alpha}}^s_\gamma)\big)_{j=1}^d\Big]\mathrm{d}s + \sqrt{2}\,\mathbf{b}^{\lfloor t\rfloor}, \\ \bar{\widehat{\alpha}}^t_\gamma &= \widehat{\alpha}^0 + \int_0^{\lfloor t\rfloor} \mathcal{G}\Big(\bar{\widehat{\alpha}}^s_\gamma, \frac{1}{d}\sum_{j=1}^d \delta_{\bar{\theta}^s_{\gamma,j}}\Big)\mathrm{d}s. \end{aligned} \tag{105}$$

We compare this to the solution $\{\boldsymbol{\theta}^t, \widehat{\alpha}^t\}_{t\geq 0}$ of the original dynamics (4–5), with $\boldsymbol{\eta}^t = \mathbf{X}\boldsymbol{\theta}^t$, to show the following lemma.

Lemma 4.6. *Let $\{\boldsymbol{\theta}^t, \boldsymbol{\eta}^t, \widehat{\alpha}^t\}$ be defined by (4–5), and let $\{\bar{\boldsymbol{\theta}}^t_\gamma, \bar{\boldsymbol{\eta}}^t_\gamma, \bar{\widehat{\alpha}}^t_\gamma\}$ be defined by the piecewise constant process (105). Then for any fixed $m \geq 1$ and $t_1, \ldots, t_m \in [0, T]$, there exists a function $\iota : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\lim_{\gamma \to 0} \iota(\gamma) = 0$ such that almost surely*

$$\limsup_{n,d\to\infty} W_2\Big(\frac{1}{d}\sum_{j=1}^d \delta_{(\theta_j^*, \theta_j^{t_1}, \ldots, \theta_j^{t_m})}, \frac{1}{d}\sum_{j=1}^d \delta_{(\theta_j^*, \bar{\theta}_{\gamma,j}^{t_1}, \ldots, \bar{\theta}_{\gamma,j}^{t_m})}\Big) < \iota(\gamma)$$

$$\limsup_{n,d\to\infty} W_2\Big(\frac{1}{n}\sum_{i=1}^n \delta_{(\eta_i^*, \varepsilon_i, \eta_i^{t_1}, \ldots, \eta_i^{t_m})}, \frac{1}{n}\sum_{i=1}^n \delta_{(\eta_i^*, \varepsilon_i, \bar{\eta}_{\gamma,i}^{t_1}, \ldots, \bar{\eta}_{\gamma,i}^{t_m})}\Big) < \iota(\gamma)$$

$$\limsup_{n,d\to\infty} \sup_{t\in[0,T]} \big\|\widehat{\alpha}^t - \bar{\widehat{\alpha}}^t_\gamma\big\| < \iota(\gamma).$$

We proceed to prove Lemma 4.6.

Lemma 4.7. *Let $\{\mathbf{b}^t\}_{t\geq 0}$ be a standard Brownian motion on $\mathbb{R}^d$. For any fixed $T > 0$, there exists a constant $C > 0$ depending on T such that almost surely*

$$\limsup_{d\to\infty} \sup_{t\in[0,T]} \frac{1}{\sqrt{d}}\|\mathbf{b}^t\| \leq C, \quad \limsup_{d\to\infty} \sup_{t\in[0,T]} \frac{1}{\sqrt{d}}\big(\|\mathbf{b}^t - \mathbf{b}^{\lfloor t\rfloor}\| + \|\mathbf{b}^t - \mathbf{b}^{\lceil t\rceil}\|\big) \leq C\sqrt{\gamma \max(\log(1/\gamma), 1)}.$$

Proof. We first show that for any $a, b \in \mathbb{R}_+$ with $a \leq b$, we have $\mathbb{P}(\sup_{t\in[a,b]} d^{-1/2}\|\mathbf{b}^t - \mathbf{b}^a\| \geq u) \leq \exp\big(-\frac{cdu^2}{b-a}\big)$ for any $u \geq \sqrt{4(b-a)}$ and some constant $c > 0$. To see this, for any $\lambda \in (0, \frac{d}{2(b-a)})$, we have

$$\begin{aligned}\mathbb{P}(\sup_{t\in[a,b]} d^{-1/2}\|\mathbf{b}^t - \mathbf{b}^a\| \geq u) &= \mathbb{P}(\sup_{t\in[a,b]} \exp(\lambda\|\mathbf{b}^t - \mathbf{b}^a\|^2/d) \geq \exp(\lambda u^2)) \\ &\overset{(*)}{\leq} e^{-\lambda u^2}\mathbb{E}[\exp(\lambda\|\mathbf{b}^b - \mathbf{b}^a\|^2/d)] \overset{(**)}{=} e^{-\lambda u^2}(1 - 2\lambda(b-a)/d)^{-d/2},\end{aligned}$$

where $(*)$ applies Doob's maximal inequality for the nonnegative submartingale $\{\exp(\lambda\|\mathbf{b}^t - \mathbf{b}^a\|^2/d)\}_{t\in[a,b]}$, and $(**)$ applies the moment generating function of the χ^2 distribution. Choosing $\lambda = cd/(b-a)$ for some small enough $c > 0$ and applying $(1-x) \geq e^{-2x}$ for small $x > 0$, we have

$$\mathbb{P}(\sup_{t\in[a,b]} d^{-1/2}\|\mathbf{b}^t - \mathbf{b}^a\| \geq u) \leq \exp\Big(-\frac{cdu^2}{b-a} + 2cd\Big) \leq \exp\Big(-\frac{cdu^2}{2(b-a)}\Big)$$

for $u \geq \sqrt{4(b-a)}$, proving the inequality. For the first claim, we apply this with $a = 0$, $b = T$, $u = \sqrt{4T}$ to yield that $\mathbb{P}(\sup_{t\in[0,T]} d^{-1/2}\|\mathbf{b}^t\| \geq \sqrt{4T}) \leq \exp(-2cd)$, so the claim follows by the Borel Cantelli lemma. For the second claim, let $N = T/\gamma$ (assumed without loss of generality to be an integer greater than 1), and $I_i = [(i-1)\gamma, i\gamma)$ for $i \in [N]$. Then applying the inequality over these intervals yields

$$\mathbb{P}(\sup_{t\in[0,T]} (d\gamma)^{-1/2}\|\mathbf{b}^t - \mathbf{b}^{\lfloor t\rfloor}\| \geq u) \leq N \max_{i\leq N} \mathbb{P}\Big(\sup_{t\in I_i} (d\gamma)^{-1/2}\|\mathbf{b}^t - \mathbf{b}^{i\gamma}\| \geq u\Big) \leq Ne^{-cdu^2}$$

for any $u \geq 2$. Hence by choosing $u = C\sqrt{\log N}$ for large enough $C > 0$, we have $\mathbb{P}(\sup_{t\in[0,T]} (d\gamma)^{-1/2}\|\mathbf{b}^t - \mathbf{b}^{\lfloor t\rfloor}\| \geq C\sqrt{\log N}) \leq \exp(-c'd)$. A similar argument applies to $\sup_{t\in[0,T]} (d\gamma)^{-1/2}\|\mathbf{b}^t - \mathbf{b}^{\lceil t\rceil}\|$, proving the second claim. ◻

Lemma 4.8. *For any fixed $T > 0$, there exists a constant $C > 0$ depending on T such that almost surely*

$$\limsup_{n,d\to\infty} \sup_{t\in[0,T]} (\|\boldsymbol{\theta}^t\|/\sqrt{d} + \|\widehat{\alpha}^t\|) \leq C.$$

Proof. Let $C > 0$ denote a constant depending on T and changing from instance to instance. Since

$$\boldsymbol{\theta}^t = \boldsymbol{\theta}^0 - \int_0^t \Big(\beta\mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}^s - \mathbf{y}) - s(\boldsymbol{\theta}^s, \widehat{\alpha}^s)\Big)\mathrm{d}s + \sqrt{2}\mathbf{b}^t,$$

and $\|s(\boldsymbol{\theta}^s,\widehat{\alpha}^s)\| \le C(\sqrt{d}+\|\boldsymbol{\theta}^s\|+\sqrt{d}\|\widehat{\alpha}^s\|)$ by Assumption 2.2, we have for every $t\in[0,T]$ that

$$\|\boldsymbol{\theta}^t\| \le \|\boldsymbol{\theta}^0\| + C(\|\mathbf{X}\|_{\mathrm{op}}^2+1)\int_0^t \Big(\|\boldsymbol{\theta}^s\|+\sqrt{d}\|\widehat{\alpha}^s\|\Big)\mathrm{d}s + C(\|\mathbf{X}^\top\mathbf{y}\|+\sqrt{d}+\sup_{t\in[0,T]}\|\mathbf{b}^t\|).$$

Next by Assumption 2.3, we have

$$\|\widehat{\alpha}^t\| \le \|\widehat{\alpha}^0\| + C\int_0^t \big(1+\|\boldsymbol{\theta}^s\|/\sqrt{d}+\|\widehat{\alpha}^s\|\big)\mathrm{d}s.$$

Combining the above two bounds yields

$$\frac{\|\boldsymbol{\theta}^t\|}{\sqrt{d}}+\|\widehat{\alpha}^t\| \le C(\|\mathbf{X}\|_{\mathrm{op}}^2+1)\int_0^t \Big(\frac{\|\boldsymbol{\theta}^s\|}{\sqrt{d}}+\|\widehat{\alpha}^s\|\Big)\mathrm{d}s + C\Big(\frac{\|\boldsymbol{\theta}^0\|}{\sqrt{d}}+\|\widehat{\alpha}^0\|+\frac{\|\mathbf{X}^\top\mathbf{y}\|}{\sqrt{d}}+\sup_{t\in[0,T]}\frac{\|\mathbf{b}^t\|}{\sqrt{d}}+1\Big). \tag{106}$$

Hence by Gronwall's inequality, we have

$$\sup_{t\in[0,T]}\frac{\|\boldsymbol{\theta}^t\|}{\sqrt{d}}+\|\widehat{\alpha}^t\| \le C\exp(C(\|\mathbf{X}\|_{\mathrm{op}}^2+1))\Big(\frac{\|\boldsymbol{\theta}^0\|}{\sqrt{d}}+\|\widehat{\alpha}^0\|+\frac{\|\mathbf{X}^\top\mathbf{y}\|}{\sqrt{d}}+\sup_{t\in[0,T]}\frac{\|\mathbf{b}^t\|}{\sqrt{d}}+1\Big).$$

Under Assumption 2.1, we have almost surely that

$$\limsup_{n,d\to\infty}\max\Big(\|\widehat{\alpha}^0\|,\frac{1}{\sqrt{d}}\|\boldsymbol{\theta}^0\|,\frac{1}{\sqrt{d}}\|\mathbf{X}^\top\mathbf{y}\|,\|\mathbf{X}\|_{\mathrm{op}}\Big)\le C, \tag{107}$$

so the conclusion follows from the first claim of Lemma 4.7. □

Proof of Lemma 4.6. Here and throughout, $C>0$ denotes a constant depending on T but not on γ, and changing from instance to instance. We restrict to the almost-sure event where Lemmas 4.7 and 4.8 hold, and (107) holds for all large n,d. Then, coupling (4) and (105) by the same Brownian motion, for any $0\le t\le T$,

$$\begin{aligned}\|\boldsymbol{\theta}^t-\bar{\boldsymbol{\theta}}^t_\gamma\| \le C\int_0^{\lfloor t\rfloor}\Big(\|\mathbf{X}^\top\mathbf{X}(\boldsymbol{\theta}^s-\bar{\boldsymbol{\theta}}^s_\gamma)\|+\|s(\boldsymbol{\theta}^s;\widehat{\alpha}^s)-s(\boldsymbol{\theta}^s_\gamma;\bar{\widehat{\alpha}}^s_\gamma)\|\Big)\mathrm{d}s + C\int_{\lfloor t\rfloor}^t\Big[\mathbf{X}^\top\mathbf{X}\boldsymbol{\theta}^s+s(\boldsymbol{\theta}^s,\widehat{\alpha}^s)\Big]\mathrm{d}s\\ +\sqrt{2}\|\mathbf{b}^t-\mathbf{b}^{\lfloor t\rfloor}\|.\end{aligned}$$

Applying Lipschitz continuity of $s(\cdot)$ in Assumption 2.2 and the bounds of Lemma 4.8 and (107), this shows

$$\|\boldsymbol{\theta}^t-\bar{\boldsymbol{\theta}}^t_\gamma\| \le C\int_0^t\big(\|\boldsymbol{\theta}^s-\bar{\boldsymbol{\theta}}^s_\gamma\|+\sqrt{d}\,\|\widehat{\alpha}^s-\bar{\widehat{\alpha}}^s_\gamma\|\big)\mathrm{d}s + C\gamma\sqrt{d}+C\sup_{t\in[0,T]}\|\mathbf{b}^t-\mathbf{b}^{\lfloor t\rfloor}\|. \tag{108}$$

Similarly, using Assumption 2.3 and $W_2^2(d^{-1}\sum_{j=1}^d\delta_{u_j},d^{-1}\sum_{j=1}^d\delta_{v_j})\le d^{-1}\|\mathbf{u}-\mathbf{v}\|^2$ for $\mathbf{u},\mathbf{v}\in\mathbb{R}^d$,

$$\begin{aligned}\|\widehat{\alpha}^t-\bar{\widehat{\alpha}}^t_\gamma\| &\le \int_0^{\lfloor t\rfloor}\Big\|\mathcal{G}\Big(\widehat{\alpha}^s,\frac{1}{d}\sum_{j=1}^d\delta_{\theta^s_j}\Big)-\mathcal{G}\Big(\bar{\widehat{\alpha}}^s_\gamma,\frac{1}{d}\sum_{j=1}^d\delta_{\bar{\theta}^s_{\gamma,j}}\Big)\Big\|\mathrm{d}s+\int_{\lfloor t\rfloor}^t\Big\|\mathcal{G}\Big(\widehat{\alpha}^s,\frac{1}{d}\sum_{j=1}^d\delta_{\theta^s_j}\Big)\Big\|\mathrm{d}s\\ &\le C\int_0^t\Big(\|\widehat{\alpha}^s-\bar{\widehat{\alpha}}^s_\gamma\|+\frac{1}{\sqrt{d}}\|\boldsymbol{\theta}^s-\bar{\boldsymbol{\theta}}^s_\gamma\|\Big)\mathrm{d}s+C\gamma.\end{aligned}$$

Combining the above display with (108), we have by Gronwall's lemma

$$\sup_{t\in[0,T]}\|\widehat{\alpha}^t-\bar{\widehat{\alpha}}^t_\gamma\|+\frac{1}{\sqrt{d}}\|\boldsymbol{\theta}^t-\bar{\boldsymbol{\theta}}^t_\gamma\|\le C\gamma+\frac{C}{\sqrt{d}}\sup_{t\in[0,T]}\|\mathbf{b}^t-\mathbf{b}^{\lfloor t\rfloor}\|. \tag{109}$$

By Lemma 4.7, there exists some $C > 0$ such that

$$\limsup_{d\to\infty} \sup_{t\in[0,T]} \frac{1}{\sqrt{d}}\|\mathbf{b}^t - \mathbf{b}^{\lfloor t\rfloor}\| \le C\sqrt{\gamma \cdot \max(\log(1/\gamma), 1)}.$$

Substituting this bound in (109) proves the claim on α. Noting that

$$W_2\Big(\frac{1}{d}\sum_{j=1}^d \delta_{(\theta_j^*,\theta_j^{t_1},\ldots,\theta_j^{t_m})}, \frac{1}{d}\sum_{j=1}^d \delta_{(\theta_j^*,\bar\theta_{\gamma,j}^{t_1},\ldots,\bar\theta_{\gamma,j}^{t_m})}\Big) \le \sqrt{\frac{1}{d}\sum_{\ell=1}^m\sum_{j=1}^d \big(\theta_j^{t_\ell} - \bar\theta_{\gamma,j}^{t_\ell}\big)^2} \le \sqrt{m}\cdot\frac{1}{\sqrt{d}}\sup_{t\in[0,T]}\|\boldsymbol{\theta}^t - \bar{\boldsymbol{\theta}}_\gamma^t\|,$$

this proves also the claim on θ, and the claim on η follows from $\|\boldsymbol{\eta}^t - \bar{\boldsymbol{\eta}}_\gamma^t\| \le \|\mathbf{X}\|_{\mathrm{op}}\|\boldsymbol{\theta}^t - \bar{\boldsymbol{\theta}}_\gamma^t\|$ and the same argument. □

4.4 Completing the proof

Proof of Theorem 2.5. For part (a), by the triangle inequality,

$$\sup_{t\in[0,T]}\|\widehat\alpha^t - \alpha^t\| \le \sup_{t\in[0,T]}\|\widehat\alpha^t - \bar{\widehat\alpha}_\gamma^t\| + \sup_{t\in[0,T]}\|\bar{\widehat\alpha}_\gamma^t - \bar\alpha_\gamma^t\| + \sup_{t\in[0,T]}\|\bar\alpha_\gamma^t - \alpha^t\|.$$

Since $\{\bar{\widehat\alpha}_\gamma^t\}$ and $\{\bar\alpha_\gamma^t\}$ are piecewise constant with values equal to those of the discrete processes of Section 4.1, Lemma 4.1 implies that the middle term converges to 0 a.s. as $n, d \to \infty$. Then, taking $n, d \to \infty$ followed by $\gamma \to 0$ and applying also Lemmas 4.5 and 4.6 to bound the first and third terms in this limit, this shows (a).

For part (b), similarly combining Lemmas 4.1, 4.5, and 4.6 shows that almost surely, for any $m \ge 1$ and $t_0, t_1, \ldots, t_m \in [0, T]$,

$$\frac{1}{d}\sum_{j=1}^d \delta_{\left(\theta_j^*,\theta_j^{t_0},\ldots,\theta_j^{t_m}\right)} \overset{W_2}{\to} \mathsf{P}(\theta^*, \theta^{t_0}, \ldots, \theta^{t_m}). \tag{110}$$

We now strengthen this to almost-sure convergence in the Wasserstein-2 sense over $\mathbb{R}\times C([0,T])$, equipped with the product norm

$$\|\theta\|_\infty := |\theta^*| + \sup_{t\in[0,T]}|\theta^t|.$$

By [54, Definition 6.8 and Theorem 6.9], it suffices to show weak convergence together with convergence of the squared norm $\|\theta\|_\infty^2$, which will be implied by convergence for all pseudo-Lipschitz test functions $f : \mathbb{R}\times C([0,T]) \to \mathbb{R}$ satisfying

$$\big|f(\theta) - f(\theta')\big| \le C\|\theta - \theta'\|_\infty(1 + \|\theta\|_\infty + \|\theta'\|_\infty). \tag{111}$$

Consider the event $\mathcal{E}$ where (110) holds for each $m \ge 2$ and $\{t_0, t_1, t_2, \ldots, t_m\} = \{0, \gamma, 2\gamma, \ldots, T\}$ with $\gamma = T/m$. Let $\{\tilde{\boldsymbol{\theta}}^t\}_{t\in[0,T]}$ be a piecewise linear interpolation of $\{\boldsymbol{\theta}^t\}_{t\in[0,T]\cap\gamma\mathbb{Z}_+}$, and similarly let $\{\tilde\theta^t\}_{t\in[0,T]}$ be a piecewise linear interpolation of the DMFT process $\{\theta^t\}_{t\in[0,T]}$. For any pseudo-Lipschitz function $f : \mathbb{R}\times C([0,T]) \to \mathbb{R}$, we have

$$\Big|\frac{1}{d}\sum_{j=1}^d f(\theta_j^*, \{\theta_j^t\}_{t\in[0,T]}) - \mathbb{E}[f(\theta^*, \{\theta^t\}_{t\in[0,T]})]\Big| \le (I) + (II) + (III) \tag{112}$$

with

$$(I) = \Big|\frac{1}{d}\sum_{j=1}^d f(\theta_j^*, \{\tilde\theta_j^t\}_{t\in[0,T]}) - \mathbb{E}[f(\theta^*, \{\tilde\theta^t\}_{t\in[0,T]})]\Big|$$

$$(II) = \Big|\frac{1}{d}\sum_{j=1}^d f(\theta_j^*, \{\theta_j^t\}_{t\in[0,T]}) - f(\theta_j^*, \{\tilde\theta_j^t\}_{t\in[0,T]})\Big|$$

$$(III) = \Big|\mathbb{E}[f(\theta^*, \{\theta^t\}_{t\in[0,T]})] - \mathbb{E}[f(\theta^*, \{\tilde\theta^t\}_{t\in[0,T]})]\Big|.$$

For a piecewise linear process $\tilde{\theta}$ with knots at $\gamma\mathbb{Z}_+$, $f(\theta_j^*, \{\tilde{\theta}^t\}_{t\in[0,T]})$ may be understood as a function of $(\theta_j^*, \tilde{\theta}^0, \tilde{\theta}^\gamma, \tilde{\theta}^{2\gamma}, \ldots, \tilde{\theta}^T)$, where this function is pseudo-Lipschitz on $\mathbb{R}^{m+2}$ by the pseudo-Lipschitz property (111) for f. Then on the above event $\mathcal{E}$, the Wasserstein-2 convergence (110) implies

$$\lim_{n,d\to\infty} (I) = 0.$$

To bound (II), let $C, C' > 0$ be constants depending on T (but not γ) and changing from instance to instance. Writing $\theta_j = (\theta_j^*, \{\theta_j^t\}_{t\in[0,T]}) \in \mathbb{R} \times C([0,T], \mathbb{R})$ and applying (111), we have

$$(II) \leq \frac{C}{d}\sum_{j=1}^d \|\theta_j - \tilde{\theta}_j\|_\infty \big(1 + \|\theta_j\|_\infty\big) \leq C'\Big(\frac{1}{d}\sum_{j=1}^d \|\theta_j - \tilde{\theta}_j\|_\infty^2\Big)^{1/2}\Big(1 + \frac{1}{d}\sum_{j=1}^d \|\theta_j\|_\infty^2\Big)^{1/2}. \tag{113}$$

Set

$$F(\boldsymbol{\theta}, \widehat{\alpha}) = -\beta \mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta} - \mathbf{y}) + s(\boldsymbol{\theta}, \widehat{\alpha})$$

so that by definition, $\theta_j^t = \theta_j^0 + \int_0^t \mathbf{e}_j^\top F(\boldsymbol{\theta}^s, \widehat{\alpha}^s)\mathrm{d}s + \sqrt{2}\, b_j^t$. Hence

$$\sup_{t\in[0,T]} (\theta_j^t)^2 \leq C\Big((\theta_j^0)^2 + \int_0^T (\mathbf{e}_j^\top F(\boldsymbol{\theta}^s, \widehat{\alpha}^s))^2 \mathrm{d}s + \|b_j\|_\infty^2\Big),$$

so

$$\frac{1}{d}\sum_{j=1}^d \|\theta_j\|_\infty^2 \leq C\Big(\frac{1}{d}\sum_{j=1}^d (\theta_j^*)^2 + (\theta_j^0)^2 + \frac{1}{d}\sup_{t\in[0,T]} \|F(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\|^2 + \frac{1}{d}\sum_{j=1}^d \|b_j\|_\infty^2\Big).$$

On an almost-sure event $\mathcal{E}'$, for all large n, d, we have that $d^{-1}\sum_j (\theta_j^*)^2 + d^{-1}\sum_j (\theta_j^0)^2 \leq C$ by Assumption 2.2, that $\sup_{t\in[0,T]} d^{-1}\|F(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\|^2 \leq C$ by the definition of $F(\cdot)$ together with Assumption 2.2 and Lemma 4.8, and that $d^{-1}\sum_j \|b_j\|_\infty^2 \leq C$ by Doob's maximal inequality $\mathbb{P}[\|b_j\|_\infty > x] \leq 2e^{-x^2/(2T)}$ and Bernstein's inequality for a sum of independent subexponential random variables [61, Theorem 2.8.1]. Thus, on $\mathcal{E}'$,

$$\frac{1}{d}\sum_{j=1}^d \|\theta_j\|_\infty^2 \leq C'. \tag{114}$$

Now fixing any $\alpha \in (0, 1/2)$, define the Hölder semi-norm $\|\theta_j\|_\alpha = \sup_{s,t\in[0,T]} |\theta_j^t - \theta_j^s|/|t-s|^\alpha$. Then, since $\tilde{\theta}_j$ linearly interpolates θ_j at the knots $\gamma\mathbb{Z}_+$,

$$\|\theta_j - \tilde{\theta}_j\|_\infty \leq \gamma^\alpha \|\theta_j\|_\alpha.$$

We have by definition $\theta_j^t - \theta_j^s = \int_s^t \mathbf{e}_j^\top F(\boldsymbol{\theta}^r, \widehat{\alpha}^r)\mathrm{d}r + \sqrt{2}(b_j^t - b_j^s)$, so by Hölder's inequality,

$$\begin{aligned}
|\theta_j^t - \theta_j^s| &\leq |t-s|^\alpha \Big(\int_s^t \big|\mathbf{e}_j^\top F(\boldsymbol{\theta}^r, \widehat{\alpha}^r)\big|^{\frac{1}{1-\alpha}} \mathrm{d}r\Big)^{1-\alpha} + \sqrt{2}|b_j^t - b_j^s| \\
&\leq C|t-s|^\alpha \Big(\Big(\int_0^T (\mathbf{e}_j^\top F(\boldsymbol{\theta}^r, \widehat{\alpha}^r))^2\Big)^{1/2} + \|b_j\|_\alpha\Big)
\end{aligned}$$

and hence

$$\frac{1}{d}\sum_{j=1}^d \|\theta_j\|_\alpha^2 \leq C\Big(\frac{1}{d}\sup_{t\in[0,T]} \|F(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\|^2 + \frac{1}{d}\sum_{j=1}^d \|b_j\|_\alpha^2\Big).$$

On an almost-sure event $\mathcal{E}''$, for all large n, d, we have $\sup_{t\in[0,T]} d^{-1}\|F(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\|^2 \leq C$ as above, and $d^{-1}\sum_j \|b_j\|_\alpha^2 \leq C$ by the tail bound $\mathbb{P}[\|b_j\|_\alpha > C + x] \leq e^{-cx^2}$ for some $C, c > 0$ (see e.g. [62, Theorem 5.32, Example 5.37]) and Bernstein's inequality. Thus on $\mathcal{E}''$,

$$\frac{1}{d}\sum_{j=1}^d \|\theta_j - \tilde{\theta}_j\|_\infty^2 \leq C\gamma^{2\alpha}. \tag{115}$$

Applying (114) and (115) to (113), on $\mathcal{E}' \cap \mathcal{E}''$,

$$\limsup_{n,d\to\infty}(II) < C\gamma^\alpha.$$

To bound (III), similarly we have

$$(III) \le C\Big(\mathbb{E}\|\theta - \tilde{\theta}\|_\infty^2\Big)^{1/2}\Big(1 + \mathbb{E}\|\theta\|_\infty^2\Big)^{1/2}. \tag{116}$$

By definition

$$\theta^t = \theta^0 + \int_0^t \Big[-\delta\beta(\theta^s - \theta^*) + s(\theta^s, \alpha^s) + \int_0^s R_\eta(s,s')(\theta^{s'} - \theta^*)\mathrm{d}s' + u^s\Big]\mathrm{d}s + \sqrt{2}b^t.$$

Hence, applying $|s(\theta^s, \alpha^s)| \le C(1+|\theta^s|+\|\alpha^s\|)$ by Assumption 2.2 and uniform boundedness of the continuous functions α^s and $R_\eta(s,s')$ over $[0,T]$,

$$(\theta^t)^2 \le C\Big(1 + (\theta^0)^2 + (\theta^*)^2 + \|u\|_\infty^2 + \|b\|_\infty^2 + \int_0^t \Big(\sup_{r\in[0,s]}(\theta^s)^2\Big)\mathrm{d}s\Big).$$

Then Gronwall's lemma gives

$$\sup_{t\in[0,T]}(\theta^t)^2 \le C\Big(1 + (\theta^0)^2 + (\theta^*)^2 + \|u\|_\infty^2 + \|b\|_\infty^2\Big).$$

We have $\mathbb{E}(\theta^0)^2, \mathbb{E}(\theta^*)^2 \le C$ by assumption. Since $\{u^t\}_{t\in[0,T]}$ has covariance $C_\eta(t,s)$ satisfying $|C_\eta(t,s)| \le C|t-s|$ by the condition (32) defining $\mathcal{S}(T)^{\mathrm{cont}}$, we have $\mathbb{P}[\|u\|_\infty \ge C+t] \le e^{-ct^2}$ for some constants $C, c > 0$ by a standard application of Dudley's inequality [61, Theorem 8.1.6], so $\mathbb{E}\|u\|_\infty^2 \le C$. Similarly $\mathbb{E}\|b\|_\infty^2 \le C$, so this gives

$$\mathbb{E}\|\theta\|_\infty^2 \le C. \tag{117}$$

By definition we have also

$$\theta^t - \theta^s = \int_s^t \Big[-\delta\beta(\theta^r - \theta^*) + s(\theta^r, \alpha^r) + \int_0^r R_\eta(r,r')(\theta^{r'} - \theta^*)\mathrm{d}r' + u^r\Big]\mathrm{d}r + \sqrt{2}(b^t - b^s),$$

so

$$\begin{aligned}|\theta^t - \theta^s| &\le C|t-s|^\alpha\Big(\int_s^t \Big(1 + |\theta^*| + \sup_{r'\in[0,r]}|\theta^{r'}| + |u^r|\Big)^{\frac{1}{1-\alpha}}\mathrm{d}r\Big)^{1-\alpha} + \sqrt{2}|b^t - b^s| \\ &\le C'|t-s|^\alpha\Big(1 + \|\theta\|_\infty + \|u\|_\infty + \|b\|_\alpha\Big).\end{aligned}$$

Then

$$\mathbb{E}\|\theta\|_\alpha^2 \le C(1 + \mathbb{E}\|\theta\|_\infty^2 + \mathbb{E}\|u\|_\infty^2 + \mathbb{E}\|b\|_\alpha^2) \le C',$$

so

$$\mathbb{E}\|\theta - \tilde{\theta}\|_\infty^2 \le \gamma^{2\alpha}\mathbb{E}\|\theta\|_\alpha^2 \le C'\gamma^{2\alpha}. \tag{118}$$

Applying (117) and (118) to (116) shows

$$(III) \le C\gamma^\alpha.$$

Applying these bounds for (I), (II), and (III) to take the limit $n, d \to \infty$ followed by $\gamma \to 0$ in (112), this shows that on the almost-sure event $\mathcal{E} \cap \mathcal{E}' \cap \mathcal{E}''$ (which does not depend on f), for every pseudo-Lipschitz function $f : \mathbb{R} \times C([0,T]) \to \mathbb{R}$,

$$\lim_{n,d\to\infty}\frac{1}{d}\sum_{j=1}^d f(\theta_j^*, \{\theta_j^t\}_{t\in[0,T]}) = \mathbb{E}[f(\theta^*, \{\theta^t\}_{t\in[0,T]})].$$

This implies on $\mathcal{E} \cap \mathcal{E}' \cap \mathcal{E}''$ that

$$\frac{1}{d}\sum_{j=1}^{d} \delta_{\theta_j^*,\{\theta_j^t\}_{t\in[0,T]}} \overset{W_2}{\to} \mathsf{P}(\theta^*, \{\theta^t\}_{t\in[0,T]}).$$

For the convergence for $\boldsymbol{\eta}$, we note that $\eta_i^t = \mathbf{e}_i^\top \mathbf{X}\boldsymbol{\theta}^t = \mathbf{e}_i^\top \mathbf{X}\boldsymbol{\theta}^t + \int_0^t \mathbf{e}_i^\top \mathbf{X} F(\boldsymbol{\theta}^s, \widehat{\alpha}^s)\mathrm{d}s + \sqrt{2}\,\mathbf{e}_i^\top \mathbf{X}\mathbf{b}^t$. Then applying similar arguments as above, fixing any $\alpha \in (0, 1/2)$, on an almost-sure event we have for all large n, d that

$$\frac{1}{n}\sum_{i=1}^{n} \|\eta_i\|_\infty^2 \le C, \qquad \frac{1}{n}\sum_{i=1}^{n} \|\eta_i\|_\alpha^2 \le C\gamma^{2\alpha}.$$

For the DMFT process we have $\eta^t = -\beta \int_0^t R_\theta(t,s)(\eta^s + w^* - \varepsilon)\mathrm{d}s - w^t$, hence

$$(\eta^t)^2 \le C\Big((w^*)^2 + (w^t)^2 + \varepsilon^2 + \int_0^t (\eta^s)^2 \mathrm{d}s\Big)$$

so Gronwall's lemma and a similar argument as above gives $\mathbb{E}\|\eta\|_\infty^2 \le C$. Also

$$|\eta^t - \eta^s| \le C|t-s|^\alpha \Big(\int_s^t \Big(|w^*| + |\varepsilon| + |\eta^r|\Big)^{\frac{1}{1-\alpha}} \mathrm{d}r\Big)^{1-\alpha} + |w^t - w^s| \le C'|t-s|^\alpha \Big(|w^*| + |\varepsilon| + \|\eta\|_\infty + \|w\|_\alpha\Big)$$

so

$$\mathbb{E}\|\eta\|_\alpha^2 \le C\Big(\mathbb{E}(w^*)^2 + \mathbb{E}\varepsilon^2 + \mathbb{E}\|\eta\|_\infty^2 + \mathbb{E}\|w\|_\alpha^2\Big).$$

We recall that $\{w^t\}_{t\in[0,T]}$ has covariance satisfying $|C_\theta(t,s)| \le C|t-s|$, so $\mathbb{P}[\|w\|_\alpha > C + x] \le 2e^{-cx^2}$ for some $C, c > 0$ (c.f. [62, Theorem 5.32]). Thus $\mathbb{E}\|\eta\|_\alpha^2 \le C$. Applying these bounds, the same arguments as above show the almost-sure convergence

$$\frac{1}{n}\sum_{i=1}^{n} \delta_{\eta_i^*, \varepsilon_i, \{\eta_i^t\}_{t\in[0,T]})} \overset{W_2}{\to} \mathsf{P}(\eta^*, \varepsilon, \{\eta^t\}_{t\in[0,T]})$$

where we recall that η^* on the right side is, by definition, $\eta^* = -w^*$. □

5 Convergence of the linear response

In this section, we prove Theorem 2.8. We assume throughout Assumptions 2.1, 2.2, 2.3 and the Hölder-continuity conditions of Theorem 2.8. We first state and prove in Section 5.1 an analogue of Theorem 2.8 for the discrete-time dynamics introduced previously in Section 4.1, and then analyze the discretization error and complete the proof of Theorem 2.8 in Section 5.2.

5.1 Convergence of response functions for discrete dynamics

We recall the discrete, integer-indexed dynamics (55–56), which we reproduce here as

$$\begin{aligned}
\boldsymbol{\theta}_\gamma^{t+1} &= \boldsymbol{\theta}_\gamma^t - \gamma\Big[\beta \mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}_\gamma^t - \mathbf{y}) - s(\boldsymbol{\theta}_\gamma^t, \widehat{\alpha}_\gamma^t)\Big] + \sqrt{2}(\mathbf{b}_\gamma^{t+1} - \mathbf{b}_\gamma^t), \quad \boldsymbol{\eta}_\gamma^t = \mathbf{X}\boldsymbol{\theta}_\gamma^t \\
\widehat{\alpha}_\gamma^{t+1} &= \widehat{\alpha}_\gamma^t + \gamma \cdot \mathcal{G}\Big(\widehat{\alpha}_\gamma^t, \widehat{\mathsf{P}}(\boldsymbol{\theta}_\gamma^t)\Big), \quad \widehat{\mathsf{P}}(\boldsymbol{\theta}) = \frac{1}{d}\sum_{j=1}^{d} \delta_{\theta_j}.
\end{aligned} \tag{119}$$

We first show an analogue of Theorem 2.8 for these discrete dynamics.

For any $s \in \mathbb{Z}_+$ and any $j \in [d]$ or $i \in [n]$, letting $\mathbf{e}_j$ denote the j^{th} standard basis vector in either $\mathbb{R}^d$ or $\mathbb{R}^n$, define two sets of perturbed dynamics

$$\begin{aligned}
\boldsymbol{\theta}_\gamma^{t+1,(s,j),\varepsilon} &= \boldsymbol{\theta}_\gamma^{t,(s,j),\varepsilon} - \gamma\Big[\beta \mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}_\gamma^{t,(s,j),\varepsilon} - \mathbf{y}) - s(\boldsymbol{\theta}_\gamma^{t,(s,j),\varepsilon}, \widehat{\alpha}_\gamma^{t,(s,j),\varepsilon}) - \varepsilon \mathbf{e}_j \mathbf{1}_{t=s}\Big] + \sqrt{2}(\mathbf{b}_\gamma^{t+1} - \mathbf{b}_\gamma^t) \\
\widehat{\alpha}_\gamma^{t+1,(s,j),\varepsilon} &= \widehat{\alpha}_\gamma^{t,(s,j),\varepsilon} + \gamma \cdot \mathcal{G}\Big(\widehat{\alpha}_\gamma^{t,(s,j),\varepsilon}, \widehat{\mathsf{P}}(\boldsymbol{\theta}_\gamma^{t,(s,j),\varepsilon})\Big),
\end{aligned} \tag{120}$$

and

$$\boldsymbol{\theta}_\gamma^{t+1,[s,i],\varepsilon} = \boldsymbol{\theta}_\gamma^{t,[s,i],\varepsilon} - \gamma\Big[\beta\mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}_\gamma^{t,[s,i],\varepsilon} - \mathbf{y}) - s(\boldsymbol{\theta}_\gamma^{t,[s,i],\varepsilon}, \widehat{\alpha}_\gamma^{t,[s,i],\varepsilon}) - \varepsilon\mathbf{X}^\top\mathbf{e}_i\mathbf{1}_{t=s}\Big] + \sqrt{2}(\mathbf{b}_\gamma^{t+1} - \mathbf{b}_\gamma^t)$$
$$\widehat{\alpha}_\gamma^{t+1,[s,i],\varepsilon} = \widehat{\alpha}_\gamma^{t,[s,i],\varepsilon} + \gamma\cdot\mathcal{G}\Big(\widehat{\alpha}_\gamma^{t,[s,i],\varepsilon}, \widehat{\mathsf{P}}(\boldsymbol{\theta}_\gamma^{t,[s,i],\varepsilon})\Big) \tag{121}$$

with the same initial conditions as (119). We set

$$\boldsymbol{\eta}_\gamma^{t,[s,i],\varepsilon} = \mathbf{X}\boldsymbol{\theta}_\gamma^{t,[s,i],\varepsilon}. \tag{122}$$

Comparing with (119), these dynamics have a perturbation to the drift in the direction of $\mathbf{e}_j$ or $\mathbf{X}^\top\mathbf{e}_i$ at the single time $s \in \mathbb{Z}_+$. Let $\mathbf{R}_\theta^\gamma(t,s) = (\mathbf{R}_\theta^\gamma(t,s))_{i,j=1}^d \in \mathbb{R}^{d\times d}$ and $\mathbf{R}_\eta^\gamma(t,s) = (\mathbf{R}_\eta^\gamma(t,s))_{i,j=1}^n \in \mathbb{R}^{n\times n}$ be matrices of response functions defined by

$$(\mathbf{R}_\theta^\gamma(t,s))_{i,j} = \partial_\varepsilon|_{\varepsilon=0}\langle\theta_{\gamma,i}^{t,(s,j),\varepsilon}\rangle, \qquad (\mathbf{R}_\eta^\gamma(t,s))_{i,j} = \delta\beta^2\cdot\partial_\varepsilon|_{\varepsilon=0}\langle\eta_{\gamma,i}^{t,[s,j],\varepsilon}\rangle,$$

where $\langle\cdot\rangle$ denotes the expectation over only the randomness of $\{\mathbf{b}_\gamma^t\}_{t\in\mathbb{Z}_+}$, i.e. conditional on $(\mathbf{X}, \boldsymbol{\theta}^*, \boldsymbol{\varepsilon})$ and on the initial conditions $(\boldsymbol{\theta}_\gamma^{0,(s,j),\varepsilon}, \widehat{\alpha}_\gamma^{0,(s,j),\varepsilon}) = (\boldsymbol{\theta}_\gamma^{0,[s,i],\varepsilon}, \widehat{\alpha}_\gamma^{0,[s,i],\varepsilon}) = (\boldsymbol{\theta}^0, \widehat{\alpha}^0) \in \mathbb{R}^{d+K}$.

Recall also the discrete-time DMFT response functions $R_\theta^\gamma(t,s), R_\eta^\gamma(t,s)$ defined by (60) and (64). The goal of this section is to prove the following analogue of the convergence statements for the response functions in Theorem 2.8.

Lemma 5.1. *For any fixed $s,t \in \mathbb{Z}_+$ with $s < t$, almost surely*

$$\lim_{n,d\to\infty}\frac{1}{d}\operatorname{Tr}\mathbf{R}_\theta^\gamma(t,s) = R_\theta^\gamma(t,s), \qquad \lim_{n,d\to\infty}\frac{1}{n}\operatorname{Tr}\mathbf{R}_\eta^\gamma(t,s) = R_\eta^\gamma(t,s).$$

To ease notation, in the remainder of this section we will drop all subscripts γ and write simply $\boldsymbol{\theta}^t = \boldsymbol{\theta}_\gamma^t$, $\widehat{\alpha}^t = \widehat{\alpha}_\gamma^t$, $\mathbf{b}^t = \mathbf{b}_\gamma^t$ etc. to refer to the above discrete-time processes. We first establish in Section 5.1.1 a set of dynamical cavity estimates, which we will then use to prove Lemma 5.1 in Section 5.1.2.

5.1.1 Dynamical cavity estimates

We introduce the following notations: For any $j \in [d]$ and $i \in [n]$, denote

$$\boldsymbol{\theta}^t = (\theta_j^t, \boldsymbol{\theta}_{-j}^t) \in \mathbb{R}^d, \qquad \theta_j^t \in \mathbb{R}, \qquad \boldsymbol{\theta}_{-j}^t \in \mathbb{R}^{d-1},$$
$$\boldsymbol{\eta}^t = (\eta_i^t, \boldsymbol{\eta}_{-i}^t) \in \mathbb{R}^n, \qquad \eta_i^t \in \mathbb{R}, \qquad \boldsymbol{\eta}_{-i}^t \in \mathbb{R}^{n-1},$$

where $\{\boldsymbol{\theta}^t, \boldsymbol{\eta}^t\}_{t\in\mathbb{Z}_+}$ are the components of the discrete-time process (119), and $\boldsymbol{\theta}_{-j}^t$ are the coordinates of $\boldsymbol{\theta}^t$ excluding the j^{th} (and similarly for $\boldsymbol{\eta}^t$).

We consider the following leave-one-out versions of (119): For $j \in [d]$, let

$$\mathbf{X}^{(j)} = (X_{ik}\mathbf{1}_{k\neq j})_{i,k} \in \mathbb{R}^{n\times d}, \qquad \mathbf{y}^{(j)} = \mathbf{X}^{(j)}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon} \in \mathbb{R}^n \tag{123}$$

where $\mathbf{X}^{(j)}$ denotes $\mathbf{X}$ with j^{th} column set to 0. Define

$$\boldsymbol{\theta}^{t+1,(j)} = \boldsymbol{\theta}^{t,(j)} - \gamma\Big[\beta(\mathbf{X}^{(j)})^\top(\mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)} - \mathbf{y}^{(j)}) - s(\boldsymbol{\theta}^{t,(j)}, \widehat{\alpha}^{t,(j)})\Big] + \sqrt{2}(\mathbf{b}^{t+1} - \mathbf{b}^t) \in \mathbb{R}^d$$
$$\widehat{\alpha}^{t+1,(j)} = \widehat{\alpha}^{t,(j)} + \gamma\cdot\mathcal{G}\Big(\widehat{\alpha}^{t,(j)}, \widehat{\mathsf{P}}(\boldsymbol{\theta}^{t,(j)})\Big) \tag{124}$$

with initialization $(\boldsymbol{\theta}^{0,(j)}, \widehat{\alpha}^{0,(j)}) = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$, and write as above

$$\boldsymbol{\theta}^{t,(j)} = (\theta_j^{t,(j)}, \boldsymbol{\theta}_{-j}^{t,(j)}) \in \mathbb{R}^d, \qquad \theta_j^{t,(j)} \in \mathbb{R}, \qquad \boldsymbol{\theta}_{-j}^{t,(j)} \in \mathbb{R}^{d-1}.$$

We note that for convenience of the proof, we define $\boldsymbol{\theta}^{t,(j)}$ to be of the same dimension as $\boldsymbol{\theta}$, where one may check from (124) that the dynamics of $\boldsymbol{\theta}_{-j}^{t,(j)}$ do not involve $\theta_j^{t,(j)}$. Similarly, for $i \in [n]$, let

$$\mathbf{X}^{[i]} = (X_{kj}\mathbf{1}_{k\neq i})_{k,j} \in \mathbb{R}^{n\times d}, \qquad \mathbf{y}^{[i]} = \mathbf{X}^{[i]}\boldsymbol{\theta}^* + \boldsymbol{\varepsilon},$$

where $\mathbf{X}^{[i]}$ sets the i^{th} row of $\mathbf{X}$ to 0. Define

$$\boldsymbol{\theta}^{t+1,[i]} = \boldsymbol{\theta}^{t,[i]} - \gamma\Big[\beta(\mathbf{X}^{[i]})^\top(\mathbf{X}^{[i]}\boldsymbol{\theta}^{t,[i]} - \mathbf{y}^{[i]}) - s(\boldsymbol{\theta}^{t,[i]}, \widehat{\alpha}^{t,[i]})\Big] + \sqrt{2}(\mathbf{b}^{t+1} - \mathbf{b}^t) \in \mathbb{R}^d$$
$$\widehat{\alpha}^{t+1,[i]} = \widehat{\alpha}^{t,[i]} + \gamma \cdot \mathcal{G}\Big(\widehat{\alpha}^{t,[i]}, \widehat{\mathsf{P}}(\boldsymbol{\theta}^{t,[i]})\Big) \tag{125}$$

also with initialization $(\boldsymbol{\theta}^{0,[i]}, \widehat{\alpha}^{0,[i]}) = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$, and write as above

$$\boldsymbol{\eta}^{t,[i]} = (\eta_i^{t,[i]}, \boldsymbol{\eta}_{-i}^{t,[i]}) \in \mathbb{R}^n, \qquad \eta_i^{t,[i]} \in \mathbb{R}, \qquad \boldsymbol{\eta}_{-i}^{t,[i]} \in \mathbb{R}^{n-1}.$$

By construction, $\{\boldsymbol{\theta}^{t,(j)}, \widehat{\alpha}^{t,(j)}\}$ is independent of the j^{th} column of $\mathbf{X}$, and $\{\boldsymbol{\theta}^{t,[i]}, \widehat{\alpha}^{t,[i]}\}$ is independent of the i^{th} row of $\mathbf{X}$.

The following lemma gives ℓ_2 estimates on the original dynamics (119) as well as on its difference with the cavity versions (124) and (125).

Lemma 5.2. *Fix any $T > 0$. Then there exists a constant $C > 0$ (depending on T but not γ) such that for any $\gamma > 0$, almost surely for all large n, d, we have for all $0 \le t \le T/\gamma$ and all $j \in [d]$, $i \in [n]$ that*

$$\frac{\|\boldsymbol{\theta}^t\|}{\sqrt{d}} + \|\widehat{\alpha}^t\| \le C, \quad \frac{\|\boldsymbol{\theta}^{t,(j)}\|}{\sqrt{d}} + \|\widehat{\alpha}^{t,(j)}\| \le C, \quad \frac{\|\boldsymbol{\theta}^{t,[i]}\|}{\sqrt{d}} + \|\widehat{\alpha}^{t,[i]}\| \le C, \tag{126}$$
$$|\theta_j^{t,(j)}| \le C(1 + |\theta_j^0| + \max_{t\in[0,T/\gamma]} |b_j^t|), \tag{127}$$
$$\|\boldsymbol{\theta}^{t,(j)} - \boldsymbol{\theta}^t\| + \sqrt{d}\,\|\widehat{\alpha}^{t,(j)} - \widehat{\alpha}^t\| \le C(|\theta_j^0| + |\theta_j^*| + \max_{t\in[0,T/\gamma]} |b_j^t| + \sqrt{\log d}), \tag{128}$$
$$\|\boldsymbol{\theta}^{t,[i]} - \boldsymbol{\theta}^t\| + \sqrt{d}\,\|\widehat{\alpha}^{t,[i]} - \widehat{\alpha}^t\| \le C(|\varepsilon_i| + \sqrt{\log d}). \tag{129}$$

Proof. Fixing a constant $C_0 > 0$ large enough (depending on T) and any $\gamma > 0$, define the event

$$\mathcal{E} = \Big\{\|\mathbf{X}\|_{\text{op}} \le C_0,\ \|\widehat{\alpha}^0\| \le C_0,\ \|\boldsymbol{\theta}^*\|_2, \|\boldsymbol{\theta}^0\|_2 \le C_0\sqrt{d},\ \|\boldsymbol{\varepsilon}\|_2 \le C_0\sqrt{d},$$
$$\max_{t\in[0,T/\gamma]} \|\mathbf{b}^t\|_2 \le C_0\sqrt{d} \text{ for all large } n, d\Big\}.$$

Note that we have $\mathbf{b}^t \sim \mathcal{N}(0, t\gamma\mathbf{I})$, so $\mathbb{P}[\|\mathbf{b}^t\|_2 > C_0\sqrt{t\gamma d}] \le e^{-cd}$ for some constants $C_0, c > 0$ and all large n, d by a chi-squared tail bound. Then, taking a union bound over all $t \in [0, T/\gamma] \cap \mathbb{Z}_+$ and applying the conditions of Assumption 2.1 together with the Borel-Cantelli lemma, we see that this event $\mathcal{E}$ holds almost surely.

We restrict to the event $\mathcal{E}$. Let $C, C' > 0$ denote constants depending on C_0, T (but not on γ) and changing from instance to instance. For (126), we have by definition of $\{\boldsymbol{\theta}^t, \widehat{\alpha}^t\}$ in (119) that

$$\boldsymbol{\theta}^t = \boldsymbol{\theta}^0 - \gamma\sum_{s=0}^{t-1}\Big[\beta\mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}^s - \mathbf{y}) - s(\boldsymbol{\theta}^s, \widehat{\alpha}^s)\Big] + \sqrt{2}\mathbf{b}^t$$
$$\widehat{\alpha}^t = \widehat{\alpha}^0 + \gamma\sum_{s=0}^{t-1}\mathcal{G}\Big(\widehat{\alpha}^s, \widehat{\mathsf{P}}(\boldsymbol{\theta}^s)\Big).$$

Applying the bounds for $s(\cdot)$ and $\mathcal{G}(\cdot)$ in Assumptions 2.2 and 2.3 and the conditions of $\mathcal{E}$,

$$\|\boldsymbol{\theta}^t\| \le C\gamma\sum_{s=0}^{t-1}\Big(\|\boldsymbol{\theta}^s\| + \sqrt{d}\,\|\widehat{\alpha}^s\| + \sqrt{d}\Big) + \|\boldsymbol{\theta}^0\| + \sqrt{2}\,\|\mathbf{b}^t\|$$
$$\|\widehat{\alpha}^t\| \le C\gamma\sum_{s=0}^{t-1}\Big(\|\widehat{\alpha}^s\| + \|\boldsymbol{\theta}^s\|/\sqrt{d} + 1\Big) + \|\widehat{\alpha}^0\|$$

so

$$1 + \frac{\|\boldsymbol{\theta}^t\|}{\sqrt{d}} + \|\widehat{\alpha}^t\| \le C\gamma\sum_{s=0}^{t-1}\Big(1 + \frac{\|\boldsymbol{\theta}^s\|}{\sqrt{d}} + \|\widehat{\alpha}^s\|\Big) + 1 + \frac{\|\boldsymbol{\theta}^0\|}{\sqrt{d}} + \|\widehat{\alpha}^0\| + \sqrt{\frac{2}{d}}\|\mathbf{b}^t\|.$$

Iterating this bound over t shows

$$\frac{\|\boldsymbol{\theta}^t\|}{\sqrt{d}} + \|\widehat{\alpha}^t\| \leq (1+C\gamma)^t\left[1 + \frac{\|\boldsymbol{\theta}^0\|}{\sqrt{d}} + \|\widehat{\alpha}^0\| + \sqrt{\frac{2}{d}}\max_{s\in[0,t]}\|\mathbf{b}^s\|\right] \leq C',$$

the last bound holding for $t \leq T/\gamma$ and on $\mathcal{E}$. This establishes the first claim of (126). The other two claims of (126) for the cavity dynamics hold by the same argument, noting that on $\mathcal{E}$ we have also $\|\mathbf{X}^{(j)}\|_{\mathrm{op}}, \|\mathbf{X}^{[i]}\|_{\mathrm{op}} \leq C_0$ for all $j \in [d]$ and $i \in [n]$.

For (127), we have by definition of (124) that

$$\theta_j^{t+1,(j)} = \theta_j^{t,(j)} + \gamma\, s(\theta_j^{t,(j)}, \widehat{\alpha}^{t,(j)}) + \sqrt{2}(b_j^{t+1} - b_j^t).$$

Then

$$\theta_j^{t,(j)} = \theta_j^0 + \gamma\sum_{s=0}^{t-1} s(\theta_j^{s,(j)}, \widehat{\alpha}^{t,(j)}) + \sqrt{2}\, b_j^t$$

so applying the bound for $s(\cdot)$ in Assumption 2.2 and the bound $\|\widehat{\alpha}^{t,(j)}\| \leq C$ already shown in (126),

$$1 + |\theta_j^{t,(j)}| \leq C\gamma\sum_{s=0}^{t-1}(1 + |\theta_j^{s,(j)}|) + 1 + |\theta_j^0| + \sqrt{2}\,|b_j^t|.$$

Iterating this bound gives, for all $t \leq T/\gamma$,

$$1 + |\theta_j^{t,(j)}| \leq (1+C\gamma)^t(1 + |\theta_j^0| + \sqrt{2}\max_{s\in[0,t]}|b_j^t|) \leq C'(1 + |\theta_j^0| + \max_{t\leq T/\gamma}|b_j^t|)$$

which shows (127).

For (128), by definition,

$$\begin{aligned}\boldsymbol{\theta}^{t+1} - \boldsymbol{\theta}^{t+1,(j)} = &\Big(\mathbf{I} - \gamma\beta\mathbf{X}^\top\mathbf{X}\Big)\boldsymbol{\theta}^t - \Big(\mathbf{I} - \gamma\beta\mathbf{X}^{(j)\top}\mathbf{X}^{(j)}\Big)\boldsymbol{\theta}^{t,(j)} + \gamma\beta\Big(\mathbf{X}^\top\mathbf{y} - \mathbf{X}^{(j)\top}\mathbf{y}^{(j)}\Big)\\ &+ \gamma\Big(s(\boldsymbol{\theta}^t, \widehat{\alpha}^t) - s(\boldsymbol{\theta}^{t,(j)}, \widehat{\alpha}^{t,(j)})\Big).\end{aligned}$$

Then, applying the Lipschitz bound for $s(\cdot)$ in Assumption 2.2 and the conditions defining $\mathcal{E}$,

$$\begin{aligned}\|\boldsymbol{\theta}^{t+1} - \boldsymbol{\theta}^{t+1,(j)}\| \leq &(1+C\gamma)\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t,(j)}\| + C\gamma\sqrt{d}\|\widehat{\alpha}^{t-1} - \widehat{\alpha}^{t-1,(j)}\|\\ &+ C\gamma\underbrace{\Big(\|(\mathbf{X}^{(j)\top}\mathbf{X}^{(j)} - \mathbf{X}^\top\mathbf{X})\boldsymbol{\theta}^{t,(j)}\| + \|\mathbf{X}^\top\mathbf{y} - \mathbf{X}^{(j)\top}\mathbf{y}^{(j)}\|\Big)}_{:=\Delta_{t,j}}.\end{aligned} \tag{130}$$

Similarly, by the Lipschitz bound for $\mathcal{G}(\cdot)$ in Assumption 2.3,

$$\|\widehat{\alpha}^{t+1} - \widehat{\alpha}^{t+1,(j)}\| \leq (1+C\gamma)\|\widehat{\alpha}^t - \widehat{\alpha}^{t,(j)}\| + C\gamma\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t,(j)}\|/\sqrt{d}.$$

Combining the above two inequalities yields

$$\|\boldsymbol{\theta}^{t+1} - \boldsymbol{\theta}^{t+1,(j)}\| + \sqrt{d}\|\widehat{\alpha}^{t+1} - \widehat{\alpha}^{t+1,(j)}\| \leq (1+C\gamma)\big(\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t,(j)}\| + \sqrt{d}\|\widehat{\alpha}^t - \widehat{\alpha}^{t,(j)}\|\big) + C\gamma\Delta_{t,j}, \tag{131}$$

and hence iterating this bound and using $(\boldsymbol{\theta}^{0,(j)}, \widehat{\alpha}^{0,(j)}) = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$, for any $t \leq T/\gamma$,

$$\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t,(j)}\| + \sqrt{d}\|\widehat{\alpha}^t - \widehat{\alpha}^{t,(j)}\| \leq \sum_{s=0}^{t-1}(1+C\gamma)^s\max_{s=0}^{t-1} C\gamma\Delta_{s,j} \leq C'\max_{s=0}^{t-1}\Delta_{s,j}.$$

Let us now bound $\Delta_{t,j}$. Writing $\mathbf{x}_j \in \mathbb{R}^n$ for the j^{th} column of $\mathbf{X}$, we have $\mathbf{X}^{(j)} = \mathbf{X} - \mathbf{x}_j\mathbf{e}_j^\top$, hence $\mathbf{X}^\top\mathbf{X} - \mathbf{X}^{(j)\top}\mathbf{X}^{(j)} = \mathbf{X}^\top\mathbf{x}_j\mathbf{e}_j^\top + \mathbf{e}_j\mathbf{x}_j^\top\mathbf{X} - \mathbf{e}_j\mathbf{x}_j^\top\mathbf{x}_j\mathbf{e}_j^\top$, and

$$\begin{aligned}\|(\mathbf{X}^{(j)\top}\mathbf{X}^{(j)} - \mathbf{X}^\top\mathbf{X})\boldsymbol{\theta}^{t,(j)}\| &\leq \|\mathbf{X}^\top\mathbf{x}_j\||\theta_j^{t,(j)}| + |\mathbf{x}_j^\top\mathbf{X}\boldsymbol{\theta}^{t,(j)}| + \|\mathbf{X}\|_{\mathrm{op}}^2|\theta_j^{t,(j)}|\\ &\leq C\|\mathbf{X}\|_{\mathrm{op}}^2|\theta_j^{t,(j)}| + |\mathbf{x}_j^\top\mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}|.\end{aligned} \tag{132}$$

Similarly, we have $\mathbf{X}^\top \mathbf{y} - \mathbf{X}^{(j)\top}\mathbf{y}^{(j)} = (\mathbf{X}^\top\mathbf{X} - \mathbf{X}^{(j)\top}\mathbf{X}^{(j)})\boldsymbol{\theta}^* + (\mathbf{X} - \mathbf{X}^{(j)})^\top \boldsymbol{\varepsilon}$ so

$$\|\mathbf{X}^\top \mathbf{y} - \mathbf{X}^{(j)\top}\mathbf{y}^{(j)}\| \le C\|\mathbf{X}\|_{\mathrm{op}}^2 |\theta_j^*| + |\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^*| + |\mathbf{x}_j^\top \boldsymbol{\varepsilon}|. \tag{133}$$

By (127), we have $|\theta_j^{t,(j)}| \le C(1 + |\theta_j^0| + \sup_{t\in[0,T/\gamma]} |b_j^t|)$ for $t \le \lfloor T/\gamma \rfloor$. Applying this in the above two bounds yields, on $\mathcal{E}$,

$$\Delta_{t,j} \le C\Big[1 + |\theta_j^0| + |\theta_j^*| + \sup_{t\in[0,T/\gamma]} |b_j^t| + |\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^*| + |\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}| + |\mathbf{x}_j^\top \boldsymbol{\varepsilon}|\Big].$$

Define the additional event $\mathcal{E}'$ where

$$\sup_{j\in[d]} \max_{t\in[0,T/\gamma]} |\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^*| + |\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}| + |\mathbf{x}_j^\top \boldsymbol{\varepsilon}| \le C_0\sqrt{\log d} \text{ for all large } n, d.$$

Then the desired bound (128) holds for all large n, d on the event $\mathcal{E} \cap \mathcal{E}'$, so it remains to show that $\mathcal{E}'$ holds almost surely for sufficiently large $C_0 > 0$. For each $j \in [d]$ and $t \in [0, T/\gamma]$, by independence between $\mathbf{x}_j$ and $\mathbf{X}^{(j)}, \boldsymbol{\theta}^{t,(j)}$, we have that $\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}$ is subgaussian conditional on $\mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}$, so

$$\mathbb{P}\left[|\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}| \ge C\sqrt{\frac{\|\mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}\| \log d}{d}}\right] \le e^{-cd}$$

for some constants $C, c > 0$ (conditional on $\mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}$, and hence also unconditionally). Then, taking a union bound over $j \in [d]$ and $t \in [0, T/\gamma] \cap \mathbb{Z}_+$ and applying the Borel-Cantelli lemma, almost surely for all large n, d,

$$\sup_{j,t} |\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}| \le C \sup_{j,t} \sqrt{\frac{\|\mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}\| \log d}{d}}.$$

On the event $\mathcal{E}$ we have $\sup_{j,t} \|\mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}\| \le C\sqrt{d}$ by (126) already shown, so $\sup_{j,t} |\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^{t,(j)}| \le C'\sqrt{\log d}$ a.s. for all large n, d. The terms $|\mathbf{x}_j^\top \mathbf{X}^{(j)}\boldsymbol{\theta}^*|$ and $|\mathbf{x}_j^\top \boldsymbol{\varepsilon}|$ are bounded similarly, verifying that $\mathcal{E}'$ holds almost surely as claimed, and concluding the proof of (128).

For (129), similar to above, we have

$$\begin{aligned}\|\boldsymbol{\theta}^{t+1} - \boldsymbol{\theta}^{t+1,[i]}\| &\le (1 + C\gamma)\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t,[i]}\| + C\gamma\sqrt{d}\|\widehat{\alpha}^t - \widehat{\alpha}^{t,[i]}\| \\ &\qquad + C\gamma\underbrace{\Big(\|(\mathbf{X}^{[i]\top}\mathbf{X}^{[i]} - \mathbf{X}^\top\mathbf{X})\boldsymbol{\theta}^{t,[i]}\| + \|\mathbf{X}^\top\mathbf{y} - \mathbf{X}^{[i]\top}\mathbf{y}^{[i]}\|\Big)}_{\Delta_{t,i}}, \\ \|\widehat{\alpha}^{t+1} - \widehat{\alpha}^{t+1,[i]}\| &\le \|\widehat{\alpha}^t - \widehat{\alpha}^{t,[i]}\| + C\gamma\big(\|\widehat{\alpha}^t - \widehat{\alpha}^{t,[i]}\| + \|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t,[i]}\|/\sqrt{d}\big).\end{aligned}$$

which implies

$$\|\boldsymbol{\theta}^t - \boldsymbol{\theta}^{t,[i]}\| + \sqrt{d}\,\|\widehat{\alpha}^t - \widehat{\alpha}^{t,[i]}\| \le C \max_{s=0}^{t-1} \Delta_{s,i}.$$

Using $\mathbf{X} = \mathbf{X}^{[i]} + \mathbf{e}_i\mathbf{x}_i^\top$, where $\mathbf{x}_i \in \mathbb{R}^d$ now denotes (the transpose of) the i^{th} row of $\mathbf{X}$, we have $\mathbf{X}^\top\mathbf{X} = \mathbf{X}^{[i]\top}\mathbf{X}^{[i]} + \mathbf{x}_i\mathbf{x}_i^\top$ and $\mathbf{X}^\top\mathbf{y} - \mathbf{X}^{[i]\top}\mathbf{y}^{[i]} = \mathbf{x}_i(\mathbf{x}_i^\top\boldsymbol{\theta}^* + \varepsilon_i)$, so

$$\Delta_{t,i} \le \|\mathbf{X}\|_{\mathrm{op}}\Big(|\mathbf{x}_i^\top\boldsymbol{\theta}^{t,[i]}| + |\mathbf{x}_i^\top\boldsymbol{\theta}^*| + |\varepsilon_i|\Big).$$

Using independence between $\mathbf{x}_i$ and $\boldsymbol{\theta}^{t,[i]}, \boldsymbol{\theta}^*$, we obtain as above that on an almost sure event $\mathcal{E}'$, for all large n, d we have $\Delta_{t,i} \le C(|\varepsilon_i| + \sqrt{\log d})$ for all $t \in [0, T/\gamma] \cap \mathbb{Z}_+$ and $i \in [n]$, showing (129). □

5.1.2 Proof of Lemma 5.1

Lemma 5.3. *For any $T > 0$, on the event where $\|\mathbf{X}\|_{\mathrm{op}} \le C_0$, there exists a constant $C > 0$ (depending on T, C_0 but not γ) such that*

$$\max_{0\le s\le t\le T/\gamma} \max_{j\in[d]} \Big[\|\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,(s,j),\varepsilon}\| + \sqrt{d}\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,(s,j),\varepsilon}\|\Big] \le C\gamma,$$

$$\max_{0\le s\le t\le T/\gamma} \max_{i\in[n]} \Big[\|\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,[s,i],\varepsilon}\| + \sqrt{d}\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,[s,i],\varepsilon}\|\Big] \le C\gamma.$$

Proof. For the first statement, we fix s, j, and shorthand $\boldsymbol{\theta}^{t,(s,j),\varepsilon}, \widehat{\alpha}^{t,(s,j),\varepsilon}$ as $\boldsymbol{\theta}^{t,\varepsilon}, \widehat{\alpha}^{t,\varepsilon}$. By definition of the process (120), we have for $t \geq s+1$

$$\begin{aligned}
\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t+1,\varepsilon} &= \Big(\mathbf{I} - \gamma\beta\mathbf{X}^\top\mathbf{X} + \gamma\,\mathrm{Diag}(\partial_\theta s(\boldsymbol{\theta}^t, \widehat{\alpha}^t))\Big)\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,\varepsilon} + \gamma\nabla_\alpha s(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon},\\
\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t+1,\varepsilon} &= \Big(1 + \gamma \mathrm{d}_\alpha\mathcal{G}(\widehat{\alpha}^t, \widehat{\mathsf{P}}(\boldsymbol{\theta}^t))\Big)\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon} + \gamma\mathrm{d}_{\boldsymbol{\theta}}\mathcal{G}(\widehat{\alpha}^t, \widehat{\mathsf{P}}(\boldsymbol{\theta}^t))\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,\varepsilon}.
\end{aligned}$$

Then applying the conditions for $s(\cdot)$ and $\mathcal{G}(\cdot)$ in Assumptions 2.2 and 2.3 and $\|\mathbf{X}\|_{\mathrm{op}} \leq C_0$,

$$\begin{aligned}
\|\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t+1,\varepsilon}\| &\leq (1 + C\gamma)\|\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,\varepsilon}\| + C\sqrt{d}\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon}\|,\\
\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t+1,\varepsilon}\| &\leq (1 + C\gamma)\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon}\| + C\gamma\|\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,\varepsilon}\|/\sqrt{d},
\end{aligned}$$

where $C > 0$ is a constant independent of γ. Combining and iterating these inequalities yields, for all $t \in [s+1, T/\gamma]$,

$$\|\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t+1,\varepsilon}\| + \sqrt{d}\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t+1,\varepsilon}\| \leq (1 + C\gamma)^{t-s}\big(\|\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{s+1,(s,j),\varepsilon}\| + \sqrt{d}\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{s+1,(s,j),\varepsilon}\|\big) \leq C'\gamma, \tag{134}$$

using $\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{s+1,(s,j),\varepsilon} = \gamma\mathbf{e}_j$ and $\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{s+1,(s,j),\varepsilon} = 0$. This holds for all $s \in [0, T/\gamma]$ and $j \in [d]$, showing the first claim. The proof of the second claim is analogous, and omitted for brevity. ☐

Lemma 5.4. *Let $\{\boldsymbol{\theta}^t, \widehat{\alpha}^t\}$ be given by (119). For each $t \in \mathbb{Z}_+$ define the matrix*

$$\boldsymbol{\Omega}^t = \mathbf{I} - \gamma\beta\mathbf{X}^\top\mathbf{X} + \gamma\,\mathrm{Diag}(\partial_\theta s(\boldsymbol{\theta}^t, \widehat{\alpha}^t)) \in \mathbb{R}^{d\times d}. \tag{135}$$

Then for any fixed $s, t \in [0, T/\gamma]$ with $t \geq s+1$, almost surely

$$\lim_{n,d\to\infty} \frac{1}{d}\sum_{j=1}^d \partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,(s,j),\varepsilon} - \frac{\gamma}{d}\operatorname{Tr}\Big(\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{s+1}\Big) = 0$$

$$\lim_{n,d\to\infty} \frac{1}{n}\sum_{i=1}^n \partial_\varepsilon|_{\varepsilon=0}\eta_i^{t,[s,i],\varepsilon} - \frac{\gamma}{n}\operatorname{Tr}\Big(\mathbf{X}\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{s+1}\mathbf{X}^\top\Big) = 0$$

where by convention we set $\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{s+1} = \mathbf{I}$ for $t = s+1$.

Proof. Let us denote

$$\nabla_\alpha s(\boldsymbol{\theta}^t, \widehat{\alpha}^t) = \Big(\nabla_\alpha s(\theta_i^t, \widehat{\alpha}^t)^\top\Big)_{i=1}^d \in \mathbb{R}^{d\times K}, \quad \mathbf{r}^{t,(s,j)} = \nabla_\alpha s(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,(s,j),\varepsilon} \in \mathbb{R}^d.$$

Then

$$\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t+1,(s,j),\varepsilon} = \underbrace{\Big(I - \gamma\beta\mathbf{X}^\top\mathbf{X} + \gamma\,\mathrm{Diag}(\partial_\theta s(\boldsymbol{\theta}^t, \widehat{\alpha}^t))\Big)}_{\boldsymbol{\Omega}^t}\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,(s,j),\varepsilon} + \gamma\mathbf{r}^{t,(s,j)}.$$

Iterating this identity with $\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{s+1,(s,j),\varepsilon} = \gamma\mathbf{e}_j$ shows

$$\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,(s,j),\varepsilon} = \gamma\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{s+1}\mathbf{e}_j + \gamma\sum_{\ell=s+1}^{t-1}\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{\ell+1}\mathbf{r}^{\ell,(s,j)} \tag{136}$$

(where $\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{\ell+1} = \mathbf{I}$ for $\ell = t-1$). This implies

$$\frac{1}{d}\sum_{j=1}^d \partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,(s,j),\varepsilon} = \frac{\gamma}{d}\operatorname{Tr}\big(\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{s+1}\big) + \frac{\gamma}{d}\sum_{\ell=s+1}^{t-1}\sum_{j=1}^d \mathbf{e}_j^\top\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{\ell+1}\mathbf{r}^{\ell,(s,j)}. \tag{137}$$

On an event where $\|\mathbf{X}\|_{\mathrm{op}} \leq C_0$ for all large n, d (which holds almost surely), by the Lipschitz continuity of $s(\cdot)$ in Assumption 2.1 and bound in Lemma 5.3, we have $\|\boldsymbol{\Omega}^t\|_{\mathrm{op}} \leq C$, $\|\nabla_\alpha s(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\|_F \leq C\sqrt{d}$, and

$\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,(s,j),\varepsilon}\| \le C/\sqrt{d}$ for all t, where $C > 0$ is a constant (possibly depending on γ) changing from instance to instance. Then by Cauchy-Schwarz,

$$\Big|\sum_{j=1}^{d} \mathbf{e}_j^\top \mathbf{\Omega}^{t-1}\dots\mathbf{\Omega}^{\ell+1}\mathbf{r}^{\ell,(s,j)}\Big| \le \sqrt{\Big\|\mathbf{\Omega}^{t-1}\dots\mathbf{\Omega}^{\ell+1}\nabla_\alpha s(\boldsymbol{\theta}^\ell,\widehat{\alpha}^\ell)\Big\|_F^2}\cdot\sqrt{\sum_{j=1}^{d}\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{\ell,(s,j),\varepsilon}\|^2} \le C\sqrt{d},$$

which implies that the second term of (137) converges to 0 a.s. as $n, d \to \infty$. This proves the first claim. The proof of the second claim is analogous, and omitted for brevity. $\square$

Let us now introduce a notation for the discrete DMFT response process (57) prior to taking an expectation. Fixing a univariate process $\theta = \{\theta^t\}_{t\in\mathbb{Z}_+}$ and $\mathbb{R}^K$-valued process $\alpha = \{\alpha^t\}_{t\in\mathbb{Z}_+}$ as inputs, define the following auxiliary process $\{r_\theta^{(\theta,\alpha)}(t,s)\}_{s<t}$:

$$r_\theta^{(\theta,\alpha)}(t+1,s) = \begin{cases}\gamma & \text{for } s=t,\\ \Big(1-\gamma\delta\beta+\gamma\partial_\theta s(\theta^t,\alpha^t)\Big)r_\theta^{(\theta,\alpha)}(t,s)+\gamma\sum_{\ell=s+1}^{t-1}R_\eta^\gamma(t,\ell)r_\theta^{(\theta,\alpha)}(\ell,s) & \text{for } s<t.\end{cases} \tag{138}$$

Note that if the inputs $\{\theta^t,\alpha^t\}$ are given by the discrete-time DMFT processes defined in (57) and (60), then

$$r_\theta^{(\theta,\alpha)}(t,s) = \frac{\partial\theta_\gamma^t}{\partial u_\gamma^s}, \qquad \mathbb{E}[r_\theta^{(\theta,\alpha)}(t,s)] = R_\theta^\gamma(t,s)$$

which are precisely the auxiliary process defined in (58) and DMFT response function in (60). We will instead consider (138) with inputs $\{\theta_j^t,\widehat{\alpha}^t\}_{t\in\mathbb{Z}_+}$ given by the coordinates of $\{\boldsymbol{\theta}^t,\widehat{\alpha}^t\}$ solving (119).

Lemma 5.5. *Let $\{\boldsymbol{\theta}^t,\widehat{\alpha}^t\}_{t\in\mathbb{Z}_+}$ be defined by (119), and let $R_\theta^\gamma(t,s)$ be the response function of its DMFT limit defined in (60). Then for any fixed $s,t\in\mathbb{Z}_+$ with $s<t$, almost surely*

$$\lim_{n,d\to\infty}\frac{1}{d}\sum_{j=1}^{d} r_\theta^{(\theta_j,\widehat{\alpha})}(t,s) = R_\theta^\gamma(t,s).$$

Proof. First note that by the Lipschitz bound for $\partial_\theta s(\cdot)$ in Assumption 2.2, the a.s. convergence $\{\widehat{\alpha}^t\}\to\{\alpha^t\}$ in Lemma 4.1, and a simple induction argument, we have almost surely

$$\lim_{n,d\to\infty}\Big(\frac{1}{d}\sum_{j=1}^{d} r_\theta^{(\theta_j,\widehat{\alpha})}(t,s) - \frac{1}{d}\sum_{j=1}^{d} r_\theta^{(\theta_j,\alpha)}(t,s)\Big) = 0.$$

By a similar induction argument using the boundedness and Lipschitz-continuity of $\partial_\theta s(\cdot)$, for each fixed $s<t$, the map $(\theta_j^0,\dots,\theta_j^t)\mapsto r_\theta^{(\theta_j,\alpha)}(t+1,s)$ is Lipschitz for each j. Then by the empirical Wasserstein-2 convergence for $\boldsymbol{\theta}^0,\dots,\boldsymbol{\theta}^T$ in (66) of Lemma 4.1, almost surely

$$\lim_{n,d\to\infty}\frac{1}{d}\sum_{j=1}^{d} r_\theta^{(\theta_j,\alpha)}(t,s) = \mathbb{E}[r_\theta^{(\theta,\alpha)}(t,s)]$$

where the inputs (θ,α) on the right side are the discrete-time DMFT processes (57) and (60), and $\mathbb{E}[\cdot]$ is the expectation over their law. The lemma follows from noting that, by definition, $R_\theta^\gamma(t,s) = \mathbb{E}[r_\theta^{(\theta,\alpha)}(t,s)]$. $\square$

We now proceed to prove Lemma 5.1.

Proof of Lemma 5.1. For any fixed $s,t\in\mathbb{Z}_+$ with $s<t$, set also

$$r_\eta(t,s) = \frac{\partial\eta^t}{\partial w^s} = (\delta\beta)^{-1}R_\eta^\gamma(t,s)$$

and define the error terms

$$E_\theta^{t,(s,j)} = \partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,(s,j),\varepsilon} - r_\theta^{(\theta_j,\widehat{\alpha})}(t,s), \tag{139}$$

$$E_\eta^{t,[s,i]} = \partial_\varepsilon|_{\varepsilon=0}\eta_i^{t,[s,i],\varepsilon} - \beta^{-1}r_\eta(t,s). \tag{140}$$

We first prove by induction on t that for any $p \geq 1$ and $s,t \in \mathbb{Z}_+$ with $s < t$, almost surely

$$\lim_{n,d\to\infty} \frac{1}{d}\sum_{j=1}^d |E_\theta^{t,(s,j)}|^p = 0, \quad \lim_{n,d\to\infty} \frac{1}{n}\sum_{i=1}^n |E_\theta^{t,[s,i]}|^p = 0. \tag{141}$$

Fixing any $s \in \mathbb{Z}_+$, the base case $t = s+1$ holds, as direct calculation via (120) shows that

$$\partial_\varepsilon|_{\varepsilon=0}\theta_j^{s+1,(s,j),\varepsilon} = \gamma = r_\theta^{(\theta_j,\widehat{\alpha})}(s+1,s), \quad j \in [d],$$

and similarly via (121),

$$\partial_\varepsilon|_{\varepsilon=0}\eta_i^{s+1,[s,i],\varepsilon} = \gamma\|\mathbf{X}^\top\mathbf{e}_i\|^2 = \gamma + E_\eta^{s+1,[s,i]}$$

where $n^{-1}\sum_{i=1}^n |E_\eta^{s+1,[s,i]}|^p \to 0$ a.s. under Assumption 2.1 while $r_\eta(s+1,s) = \beta R_\theta^\gamma(s+1,s) = \gamma\beta$.

Suppose by induction that (141) holds for this fixed $s \in \mathbb{Z}_+$ up to time t. Note that Lemmas 5.4 and 5.5 then imply, with the matrix $\mathbf{\Omega}^t$ defined in (135), for each $s < t$, almost surely

$$\begin{aligned}\lim_{n,d\to\infty}\left\{\gamma \cdot \frac{1}{d}\operatorname{Tr}\left(\mathbf{\Omega}^{t-1}\dots\mathbf{\Omega}^{s+1}\right),\ \frac{1}{d}\sum_{j=1}^d \partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,(s,j),\varepsilon}\right\} &= R_\theta^\gamma(t,s),\\ \lim_{n,d\to\infty}\left\{\gamma \cdot \frac{1}{n}\operatorname{Tr}\left(\mathbf{X}\mathbf{\Omega}^{t-1}\dots\mathbf{\Omega}^{s+1}\mathbf{X}^\top\right),\ \frac{1}{n}\sum_{i=1}^n \partial_\varepsilon|_{\varepsilon=0}\eta_i^{t,[s,i],\varepsilon}\right\} &= \beta^{-1}r_\eta(t,s) = (\delta\beta^2)^{-1}R_\eta^\gamma(t,s).\end{aligned} \tag{142}$$

Claim for $\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,(s,j),\varepsilon}$. We establish the claim (141) for $E_\theta^{t+1,(s,j)}$. Fixing both $s \in \mathbb{Z}_+$ and $j \in [d]$, let us shorthand $\boldsymbol{\theta}^{t,(s,d),\varepsilon}$ as $\boldsymbol{\theta}^{t,\varepsilon}$ and recall the notations $\boldsymbol{\theta}^t = (\theta_j^t, \boldsymbol{\theta}_{-j}^t)$ from Section 5.1.1 where θ_j^t is the jth coordinate of $\boldsymbol{\theta}^t$. Writing correspondingly $\mathbf{X} = (\mathbf{x}_j, \mathbf{X}_{-j})$, we have

$$\begin{aligned}\boldsymbol{\theta}_{-j}^{t+1,\varepsilon} &= \left(\mathbf{I} - \gamma\beta\mathbf{X}_{-j}^\top\mathbf{X}_{-j}\right)\boldsymbol{\theta}_{-j}^{t,\varepsilon} - \gamma\beta\mathbf{X}_{-j}^\top(\mathbf{x}_j\theta_j^{t,\varepsilon} - \mathbf{y}) + \gamma s(\boldsymbol{\theta}_{-j}^{t,\varepsilon}, \widehat{\alpha}^{t,\varepsilon}) + \sqrt{2}(\mathbf{b}_{-j}^{t+1} - \mathbf{b}_{-j}^t)\\ \theta_j^{t+1,\varepsilon} &= \left(1 - \gamma\beta\|\mathbf{x}_j\|^2\right)\theta_j^{t,\varepsilon} - \gamma\beta\mathbf{x}_j^\top(\mathbf{X}_{-j}\boldsymbol{\theta}_{-j}^{t,\varepsilon} - \mathbf{y}) + \gamma s(\theta_j^{t,\varepsilon}, \widehat{\alpha}^{t,\varepsilon}) + \sqrt{2}(b_j^{t+1} - b_j^t)\\ \widehat{\alpha}^{t+1,\varepsilon} &= \widehat{\alpha}^{t,\varepsilon} + \gamma \cdot \mathcal{G}(\widehat{\alpha}^{t,\varepsilon}, \widehat{\mathsf{P}}(\boldsymbol{\theta}^{t,\varepsilon}))\end{aligned}$$

Define

$$\nabla_\alpha s(\boldsymbol{\theta}^t, \widehat{\alpha}^t) = \left(\nabla_\alpha s(\theta_i^t, \widehat{\alpha}^t)^\top\right)_{i=1}^d \in \mathbb{R}^{d\times K}, \quad \mathbf{r}^t = \nabla_\alpha s(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon} \in \mathbb{R}^d.$$

Then, taking the derivative of $\theta_j^{t+1,\varepsilon}$ in ε,

$$\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t+1,\varepsilon} = \left(1 - \gamma\beta\|\mathbf{x}_j\|^2 + \gamma\partial_\theta s(\theta_j^t, \widehat{\alpha}^t)\right)\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,\varepsilon} + \gamma r_j^t - \gamma\beta\mathbf{x}_j^\top\mathbf{X}_{-j}\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}_{-j}^{t,\varepsilon}. \tag{143}$$

Taking the derivative of $\boldsymbol{\theta}_{-j}^{t,\varepsilon}$ in ε,

$$\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}_{-j}^{t,\varepsilon} = \underbrace{\left(\mathbf{I} - \gamma\beta\mathbf{X}_{-j}^\top\mathbf{X}_{-j} + \gamma\operatorname{Diag}(\partial_\theta s(\boldsymbol{\theta}_{-j}^{t-1}, \widehat{\alpha}^{t-1}))\right)}_{:=\mathbf{\Omega}_{-j}^{t-1}}\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}_{-j}^{t-1,\varepsilon} - \gamma\beta\mathbf{X}_{-j}^\top\mathbf{x}_j\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t-1,\varepsilon} + \gamma\mathbf{r}_{-j}^{t-1}.$$

Then, iterating this equality and using $\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}_{-j}^{s+1,(s,j),\varepsilon} = 0$ gives

$$\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}_{-j}^{t,\varepsilon} = -\gamma\beta\sum_{k=s+1}^{t-1}\mathbf{\Omega}_{-j}^{t-1}\dots\mathbf{\Omega}_{-j}^{k+1}\mathbf{X}_{-j}^\top\mathbf{x}_j \cdot \partial_\varepsilon|_{\varepsilon=0}\theta_j^{k,\varepsilon} + \gamma\sum_{k=s+1}^{t-1}\mathbf{\Omega}_{-j}^{t-1}\dots\mathbf{\Omega}_{-j}^{k+1,\varepsilon}\mathbf{r}_{-j}^k.$$

Plugging the above expression into (143), we have

$$\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t+1,\varepsilon} = \underbrace{\Big(1-\gamma\beta\|\mathbf{x}_j\|^2+\gamma\partial_\theta s(\theta_j^t,\widehat{\alpha}^t)\Big)\cdot\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,\varepsilon}}_{(I_j)}$$
$$+(\gamma\beta)^2\sum_{k=s+1}^{t-1}\underbrace{\mathbf{x}_j^\top\mathbf{X}_{-j}\mathbf{\Omega}_{-j}^{t-1}\ldots\mathbf{\Omega}_{-j}^{k+1}\mathbf{X}_{-j}^\top\mathbf{x}_j\cdot\partial_\varepsilon|_{\varepsilon=0}\theta_j^{k,\varepsilon}}_{(II_{j,k})}$$
$$-\gamma^2\beta\sum_{k=s+1}^{t-1}\underbrace{\mathbf{x}_j^\top\mathbf{X}_{-j}\mathbf{\Omega}_{-j}^{t-1}\ldots\mathbf{\Omega}_{-j}^{k+1}\mathbf{r}_{-j}^k}_{(III_{j,k})}+\underbrace{\gamma r_j^t}_{(IV_j)}. \tag{144}$$

Analysis of (I_j). We have

$$(I_j)=(1-\gamma\delta\beta+\gamma\partial_\theta s(\theta_j^t,\widehat{\alpha}^t))r_\theta^{(\theta_j,\widehat{\alpha})}(t,s)+\mathsf{r}_1^{(j)}, \tag{145}$$

where

$$\mathsf{r}_1^{(j)}=\gamma\beta(\delta-\|\mathbf{x}_j\|^2)(\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,\varepsilon})+\Big(1-\gamma\delta\beta+\gamma\partial_\theta s(\theta_j^t,\widehat{\alpha}^t)\Big)E_\theta^{t,(s,j)}.$$

For any $p\geq 1$, by the induction hypothesis we have $d^{-1}\sum_j|E_\theta^{t,(s,j)}|^p\to 0$ a.s. By the conditions for $\mathbf{X}$ in Assumption 2.1, $d^{-1}\sum_j|\delta-\|\mathbf{x}_j\|^2|^p\to 0$ a.s. By Lemma 5.3, $\max_j|\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,\varepsilon}|\leq C$ a.s. for all large n,d, while by Assumption 2.2, $\partial_\theta s(\cdot)$ is also bounded. Combining these bounds gives $d^{-1}\sum_j|\mathsf{r}_1^{(j)}|^p\to 0$ a.s. for any $p\geq 1$.

Analysis of $(II_{j,k})$. Let

$$\mathbf{\Omega}_{-j}^{t,(j)}=\mathbf{I}-\gamma\beta\mathbf{X}_{-j}^\top\mathbf{X}_{-j}+\gamma\operatorname{Diag}(\partial_\theta s(\boldsymbol{\theta}_{-j}^{t,(j)},\widehat{\alpha}^{t,(j)}))$$

be the analogue of $\mathbf{\Omega}_{-j}^t$ defined by the cavity dynamics $\{\boldsymbol{\theta}^{t,(j)},\widehat{\alpha}^{t,(j)}\}$. We first show that a.s. for all large n,d, we have for every $j\in[d]$ that

$$|\mathsf{r}_{2,1}^{(j,k)}|:=\Big|\mathbf{x}_j^\top\mathbf{X}_{-j}\mathbf{\Omega}_{-j}^{t-1}\ldots\mathbf{\Omega}_{-j}^{k+1}\mathbf{X}_{-j}^\top\mathbf{x}_j-\mathbf{x}_j^\top\mathbf{X}_{-j}\mathbf{\Omega}_{-j}^{t-1,(j)}\ldots\mathbf{\Omega}_{-j}^{k+1,(j)}\mathbf{X}_{-j}^\top\mathbf{x}_j\Big|$$
$$\leq C\sqrt{\frac{\log d}{d}}\Big(|\theta_j^0|+|\theta_j^*|+\max_{u\in[0,t]}|b_j^u|+\sqrt{\log d}\Big). \tag{146}$$

To see this, note that

$$|\mathsf{r}_{2,1}^{(j,k)}|\leq\sum_{\ell=k+1}^{t-1}\underbrace{\Big|\mathbf{x}_j^\top\mathbf{X}_{-j}\mathbf{\Omega}_{-j}^{t-1,(j)}\ldots\mathbf{\Omega}_{-j}^{\ell+1,(j)}(\mathbf{\Omega}_{-j}^\ell-\mathbf{\Omega}_{-j}^{\ell,(j)})\mathbf{\Omega}_{-j}^{\ell-1}\ldots\mathbf{\Omega}_{-j}^{k+1}\mathbf{X}_{-j}^\top\mathbf{x}_j\Big|}_{:=T_{(j,k)}^\ell}.$$

Here $\mathbf{\Omega}_{-j}^\ell-\mathbf{\Omega}_{-j}^{\ell,(j)}=\gamma\operatorname{Diag}(\partial_\theta s(\boldsymbol{\theta}_{-j}^\ell,\widehat{\alpha}^\ell)-\partial_\theta s(\boldsymbol{\theta}_{-j}^{\ell,(j)},\widehat{\alpha}^{\ell,(j)}))$. Then we may bound

$$|T_{(j,k)}^\ell|\leq\gamma\|\mathbf{\Omega}_{-j}^{\ell+1,(j)}\ldots\mathbf{\Omega}_{-j}^{t-1,(j)}\mathbf{X}_{-j}^\top\mathbf{x}_j\|_\infty\cdot\|\mathbf{\Omega}_{-j}^{\ell-1}\ldots\mathbf{\Omega}_{-j}^{k+1}\mathbf{X}_{-j}^\top\mathbf{x}_j\|_2\cdot\|\partial_\theta s(\boldsymbol{\theta}_{-j}^\ell,\widehat{\alpha}^\ell)-\partial_\theta s(\boldsymbol{\theta}_{-j}^{\ell,(j)},\widehat{\alpha}^{\ell,(j)})\|_2$$
$$\leq C\gamma\|\mathbf{\Omega}_{-j}^{\ell+1,(j)}\ldots\mathbf{\Omega}_{-j}^{t-1,(j)}\mathbf{X}_{-j}^\top\mathbf{x}_j\|_\infty\cdot\|\mathbf{\Omega}_{-j}^{\ell-1}\|_{\mathrm{op}}\ldots\|\mathbf{\Omega}_{-j}^{k+1}\|_{\mathrm{op}}\|\mathbf{X}_{-j}^\top\mathbf{x}_j\|_2$$
$$\cdot(\|\boldsymbol{\theta}_{-j}^{\ell,(j)}-\boldsymbol{\theta}_{-j}^\ell\|_2+\sqrt{d}\|\widehat{\alpha}^{\ell,(j)}-\widehat{\alpha}^\ell\|_2). \tag{147}$$

Since $\mathbf{x}_j$ is independent of $\mathbf{\Omega}_{-j}^{t,(j)}$ and $\mathbf{X}_{-j}$, we have by a subgaussian tail bound

$$\mathbb{P}\Big[\mathbf{e}_i^\top\mathbf{\Omega}_{-j}^{\ell+1,(j)}\ldots\mathbf{\Omega}_{-j}^{t-1,(j)}\mathbf{X}_{-j}^\top\mathbf{x}_j\geq C\sqrt{\frac{\log d}{d}}\,\|\mathbf{e}_i^\top\mathbf{\Omega}_{-j}^{\ell+1,(j)}\ldots\mathbf{\Omega}_{-j}^{t-1,(j)}\mathbf{X}_{-j}^\top\|_2\Big]\leq e^{-cd}$$

for each $i = 1, \ldots, d$ and some constants $C, c > 0$. Then, taking a union bound and applying the Borel-Cantelli lemma, almost surely for all large n, d,

$$\sup_{j,\ell} \|\mathbf{\Omega}_{-j}^{\ell+1,(j)} \ldots \mathbf{\Omega}_{-j}^{t-1,(j)} \mathbf{X}_{-j}^\top \mathbf{x}_j\|_\infty \leq \sup_{j,\ell} C\sqrt{\frac{\log d}{d}}\, \|\mathbf{\Omega}_{-j}^{\ell+1,(j)}\|_{\mathrm{op}} \ldots \|\mathbf{\Omega}_{-j}^{t-1,(j)}\|_{\mathrm{op}} \|\mathbf{X}_{-j}\|_{\mathrm{op}}.$$

The right side is bounded by $C'\sqrt{(\log d)/d}$ on an almost-sure event where $\|\mathbf{X}\|_{\mathrm{op}} \leq C_0$ holds for all large n, d. Then, applying this to (147) and applying also Lemma 5.2 to bound the last term of (147), this shows (146). By the conditions of Assumption 2.1 and the tail estimates of the Brownian motion in Lemma 4.7, this bound (146) in turn implies $d^{-1}\sum_j |\mathsf{r}_{2,1}^{(j,k)}|^p \to 0$ a.s. for any $p \geq 1$.

Now consider

$$\mathsf{r}_{2,2}^{(j,k)} := \mathbf{x}_j^\top \mathbf{X}_{-j} \mathbf{\Omega}_{-j}^{t-1,(j)} \ldots \mathbf{\Omega}_{-j}^{k+1,(j)} \mathbf{X}_{-j}^\top \mathbf{x}_j - \frac{1}{d} \operatorname{Tr}\Big(\mathbf{X}_{-j} \mathbf{\Omega}_{-j}^{t-1,(j)} \ldots \mathbf{\Omega}_{-j}^{k+1,(j)} \mathbf{X}_{-j}^\top\Big).$$

Since $\mathbf{x}_j$ is independent of $\mathbf{\Omega}_{-j}^{t,(j)}$ and $\mathbf{X}_{-j}$, the Hanson-Wright inequality yields

$$\mathbb{P}\Big[|\mathsf{r}_{2,2}^{(j,k)}| \geq \max\Big(\frac{C\sqrt{\log d}}{d}\|\mathbf{W}\|_F, \frac{C\log d}{d}\|\mathbf{W}\|_{\mathrm{op}}\Big)\Big] \leq e^{-cd}$$

for some $C, c > 0$, where $\mathbf{W} = \mathbf{X}_{-j}\mathbf{\Omega}_{-j}^{t-1,(j)} \ldots \mathbf{\Omega}_{-j}^{k+1,(j)} \mathbf{X}_{-j}^\top$. Again taking a union bound over $j \in [d]$ and applying $\|\mathbf{W}\|_{\mathrm{op}} \leq C_0$ and $\|\mathbf{W}\|_F \leq C_0\sqrt{d}$ a.s. for all large n, d, this implies $d^{-1}\sum_j |\mathsf{r}_{2,2}^{(j,k)}|^p \to 0$ a.s. for any $p \geq 1$.

Finally, let $\mathbf{X}^{(j)} \in \mathbb{R}^{n \times d}$ be the embedding of $\mathbf{X}_{-j}$ with j^{th} column set to 0 as defined in (123), and let

$$\mathbf{\Omega}^{t,(j)} = \mathbf{I} - \gamma\beta \mathbf{X}^{(j)\top}\mathbf{X}^{(j)} + \gamma \operatorname{Diag}(\partial_\theta s(\boldsymbol{\theta}^{t,(j)}, \widehat{\alpha}^{t,(j)})) \in \mathbb{R}^{d \times d}.$$

Consider

$$\begin{aligned}
\mathsf{r}_{2,3}^{(j,k)} &:= \frac{1}{d}\operatorname{Tr}\Big(\mathbf{X}_{-j}\mathbf{\Omega}_{-j}^{t-1,(j)} \ldots \mathbf{\Omega}_{-j}^{k+1,(j)}\mathbf{X}_{-j}^\top\Big) - \frac{1}{d}\operatorname{Tr}\Big(\mathbf{X}\mathbf{\Omega}^{t-1}\ldots\mathbf{\Omega}^{k+1}\mathbf{X}^\top\Big) \\
&= \frac{1}{d}\operatorname{Tr}\Big(\mathbf{X}^{(j)}\mathbf{\Omega}^{t-1,(j)}\ldots\mathbf{\Omega}^{k+1,(j)}\mathbf{X}^{(j)\top}\Big) - \frac{1}{d}\operatorname{Tr}\Big(\mathbf{X}\mathbf{\Omega}^{t-1}\ldots\mathbf{\Omega}^{k+1}\mathbf{X}^\top\Big) \\
&= \frac{1}{d}\operatorname{Tr}\Big(\mathbf{X}^{(j)}\mathbf{\Omega}^{t-1,(j)}\ldots\mathbf{\Omega}^{k+1,(j)}(\mathbf{X}^{(j)} - \mathbf{X})^\top\Big) \\
&\qquad + \sum_{\ell=k+1}^{t-1}\frac{1}{d}\operatorname{Tr}\Big(\mathbf{X}^{(j)}\mathbf{\Omega}^{t-1,(j)}\ldots\mathbf{\Omega}^{\ell+1,(j)}(\mathbf{\Omega}^\ell - \mathbf{\Omega}^{\ell,(j)})\mathbf{\Omega}^{\ell-1}\ldots\mathbf{\Omega}^{k+1}\mathbf{X}^\top\Big) \\
&\qquad + \frac{1}{d}\operatorname{Tr}\Big((\mathbf{X}^{(j)} - \mathbf{X})\mathbf{\Omega}^{t-1}\ldots\mathbf{\Omega}^{k+1}\mathbf{X}^\top\Big)
\end{aligned}$$

Almost surely for all large n, d, for every $j \in [d]$ we have $\|\mathbf{X}^{(j)} - \mathbf{X}\|_F = \|\mathbf{x}_j\| \leq C$ and

$$\begin{aligned}
\|\mathbf{\Omega}^{t,(j)} - \mathbf{\Omega}^t\|_F &\leq \gamma\beta\|\mathbf{X}^{(j)\top}\mathbf{X}^{(j)} - \mathbf{X}^\top\mathbf{X}\|_F + \gamma\|\partial_\theta s(\boldsymbol{\theta}^{t,(j)}, \widehat{\alpha}^{t,(j)}) - \partial_\theta s(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\| \\
&\leq C\Big(1 + |\theta_j^0| + |\theta_j^*| + \max_{u\in[0,t]}|b_j^u| + \sqrt{\log d}\Big),
\end{aligned}$$

the second inequality applying the Lipschitz continuity of $s(\cdot)$ in Assumption 2.2 and Lemma 5.2. Then, applying $\operatorname{Tr}(A-B)C \leq \|A-B\|_F\|C\|_F \leq \sqrt{d}\,\|A-B\|_F\|C\|_{\mathrm{op}}$, we obtain a.s. for all large n, d that for every $j \in [d]$,

$$|r_{2,3}^{(j,k)}| \leq \frac{C}{\sqrt{d}}\Big(1 + |\theta_j^0| + |\theta_j^*| + \max_{u\in[0,t]}|b_j^u| + \sqrt{\log d}\Big),$$

which implies as above that $d^{-1}\sum_j |\mathsf{r}_{2,3}^{(j,k)}|^p \to 0$ a.s. for any $p \geq 1$.

Combining these bounds for $\mathsf{r}_{2,1}^{(j,k)}, \mathsf{r}_{2,2}^{(j,k)}, \mathsf{r}_{2,3}^{(j,k)}$, the second statement of (142) for almost sure convergence of $d^{-1}\operatorname{Tr}(\mathbf{X}\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{k+1}\mathbf{X}^\top)$, the induction hypothesis for approximation of $\partial_\varepsilon|_{\varepsilon=0}\theta_j^{k,\varepsilon}$ by $r_\theta^{(\theta_j,\widehat{\alpha})}$, and the bound $|\partial_\varepsilon|_{\varepsilon=0}\theta_j^{k,\varepsilon}| \le C$ a.s. for all large n, d by Lemma 5.3, we get that

$$(II_{j,k}) = \frac{1}{\gamma\beta^2} R_\eta^\gamma(t,k) r_\theta^{(\theta_j,\widehat{\alpha})}(k,s) + \mathsf{r}_2^{(j,k)} \tag{148}$$

where $d^{-1}\sum_j |\mathsf{r}_2^{(j,k)}|^p \to 0$ a.s. for any $p \ge 1$.

Analysis of $(III_{j,k})$**.** We apply a similar leave-one-out argument as above. Let

$$\mathbf{r}_{-j}^{t,(j)} = \nabla_\alpha s(\boldsymbol{\theta}_{-j}^{t,(j)}, \widehat{\alpha}^{t,(j)})\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon} \in \mathbb{R}^{d-1}.$$

(Note that we replace only the first factor $\nabla_\alpha s(\boldsymbol{\theta}_{-j}^{t,(j)}, \widehat{\alpha}^{t,(j)})$ by the cavity dynamics, leaving the second factor unchanged.) Then

$$\begin{aligned}
&\left|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1}\ldots\boldsymbol{\Omega}_{-j}^{k+1}\mathbf{r}_{-j}^k - \mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{k+1,(j)}\mathbf{r}_{-j}^{k,(j)}\right| \\
&\le \underbrace{\sum_{\ell=k+1}^{t-1}\left|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{\ell+1,(j)}(\boldsymbol{\Omega}_{-j}^{\ell} - \boldsymbol{\Omega}_{-j}^{\ell,(j)})\boldsymbol{\Omega}_{-j}^{\ell-1}\ldots\boldsymbol{\Omega}_{-j}^{k+1}\mathbf{r}_{-j}^k\right|}_{:=u^{(j,k)}} \\
&\qquad + \underbrace{\left|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{k+1,(j)}(\mathbf{r}_{-j}^k - \mathbf{r}_{-j}^{k,(j)})\right|}_{:=v^{(j,k)}}.
\end{aligned} \tag{149}$$

We note that a.s. for all large n, d, we have $\|\mathbf{r}^k\| \le C\sqrt{d}\cdot C/\sqrt{d} \le C'$ by the Lipschitz bound for $s(\cdot)$ in Assumption 2.2 and Lemma 5.3. Then, using similar arguments as in the analysis of $\mathsf{r}_{2,1}^{(j,k)}$ above, we have $d^{-1}\sum_j |u^{(j,k)}|^p \to 0$ a.s. for any $p \ge 1$. For the second term, we have

$$\begin{aligned}
v^{(j,k)} &= \left|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{j+1,(j)}\Big(\nabla_\alpha s(\boldsymbol{\theta}_{-j}^k, \widehat{\alpha}^k) - \nabla_\alpha s(\boldsymbol{\theta}_{-j}^{k,(j)}, \widehat{\alpha}^{k,(j)})\Big)\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{k,\varepsilon}\right| \\
&\le C\|\mathbf{x}_j\|\|\mathbf{X}_{-j}\|_{\text{op}}\|\boldsymbol{\Omega}_{-j}^{t-1,(j)}\|_{\text{op}}\ldots\|\boldsymbol{\Omega}_{-j}^{k+1,(j)}\|_{\text{op}}(\|\boldsymbol{\theta}_{-j}^k - \boldsymbol{\theta}_{-j}^{k,(j)}\| + \sqrt{d}\|\widehat{\alpha}^k - \widehat{\alpha}^{k,(j)}\|)\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{k,\varepsilon}\|,
\end{aligned}$$

which satisfies $d^{-1}\sum_j |v^{(j,k)}|^p \to 0$ a.s. for all $p \ge 1$ by Lemmas 5.2 and 5.3. Thus

$$\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1}\ldots\boldsymbol{\Omega}_{-j}^{k+1}\mathbf{r}_{-j}^k = \mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{k+1,(j)}\mathbf{r}_{-j}^{k,(j)} + \mathsf{r}^{(j,k)}$$

where $d^{-1}\sum_j |\mathsf{r}^{(j,k)}|^p \to 0$ a.s. On the other hand, we have

$$\begin{aligned}
\left|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{k+1,(j)}\mathbf{r}_{-j}^{k,(j)}\right| &= \left|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{k+1,(j)}\nabla_\alpha s(\boldsymbol{\theta}_{-j}^{k,(j)}, \widehat{\alpha}^{k,(j)})\cdot\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{k,\varepsilon}\right| \\
&\le \|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{k+1,(j)}\nabla_\alpha s(\boldsymbol{\theta}_{-j}^{k,(j)}, \widehat{\alpha}^{k,(j)})\|\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{k,\varepsilon}\|.
\end{aligned}$$

Since $\mathbf{X}_{-j}, \boldsymbol{\Omega}_{-j}^{t,(j)}, \boldsymbol{\theta}_{-j}^{k,(j)}, \widehat{\alpha}^{k,(j)}$ are all independent of $\mathbf{x}_j$, a subgaussian tail bound and union bound shows, a.s. for all large n, d, that for every $j \in [d]$,

$$\|\mathbf{x}_j^\top \mathbf{X}_{-j}\boldsymbol{\Omega}_{-j}^{t-1,(j)}\ldots\boldsymbol{\Omega}_{-j}^{k+1,(j)}\nabla_\alpha s(\boldsymbol{\theta}_{-j}^{k,(j)}, \widehat{\alpha}^{k,(j)})\| \le C\sqrt{\log d}.$$

Since $\|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{k,\varepsilon}\| \le C/\sqrt{d}$ by Lemma 5.3, this shows

$$(III_{j,k}) = \mathsf{r}_3^{(j,k)} \tag{150}$$

where $\lim_{n,d\to\infty} d^{-1}\sum_{j=1}^d |\mathsf{r}_3^{(j,k)}|^p = 0$ a.s. for any $p \ge 1$.

Analysis of (IV_j)**.** By Assumption 2.2 and Lemma 5.3, $|(IV_j)| \le \gamma \|\nabla_\alpha s(\theta_j^t, \widehat{\alpha}^t)\| \|\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon}\| \le C/\sqrt{d}$ a.s. for all large n, d, hence

$$(IV_j) = \mathsf{r}_4^{(j)} \tag{151}$$

where $\lim_{n,d\to\infty} d^{-1}\sum_{j=1}^d |\mathsf{r}_4^{(j)}|^p = 0$ a.s. for any $p \ge 1$.

Applying (145), (148), (150), and (151) back to (144),

$$\begin{aligned}\partial_\varepsilon|_{\varepsilon=0}\theta_j^{t+1,\varepsilon} &= \big(1 - \gamma\delta\beta + \gamma\partial_\theta s(\theta_j^t, \widehat{\alpha}^t)\big) r_\theta^{(\theta_j,\widehat{\alpha})}(t,s) + \gamma \sum_{k=s+1}^{t-1} R_\eta^\gamma(t,k) r_\theta^{(\theta_j,\widehat{\alpha})}(k,s) + E_\theta^{t+1,(s,j)} \\ &= r_\theta^{(\theta_j,\widehat{\alpha})}(t+1,s) + E_\theta^{t+1,(s,j)}\end{aligned}$$

where $\lim_{n,d\to\infty} d^{-1}\sum_{j=1}^d |E_\theta^{t+1,(s,j)}|^p = 0$ a.s. for each $p \ge 1$, concluding the proof the inductive claim (141) for $E_\theta^{t+1,(s,j)}$.

Claim for $\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\eta}_i^{t,[s,i],\varepsilon}$**.** We now show the claim (141) for $E_\eta^{t+1,[s,i]}$. Again fixing $s \in \mathbb{Z}_+$ and $i \in [n]$, let us shorthand $\boldsymbol{\eta}^{t,[s,i],\varepsilon}$ and $\boldsymbol{\theta}^{t,[s,i],\varepsilon}$ as $\boldsymbol{\eta}^{t,\varepsilon}$ and $\boldsymbol{\theta}^{t,\varepsilon}$. Let us write $\boldsymbol{\eta}^t = (\eta_i^t, \boldsymbol{\eta}_{-i}^t)$ as in Section 5.1.1, and write correspondingly $\mathbf{y} = (y_i, \mathbf{y}_{-i})$, $\boldsymbol{\varepsilon} = (\varepsilon_i, \boldsymbol{\varepsilon}_{-i})$, and $\mathbf{X} = [\mathbf{x}_i, \mathbf{X}_{-i}^\top]^\top$ where $\mathbf{x}_i \in \mathbb{R}^d$ denotes now (the transpose of) the i^{th} *row* of $\mathbf{X}$, and $\mathbf{X}_{-i} \in \mathbb{R}^{(n-1)\times d}$. Then

$$\begin{aligned}\boldsymbol{\theta}^{t+1,\varepsilon} &= \boldsymbol{\theta}^{t,\varepsilon} + \gamma\Big[-\beta\big(\mathbf{X}_{-i}^\top(\mathbf{X}_{-i}\boldsymbol{\theta}^{t,\varepsilon} - \mathbf{y}_{-i}) + \mathbf{x}_i(\eta_i^{t,\varepsilon} - y_i)\big) + s(\boldsymbol{\theta}^{t,\varepsilon}, \widehat{\alpha}^{t,\varepsilon})\Big] + \sqrt{2}(\mathbf{b}^{t+1} - \mathbf{b}^t) \\ \eta_i^{t+1,\varepsilon} &= \eta_i^{t,\varepsilon} + \gamma\Big[-\beta\mathbf{x}_i^\top\big(\mathbf{X}_{-i}^\top(\mathbf{X}_{-i}\boldsymbol{\theta}^{t,\varepsilon} - \mathbf{y}_{-i}) + \mathbf{x}_i(\eta_i^{t,\varepsilon} - y_i)\big) + \mathbf{x}_i^\top s(\boldsymbol{\theta}^{t,\varepsilon}, \widehat{\alpha}^{t,\varepsilon})\Big] + \sqrt{2}\mathbf{x}_i^\top(\mathbf{b}^{t+1} - \mathbf{b}^t) \\ \widehat{\alpha}^{t+1,\varepsilon} &= \widehat{\alpha}^{t,\varepsilon} + \gamma \cdot \mathcal{G}(\widehat{\alpha}^{t,\varepsilon}, \widehat{\mathsf{P}}(\boldsymbol{\theta}^{t,\varepsilon})).\end{aligned}$$

Set

$$\mathbf{r}^t = \nabla_\alpha s(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\partial_\varepsilon|_{\varepsilon=0}\widehat{\alpha}^{t,\varepsilon} \in \mathbb{R}^d.$$

Then, taking the derivative of $\eta_i^{t+1,\varepsilon}$ yields

$$\partial_\varepsilon|_{\varepsilon=0}\eta_i^{t+1,\varepsilon} = \Big(1 - \gamma\beta\|\mathbf{x}_i\|^2\Big)\partial_\varepsilon|_{\varepsilon=0}\eta_i^{t,\varepsilon} + \mathbf{x}_i^\top\Big(-\gamma\beta\mathbf{X}_{-i}^\top\mathbf{X}_{-i} + \gamma\,\mathrm{Diag}(\partial_\theta s(\boldsymbol{\theta}^t, \widehat{\alpha}^t))\Big)\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,\varepsilon} + \gamma\mathbf{x}_i^\top\mathbf{r}^t. \tag{152}$$

Taking derivative of $\boldsymbol{\theta}^{t,\varepsilon}$ yields

$$\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,\varepsilon} = \underbrace{\Big(\mathbf{I} - \gamma\beta\mathbf{X}_{-i}^\top\mathbf{X}_{-i} + \gamma\,\mathrm{Diag}(\partial_\theta s(\boldsymbol{\theta}^{t-1}, \widehat{\alpha}^{t-1}))\Big)}_{:=\boldsymbol{\Omega}_{-i}^{t-1}}\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t-1,\varepsilon} - \gamma\beta\mathbf{x}_i\partial_\varepsilon|_{\varepsilon=0}\eta_i^{t-1,\varepsilon} + \gamma\mathbf{r}^{t-1}.$$

Iterating this equality and using $\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{s+1,[s,i],\varepsilon} = \gamma\mathbf{x}_i$ gives

$$\partial_\varepsilon|_{\varepsilon=0}\boldsymbol{\theta}^{t,\varepsilon} = \gamma\boldsymbol{\Omega}_{-i}^{t-1}\ldots\boldsymbol{\Omega}_{-i}^{s+1}\mathbf{x}_i - \gamma\beta\sum_{k=s+1}^{t-1}\boldsymbol{\Omega}_{-i}^{t-1}\ldots\boldsymbol{\Omega}_{-i}^{k+1}\mathbf{x}_i \cdot \partial_\varepsilon|_{\varepsilon=0}\eta_i^{k,\varepsilon} + \gamma\sum_{k=s+1}^{t-1}\boldsymbol{\Omega}_{-i}^{t-1}\ldots\boldsymbol{\Omega}_{-i}^{k+1}\mathbf{r}^k.$$

Plugging the above expression into (152), we have

$$\begin{aligned}\partial_\varepsilon|_{\varepsilon=0}\eta_i^{t+1,\varepsilon} &= \underbrace{\Big(1 - \gamma\beta\|\mathbf{x}_i\|^2\Big)\partial_\varepsilon|_{\varepsilon=0}\eta_i^{t,\varepsilon}}_{(I_i)} + \gamma\underbrace{\mathbf{x}_i^\top(\boldsymbol{\Omega}_{-i}^t - \mathbf{I})\boldsymbol{\Omega}_{-i}^{t-1}\ldots\boldsymbol{\Omega}_{-i}^{s+1}\mathbf{x}_i}_{(II_i)} \\ &- \gamma\beta\sum_{k=s+1}^{t-1}\underbrace{\mathbf{x}_i^\top(\boldsymbol{\Omega}_{-i}^t - \mathbf{I})\boldsymbol{\Omega}_{-i}^{t-1}\ldots\boldsymbol{\Omega}_{-i}^{k+1}\mathbf{x}_i \cdot \partial_\varepsilon|_{\varepsilon=0}\eta_i^{k,\varepsilon}}_{(III_{i,k})} + \gamma\sum_{k=s+1}^{t-1}\underbrace{\mathbf{x}_i^\top(\boldsymbol{\Omega}_{-i}^t - \mathbf{I})\boldsymbol{\Omega}^{t-1}\ldots\boldsymbol{\Omega}^{k+1}\mathbf{r}^k}_{(IV)_{i,k}} + \underbrace{\gamma\mathbf{x}_i^\top\mathbf{r}^t}_{(V_i)}.\end{aligned} \tag{153}$$

The arguments to analyze these terms are similar to the above, and we will omit some details.

Analysis of (I_i)**.** By the induction hypothesis and concentration of $\|\mathbf{x}_i\|^2$ around 1,

$$(I_i) = (1-\gamma\beta)\beta^{-1} r_\eta(t,s) + \mathsf{r}_1^{[i]} \tag{154}$$

where $n^{-1}\sum_{i=1}^n |\mathsf{r}_1^{[i]}|^p \to 0$ a.s. for any $p \geq 1$.

Analysis of (II_i)**.** Let $\mathbf{\Omega}_{-i}^{t,[i]} = \mathbf{I} - \gamma\beta\mathbf{X}_{-i}^\top \mathbf{X}_{-i} + \gamma\,\mathrm{Diag}(\partial_\theta s(\boldsymbol{\theta}^{t,[i]}, \widehat{\alpha}^{t,[i]}))$ with $\boldsymbol{\theta}^{t,[i]}, \widehat{\alpha}^{t,[i]}$ given by the cavity dynamics of (125). Set

$$\begin{aligned}
\mathsf{r}_{2,1}^{[i]} &= \mathbf{x}_i^\top \mathbf{\Omega}_{-i}^{t-1} \dots \mathbf{\Omega}_{-i}^{s+1} \mathbf{x}_i - \mathbf{x}_i^\top \mathbf{\Omega}_{-i}^{t-1,[i]} \dots \mathbf{\Omega}_{-i}^{s+1,[i]} \mathbf{x}_i \\
\mathsf{r}_{2,2}^{[i]} &= \mathbf{x}_i^\top \mathbf{\Omega}_{-i}^{t-1,[i]} \dots \mathbf{\Omega}_{-i}^{s+1,[i]} \mathbf{x}_i - \frac{1}{d}\operatorname{Tr} \mathbf{\Omega}_{-i}^{t-1,[i]} \dots \mathbf{\Omega}_{-i}^{s+1,[i]} \\
\mathsf{r}_{2,3}^{[i]} &= \frac{1}{d}\operatorname{Tr} \mathbf{\Omega}_{-i}^{t-1,[i]} \dots \mathbf{\Omega}_{-i}^{s+1} - \frac{1}{d}\operatorname{Tr} \mathbf{\Omega}^{t-1} \dots \mathbf{\Omega}^{s+1,[i]}.
\end{aligned}$$

Then the same arguments above yield $n^{-1}\sum_{i=1}^n |\mathsf{r}_{2,j}^{[i]}|^p \to 0$ for each $j = 1, 2, 3$. Applying the same arguments for t in place of $t-1$, and the first statement of (142) for both t and $t-1$,

$$(II_i) = \gamma^{-1} R_\theta^\gamma(t+1,s) - \gamma^{-1} R_\theta^\gamma(t,s) + \mathsf{r}_2^{[i]} \tag{155}$$

where $n^{-1}\sum_{i=1}^n |\mathsf{r}_2^{[i]}|^p \to 0$ a.s.

Analysis of $(III_{i,k})$**,** $(IV_{i,k})$**, and** (V_i)**.** Similar arguments as above show

$$(III_{i,k}) = (\gamma\beta)^{-1} R_\theta^\gamma(t+1,k) r_\eta(k,s) - (\gamma\beta)^{-1} R_\theta^\gamma(t,k) r_\eta(k,s) + \mathsf{r}_3^{[i,k]} \tag{156}$$

$$(IV)_{i,k} = \mathsf{r}_4^{[i,k]} \tag{157}$$

$$(V)_i = \mathsf{r}_5^{[i]} \tag{158}$$

where $n^{-1}\sum_{i=1}^n |\mathsf{r}_3^{[i,k]}|^p \to 0$, $n^{-1}\sum_{i=1}^n |\mathsf{r}_4^{[i,k]}|^p \to 0$, and $n^{-1}\sum_{i=1}^n |\mathsf{r}_5^{[i]}|^p \to 0$ a.s.

Applying (154), (155), (156), (157), and (158) back to (153), for an error term $E_\eta^{t+1,[s,i]}$ satisfying $n^{-1}\sum_{i=1}^n |E_\eta^{t+1,[s,i]}|^p \to 0$ a.s., we have

$$\begin{aligned}
\partial_\varepsilon|_{\varepsilon=0}\eta_i^{t+1,\varepsilon} &= (1-\gamma\beta)\beta^{-1} r_\eta(t,s) + R_\theta^\gamma(t+1,s) - R_\theta^\gamma(t,s) \\
&\qquad - \sum_{k=s+1}^{t-1} (R_\theta^\gamma(t+1,k) - R_\theta^\gamma(t,s)) r_\eta(k,s) + E_\eta^{t+1,[s,i]} \\
&= \Big[-\gamma r_\eta(t,s) + R_\theta^\gamma(t+1,s) - \sum_{k=s+1}^{t-1} R_\theta^\gamma(t+1,k) r_\eta(k,s)\Big] \\
&\qquad + \underbrace{\Big[\beta^{-1} r_\eta(t,s) - R_\theta^\gamma(t,s) + \sum_{k=s+1}^{t-1} R_\theta^\gamma(t,k) r_\eta(k,s)\Big]}_{=0} + E_\eta^{t+1,[s,i]} \\
&= R_\theta^\gamma(t+1,s) - \sum_{k=s+1}^{t} R_\theta^\gamma(t+1,k) r_\eta(k,s) + E_\eta^{t+1,[s,i]} \\
&= \beta^{-1} r_\eta(t+1,s) + E_\eta^{t+1,[s,i]}.
\end{aligned}$$

This shows the inductive claim (141) for $E_\eta^{t+1,[s,i]}$, and hence concludes the induction.

To conclude the proof, by boundedness of $\partial_\theta s(\cdot)$ and the definition (138), $d^{-1}\sum_{j=1}^d r_\theta^{(\theta_j,\widehat{\alpha})}(t,s)$ is bounded by a constant. Furthermore, by the expansion (137) and its following arguments (which hold also at non-zero $\varepsilon > 0$), on the event $\|\mathbf{X}\|_{\mathrm{op}} \le C_0$, we have that $d^{-1}\sum_{j=1}^d \partial_\varepsilon \theta_j^{t,(s,j),\varepsilon}$ is also bounded by a constant for all sufficiently small $\varepsilon \ge 0$. Then, writing $\langle \cdot \rangle$ for the expectation over only the discrete Brownian motion $\{\mathbf{b}^t\}_{t\in\mathbb{Z}_+}$, we may apply the dominated convergence theorem to the first statement of (141) to get, almost surely,

$$\begin{aligned}
\lim_{n,d\to\infty} d^{-1}\operatorname{Tr}\mathbf{R}_\theta(t,s) &= \lim_{n,d\to\infty} \frac{1}{d}\sum_{j=1}^d \partial_\varepsilon|_{\varepsilon=0}\langle \theta_j^{t,(s,j),\varepsilon}\rangle = \lim_{n,d\to\infty} \frac{1}{d}\sum_{j=1}^d \langle \partial_\varepsilon|_{\varepsilon=0}\theta_j^{t,(s,j),\varepsilon}\rangle \\
&= \lim_{n,d\to\infty} \frac{1}{d}\sum_{j=1}^d \langle r_\theta^{(\theta_j,\alpha)}(t,s)\rangle = R_\theta^\gamma(t,s),
\end{aligned}$$

the last equality holding by Lemma 5.5. Similarly we may apply the dominated convergence theorem to the second statement of (141) to get, almost surely,

$$\lim_{n,d\to\infty} n^{-1}\operatorname{Tr}\mathbf{R}_\eta(t,s) = \lim_{n,d\to\infty} \delta\beta^2\cdot\frac{1}{n}\sum_{i=1}^n \partial_\varepsilon|_{\varepsilon=0}\langle \eta_i^{t,[s,i],\varepsilon}\rangle = \lim_{n,d\to\infty} \delta\beta\cdot\frac{1}{n}\sum_{i=1}^n \langle r_\eta(t,s)\rangle = R_\eta^\gamma(t,s),$$

concluding the proof. □

5.2 Discretization of Langevin response function

In the following, we denote $\mathbf{x} = (\boldsymbol{\theta},\widehat{\alpha}) \in \mathbb{R}^{d+K}$ and consider (4–5) as a joint diffusion in the variables $\mathbf{x}^t = (\boldsymbol{\theta}^t,\widehat{\alpha}^t)$. Let $u : \mathbb{R}^{d+K} \to \mathbb{R}^{d+K}$ and $\mathbf{M} \in \mathbb{R}^{(d+K)\times(d+K)}$ be defined by

$$\begin{aligned}
u(\mathbf{x}) = u(\boldsymbol{\theta},\widehat{\alpha}) &= \Big(-\beta\mathbf{X}^\top(\mathbf{X}\boldsymbol{\theta}-\mathbf{y}) + \big(s(\theta_j,\widehat{\alpha})\big)_{j=1}^d,\ \mathcal{G}(\widehat{\alpha},\widehat{\mathsf{P}}(\boldsymbol{\theta}))\Big) \\
\mathbf{M} &= \operatorname{Diag}(\mathbf{I}_{d\times d}, 0_{K\times K})
\end{aligned} \tag{159}$$

Given an initial condition $\mathbf{x}^0 \in \mathbb{R}^{d+K}$, we consider the continuous-time dynamics for $\mathbf{x}^t \in \mathbb{R}^{d+K}$ and $\mathbf{V}^t \in \mathbb{R}^{(d+K)\times(d+K)}$ defined by

$$\begin{aligned}
\mathbf{x}^t &= \mathbf{x}^0 + \int_0^t u(\mathbf{x}^s)\mathrm{d}s + \sqrt{2}\,\mathbf{M}\mathbf{b}^t \\
\mathbf{V}^t &= \mathbf{I}_{d+K} + \int_0^t [\mathrm{d}u(\mathbf{x}^s)\mathbf{V}^s]\mathrm{d}s
\end{aligned} \tag{160}$$

where $\mathrm{d}u(\mathbf{x}) \in \mathbb{R}^{(d+K)\times(d+K)}$ is the derivative of $u(\cdot)$ at $\mathbf{x}$. We consider also a piecewise-constant version of these dynamics

$$\begin{aligned}
\bar{\mathbf{x}}_\gamma^t &= \mathbf{x} + \int_0^{\lfloor t\rfloor} u(\bar{\mathbf{x}}_\gamma^s)\mathrm{d}s + \sqrt{2}\,\mathbf{M}\mathbf{b}^{\lfloor t\rfloor} \\
\bar{\mathbf{V}}_\gamma^t &= \mathbf{I}_{d+K} + \int_0^{\lfloor t\rfloor} [\mathrm{d}u(\bar{\mathbf{x}}_\gamma^s)\bar{\mathbf{V}}_\gamma^s]\mathrm{d}s
\end{aligned} \tag{161}$$

where $\lfloor t\rfloor \in \gamma\mathbb{Z}_+$ is as previously defined in (92). We note that the process $\mathbf{x}^t = (\boldsymbol{\theta}^t,\widehat{\alpha}^t)$ in (160) is precisely our adaptive Langevin process of interest (4–5). Similarly, the process $\bar{\mathbf{x}}_\gamma^t$ in (161) is the piecewise-constant embedding from Section 4.3 of the discrete dynamics for $\mathbf{x}_\gamma^t = (\boldsymbol{\theta}_\gamma^t,\widehat{\alpha}_\gamma^t)$ which we have rewritten in (119). Denoting $[t] = \lfloor t\rfloor/\gamma \in \mathbb{Z}_+$ as in (92), we have

$$\bar{\mathbf{x}}_\gamma^t = \mathbf{x}_\gamma^{[t]} = (\boldsymbol{\theta}_\gamma^{[t]},\widehat{\alpha}_\gamma^{[t]}) \text{ for all } t \ge 0. \tag{162}$$

Throughout, we will write $\langle\cdot\rangle_{\mathbf{x}^0}$ for expectations only over the Brownian motion $\mathbf{b}^t$, i.e. conditional on $\mathbf{X},\boldsymbol{\theta}^*,\boldsymbol{\varepsilon}$ and the initial condition $\mathbf{x}^0$.

Lemma 5.6. *Let us write the block forms*

$$\mathbf{V}^t = \begin{pmatrix} \mathbf{U}^t & * \\ \mathbf{W}^t & * \end{pmatrix}, \quad \bar{\mathbf{V}}^t_\gamma = \begin{pmatrix} \bar{\mathbf{U}}^t_\gamma & * \\ \bar{\mathbf{W}}^t_\gamma & * \end{pmatrix}, \quad \mathrm{d}u(\mathbf{x}^t) = \begin{pmatrix} \mathbf{J}^t_1 & \mathbf{J}^t_2 \\ \mathbf{J}^t_3 & \mathbf{J}^t_4 \end{pmatrix}, \quad \mathrm{d}u(\bar{\mathbf{x}}^t_\gamma) = \begin{pmatrix} \bar{\mathbf{J}}^t_{\gamma,1} & \bar{\mathbf{J}}^t_{\gamma,2} \\ \bar{\mathbf{J}}^t_{\gamma,3} & \bar{\mathbf{J}}^t_{\gamma,4} \end{pmatrix}$$

with blocks of sizes d and K. Fixing any $T > 0$, on the event $\{\|\mathbf{X}\|_{\mathrm{op}} \leq C_0, \|\mathbf{y}\| \leq C_0\sqrt{d}\}$, there is a constant $C > 0$ (depending on T, C_0 but not on γ) such that for any $\gamma > 0$, we have

$$\begin{aligned} &\sup_{t\in[0,T]} \|\mathbf{J}^t_1\|_{\mathrm{op}}, \|\bar{\mathbf{J}}^t_{\gamma,1}\|_{\mathrm{op}} \leq C, \qquad \sup_{t\in[0,T]} \|\mathbf{J}^t_2\|_F, \|\bar{\mathbf{J}}^t_{\gamma,2}\|_F \leq C\sqrt{d}, \\ &\sup_{t\in[0,T]} \|\mathbf{J}^t_3\|_F, \|\bar{\mathbf{J}}^t_{\gamma,3}\|_F \leq C/\sqrt{d}, \qquad \sup_{t\in[0,T]} \|\mathbf{J}^t_4\|_F, \|\bar{\mathbf{J}}^t_{\gamma,4}\|_F \leq C, \end{aligned} \tag{163}$$

$$\sup_{t\in[0,T]} \{\|\mathbf{U}^t\|_{\mathrm{op}}, \sqrt{d}\|\mathbf{W}^t\|_F, \|\bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}}, \sqrt{d}\|\bar{\mathbf{W}}^t_\gamma\|_F\} \leq C, \tag{164}$$

$$\sup_{t\in[0,T]} \|\bar{\mathbf{U}}^{t+\gamma}_\gamma - \bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}} \leq C\gamma, \qquad \sup_{t\in[0,T]} \|\bar{\mathbf{W}}^{t+\gamma}_\gamma - \bar{\mathbf{W}}^t_\gamma\|_F \leq C\gamma/\sqrt{d}. \tag{165}$$

Furthermore, for some $\iota : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\lim_{\gamma\to 0} \iota(\gamma) = 0$ and for any initial condition $\mathbf{x}^0 = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$,

$$\sup_{t\in[0,T]} \frac{\langle \|\mathbf{J}^t_1 - \bar{\mathbf{J}}^t_{\gamma,1}\|_F\rangle_{\mathbf{x}^0}}{\sqrt{d}}, \frac{\langle \|\mathbf{J}^t_2 - \bar{\mathbf{J}}^t_{\gamma,2}\|_F\rangle_{\mathbf{x}^0}}{\sqrt{d}}, \sqrt{d}\langle \|\mathbf{J}^t_3 - \bar{\mathbf{J}}^t_{\gamma,3}\|_F\rangle_{\mathbf{x}^0}, \langle \|\mathbf{J}^t_4 - \bar{\mathbf{J}}^t_{\gamma,4}\|_F\rangle_{\mathbf{x}^0} \leq \iota(\gamma)\Big(\frac{\|\boldsymbol{\theta}^0\|}{\sqrt{d}} + \|\widehat{\alpha}^0\| + 1\Big). \tag{166}$$

Proof. For (163), we have by definition that

$$\begin{aligned} &\mathbf{J}^t_1 = -\beta\mathbf{X}^\top\mathbf{X} + \mathrm{Diag}\Big[\Big(\partial_\theta s(\theta^t_j, \widehat{\alpha}^t)\Big)_{j=1}^d\Big], \quad \mathbf{J}^t_2 = \Big(\nabla_\alpha s(\theta^t_j, \widehat{\alpha}^t)^\top\Big)_{j=1}^d, \\ &\mathbf{J}^t_3 = \mathrm{d}_{\boldsymbol{\theta}}\mathcal{G}(\widehat{\alpha}^t, \mathsf{P}(\boldsymbol{\theta}^t)), \quad \mathbf{J}^t_4 = \mathrm{d}_\alpha\mathcal{G}(\widehat{\alpha}^t, \mathsf{P}(\boldsymbol{\theta}^t)), \end{aligned}$$

and similarly for $\bar{\mathbf{J}}^t_{\gamma,1}, \bar{\mathbf{J}}^t_{\gamma,2}, \bar{\mathbf{J}}^t_{\gamma,3}, \bar{\mathbf{J}}^t_{\gamma,4}$. Then the desired bounds (163) hold on the event where $\|\mathbf{X}\|_{\mathrm{op}} \leq C_0$, by Assumptions 2.2 and 2.3 for the derivatives of $s(\cdot)$ and $\mathcal{G}(\cdot)$.

For (164), let us first prove the bounds for the discrete dynamics $\|\bar{\mathbf{U}}_\gamma\|_{\mathrm{op}}$ and $\|\bar{\mathbf{W}}_\gamma\|_F$. By definition, for each $t \in \gamma\mathbb{Z}_+$,

$$\bar{\mathbf{U}}^{t+\gamma}_\gamma = (\mathbf{I} + \gamma\bar{\mathbf{J}}^t_{\gamma,1})\bar{\mathbf{U}}^t_\gamma + \gamma\bar{\mathbf{J}}^t_{\gamma,2}\bar{\mathbf{W}}^t_\gamma, \quad \bar{\mathbf{W}}^{t+1}_\gamma = \gamma\bar{\mathbf{J}}^t_{\gamma,3}\bar{\mathbf{U}}^t_\gamma + (\mathbf{I} + \gamma\bar{\mathbf{J}}^t_{\gamma,4})\bar{\mathbf{W}}^t_\gamma. \tag{167}$$

Then applying (163),

$$\|\bar{\mathbf{U}}^{t+\gamma}_\gamma\|_{\mathrm{op}} \leq (1 + C\gamma)\|\bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}} + C\gamma\sqrt{d}\|\bar{\mathbf{W}}^t_\gamma\|_F, \quad \|\bar{\mathbf{W}}^{t+\gamma}_\gamma\|_F \leq \frac{C\gamma}{\sqrt{d}}\|\bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}} + (1 + C\gamma)\|\bar{\mathbf{W}}^t_\gamma\|_F,$$

which further implies that

$$\|\bar{\mathbf{U}}^{t+\gamma}_\gamma\|_{\mathrm{op}} + \sqrt{d}\|\bar{\mathbf{W}}^{t+\gamma}_\gamma\|_F \leq (1 + 2C\gamma)\big(\|\bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}} + \sqrt{d}\|\bar{\mathbf{W}}^t_\gamma\|_F\big).$$

Iterating this bound from the initial conditions $\bar{\mathbf{U}}^0_\gamma = \mathbf{I}_d$ and $\bar{\mathbf{W}}^0_\gamma = 0_{K\times d}$ shows (164) for $\bar{\mathbf{U}}^t_\gamma, \bar{\mathbf{W}}^t_\gamma$ and all $t \leq T$. For the continuous version $\|\mathbf{U}^t\|_{\mathrm{op}}$ and $\|\mathbf{W}^t\|_F$, note that analogously

$$\mathbf{U}^t = \mathbf{U}^0 + \int_0^t (\mathbf{J}^s_1\mathbf{U}^s + \mathbf{J}^s_2\mathbf{W}^s)\mathrm{d}s, \quad \mathbf{W}^t = \mathbf{W}^0 + \int_0^t (\mathbf{J}^s_3\mathbf{U}^s + \mathbf{J}^s_4\mathbf{W}^s)\mathrm{d}s$$

so $\frac{\mathrm{d}}{\mathrm{d}t}(\|\mathbf{U}^t\|_{\mathrm{op}} + \sqrt{d}\|\mathbf{W}^t\|_F) \leq C(\|\mathbf{U}^t\|_{\mathrm{op}} + \sqrt{d}\|\mathbf{W}^t\|_F)$. Then (164) follows by Gronwall's lemma.

For (165), we have by (167) and (163)

$$\begin{aligned} &\|\bar{\mathbf{U}}^{t+\gamma}_\gamma - \bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}} \leq \gamma\Big(\|\bar{\mathbf{J}}^t_{\gamma,1}\|_{\mathrm{op}}\|\bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}} + \|\bar{\mathbf{J}}^t_{\gamma,2}\|_F\|\bar{\mathbf{W}}^t_\gamma\|_F\Big) \leq C\gamma, \\ &\|\bar{\mathbf{W}}^{t+\gamma}_\gamma - \bar{\mathbf{W}}^t_\gamma\|_F \leq \gamma\Big(\|\bar{\mathbf{J}}^t_{\gamma,3}\|_F\|\bar{\mathbf{U}}^t_\gamma\|_{\mathrm{op}} + \|\bar{\mathbf{J}}^t_{\gamma,4}\|_F\|\bar{\mathbf{W}}^t_\gamma\|_F\Big) \leq C\gamma/\sqrt{d}. \end{aligned}$$

For (166), we have by the Lipschitz continuity of $s(\cdot)$ in Assumption 2.2 that $\mathbf{J}_1^t - \bar{\mathbf{J}}_{\gamma,1}^t$ is diagonal with $\|\mathbf{J}_1^t - \bar{\mathbf{J}}_{\gamma,1}^t\|_F \le C(\|\boldsymbol{\theta}^t - \bar{\boldsymbol{\theta}}_\gamma^t\| + \sqrt{d}\|\widehat{\alpha}^t - \bar{\widehat{\alpha}}_\gamma^t\|)$. Next using the arguments that led to (109), we have that on the event $\{\|\mathbf{X}\|_{\text{op}} \le C_0, \|\mathbf{y}\| \le C_0\sqrt{d}\}$, with $\mathbf{x}^0 = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$,

$$\langle \|\boldsymbol{\theta}^t\| \rangle_{\mathbf{x}^0} + \langle \|\bar{\boldsymbol{\theta}}_\gamma^t\| \rangle_{\mathbf{x}^0} + \sqrt{d}\langle \|\widehat{\alpha}^t\| \rangle_{\mathbf{x}^0} + \sqrt{d}\langle \|\bar{\widehat{\alpha}}_\gamma^t\| \rangle_{\mathbf{x}^0} \le C(\|\boldsymbol{\theta}^0\| + \sqrt{d}\|\widehat{\alpha}^0\| + \sqrt{d}) \tag{168}$$

and

$$\langle \|\boldsymbol{\theta}^t - \bar{\boldsymbol{\theta}}_\gamma^t\| \rangle_{\mathbf{x}^0} + \sqrt{d}\langle \|\widehat{\alpha}^t - \bar{\widehat{\alpha}}_\gamma^t\| \rangle_{\mathbf{x}^0} \le \iota(\gamma)(\|\boldsymbol{\theta}^0\| + \sqrt{d}\|\widehat{\alpha}^0\| + \sqrt{d}). \tag{169}$$

This implies the desired bound for $\langle \|\mathbf{J}_1^t - \bar{\mathbf{J}}_{\gamma,1}^t\|_F \rangle_{\mathbf{x}}$, and a similar argument leads to the bound for $\langle \|\mathbf{J}_2^t - \bar{\mathbf{J}}_{\gamma,2}^t\|_F \rangle_{\mathbf{x}}$. Next, by the derivative bounds for $\mathcal{G}(\cdot)$ in Assumption 2.3, we have $\|\mathbf{J}_t^3 - \bar{\mathbf{J}}_{\gamma,3}^t\|_F \le C(\|\boldsymbol{\theta}^t - \bar{\boldsymbol{\theta}}_\gamma^t\|/d + \|\widehat{\alpha}^t - \bar{\widehat{\alpha}}_\gamma^t\|/\sqrt{d})$ and $\|\mathbf{J}_4^t - \bar{\mathbf{J}}_{\gamma,4}^t\|_F \le C(\|\boldsymbol{\theta}^t - \bar{\boldsymbol{\theta}}_\gamma^t\|/\sqrt{d} + \|\widehat{\alpha}^t - \bar{\widehat{\alpha}}_\gamma^t\|)$, hence the desired bounds also follow by (169). □

Lemma 5.7. *Define*

$$\mathcal{E} = \{\|\mathbf{X}\|_{\text{op}} \le C_0, \|\mathbf{y}\| \le C_0\sqrt{d}, \|\boldsymbol{\theta}^0\| \le C_0\sqrt{d}, \|\widehat{\alpha}^0\| \le C_0 \textit{ for all large } n, d\}. \tag{170}$$

Fixing any $T > 0$, there exists a constant $C > 0$ (depending on T, C_0 but not on γ) and a function $\iota : \mathbb{R}_+ \to \mathbb{R}_+$ satisfying $\lim_{\gamma \to 0} \iota(\gamma) = 0$, such that on $\mathcal{E}$, for any $\gamma > 0$ and all $0 \le s \le t \le T$,

$$|d^{-1} \operatorname{Tr} \mathbf{R}_\theta(t, s) - d^{-1}\gamma^{-1} \operatorname{Tr} \mathbf{R}_\theta^\gamma([t] + 1, [s])| \le \iota(\gamma) \tag{171}$$

$$|n^{-1} \operatorname{Tr} \mathbf{R}_\eta(t, s) - n^{-1}\gamma^{-1} \operatorname{Tr} \mathbf{R}_\eta^\gamma([t] + 1, [s])| \le \iota(\gamma). \tag{172}$$

Proof. **Discretization of $\mathbf{R}_\theta$.** Let $\{P_t^\gamma\}_{t \in \mathbb{Z}_+}$ be the Markov semigroup for the discrete dynamics (119), i.e. $P_t^\gamma f(\mathbf{x}) = \langle f(\mathbf{x}_\gamma^t) \rangle_{\mathbf{x}}$. Then applying Proposition A.4, for any $s, t \in \mathbb{Z}_+$ with $s < t$,

$$\partial_\varepsilon|_{\varepsilon=0} \langle \theta_{\gamma,j}^{t,(s,j),\varepsilon} \rangle_{\mathbf{x}} = \gamma P_{s+1}^\gamma \partial_j P_{t-s-1}^\gamma e_j(\mathbf{x}).$$

This implies, for the given initial condition of the dynamics $\mathbf{x}^0 = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$, that

$$\gamma^{-1} \operatorname{Tr} \mathbf{R}_\theta^\gamma(t, s) = \sum_{j=1}^d P_{s+1}^\gamma \partial_j P_{t-s-1}^\gamma e_j(\mathbf{x}^0).$$

Let $\{P_t\}_{t \ge 0}$ analogously denote the Markov semigroup of the continuous dynamics (4–5), i.e. $P_t f(\mathbf{x}) = \langle f(\mathbf{x}^t) \rangle_{\mathbf{x}}$. Then applying Proposition A.1, for any $s, t \in \mathbb{R}_+$ with $s \le t$,

$$\operatorname{Tr} \mathbf{R}_\theta(t, s) = \sum_{j=1}^d P_s \partial_j P_{t-s} e_j(\mathbf{x}^0).$$

Thus, for all $s, t \in \mathbb{R}_+$ with $s \le t$,

$$\begin{aligned}\Big|\operatorname{Tr} \mathbf{R}_\theta(t, s) - \gamma^{-1} \operatorname{Tr} \mathbf{R}_\theta^\gamma([t] + 1, [s])\Big| &= \Big|\sum_{j=1}^d P_s \partial_j P_{t-s} e_j(\mathbf{x}^0) - \sum_{j=1}^d P_{[s]+1}^\gamma \partial_j P_{[t]-[s]}^\gamma e_j(\mathbf{x}^0)\Big| \\ &\le \underbrace{\Big|P_s\Big(\sum_{j=1}^d \partial_j P_{t-s} e_j - \sum_{j=1}^d \partial_j P_{[t]-[s]}^\gamma e_j\Big)(\mathbf{x}^0)\Big|}_{(I)} + \underbrace{\Big|(P_s - P_{[s]+1}^\gamma)\Big(\sum_{j=1}^d \partial_j P_{[t]-[s]}^\gamma e_j\Big)(\mathbf{x}^0)\Big|}_{(II)}.\end{aligned}$$

Bound of (I)**.** By Proposition A.2(c), $\sum_{j=1}^d \partial_j P_{t-s} e_j(\mathbf{x}) = \sum_{j=1}^d \langle (\mathbf{V}^{t-s})_{jj} \rangle_{\mathbf{x}}$, where $\{\mathbf{x}^t, \mathbf{V}^t\}_{t \geq 0}$ are the solution to (160) with initial condition $\mathbf{x}^0 = \mathbf{x}$. Similarly, by Lemma A.5 and the identification (162), $\sum_{j=1}^d \partial_j P^\gamma_{[t]-[s]} e_j(\mathbf{x}) = \sum_{j=1}^d \langle (\bar{\mathbf{V}}_\gamma^{\lfloor t \rfloor - \lfloor s \rfloor})_{jj} \rangle_{\mathbf{x}}$, where $\{\bar{\mathbf{x}}^t_\gamma, \bar{\mathbf{V}}^t_\gamma\}_{t \geq 0}$ are the solution to (161). Let us write

$$\mathbf{V}^t = \begin{pmatrix} \mathbf{U}^t & * \\ \mathbf{W}^t & * \end{pmatrix}, \quad \bar{\mathbf{V}}^t_\gamma = \begin{pmatrix} \bar{\mathbf{U}}^t_\gamma & * \\ \bar{\mathbf{W}}^t_\gamma & * \end{pmatrix}, \quad \mathrm{d}u(\mathbf{x}^t) = \begin{pmatrix} \mathbf{J}_1^t & \mathbf{J}_2^t \\ \mathbf{J}_3^t & \mathbf{J}_4^t \end{pmatrix}, \quad \mathrm{d}u(\bar{\mathbf{x}}^t_\gamma) = \begin{pmatrix} \bar{\mathbf{J}}^t_{\gamma,1} & \bar{\mathbf{J}}^t_{\gamma,2} \\ \bar{\mathbf{J}}^t_{\gamma,3} & \bar{\mathbf{J}}^t_{\gamma,4} \end{pmatrix}$$

with blocks of sizes d and K. Then

$$\begin{aligned} \Big| \sum_{j=1}^d \partial_j P_{t-s} e_j(\mathbf{x}) - \partial_j P^\gamma_{[t]-[s]} \mathbf{e}_j(\mathbf{x}) \Big| &= \Big| \langle \operatorname{Tr} \mathbf{U}^{t-s} - \operatorname{Tr} \bar{\mathbf{U}}_\gamma^{\lfloor t \rfloor - \lfloor s \rfloor} \rangle_{\mathbf{x}} \Big| \\ &\leq \sqrt{d} \langle \| \mathbf{U}^{t-s} - \bar{\mathbf{U}}_\gamma^{t-s} \|_F \rangle_{\mathbf{x}} + d \langle \| \bar{\mathbf{U}}_\gamma^{t-s} - \bar{\mathbf{U}}_\gamma^{\lfloor t \rfloor - \lfloor s \rfloor} \|_{\mathrm{op}} \rangle_{\mathbf{x}}. \end{aligned} \tag{173}$$

Since $|(t-s) - (\lfloor t \rfloor - \lfloor s \rfloor)| \leq C\gamma$, the second term satisfies $\|\bar{\mathbf{U}}^{t-s} - \bar{\mathbf{U}}^{t-s}\|_{\mathrm{op}} \leq C\gamma$ by (165). For the first term, note that by definition

$$\begin{gathered} \mathbf{U}^t = \mathbf{U}^0 + \int_0^t (\mathbf{J}_1^s \mathbf{U}^s + \mathbf{J}_2^s \mathbf{W}^s) \mathrm{d}s, \quad \mathbf{W}^t = \mathbf{W}^0 + \int_0^t (\mathbf{J}_3^s \mathbf{U}^s + \mathbf{J}_4^s \mathbf{W}^s) \mathrm{d}s, \\ \bar{\mathbf{U}}^t_\gamma = \mathbf{U}^0 + \int_0^{\lfloor t \rfloor} (\bar{\mathbf{J}}^s_{\gamma,1} \bar{\mathbf{U}}^s_\gamma + \bar{\mathbf{J}}^s_{\gamma,2} \bar{\mathbf{W}}^s_\gamma) \mathrm{d}s, \quad \bar{\mathbf{W}}^t_\gamma = \mathbf{W}^0 + \int_0^{\lfloor t \rfloor} (\bar{\mathbf{J}}^s_{\gamma,3} \bar{\mathbf{U}}^s_\gamma + \bar{\mathbf{J}}^s_{\gamma,4} \bar{\mathbf{W}}^s_\gamma) \mathrm{d}s. \end{gathered}$$

Hence

$$\begin{aligned} \langle \| \mathbf{U}^t - \bar{\mathbf{U}}^t_\gamma \|_F \rangle_{\mathbf{x}} \leq \int_0^{\lfloor t \rfloor} \Big[& \langle \| \mathbf{J}_1^s - \bar{\mathbf{J}}^s_{\gamma,1} \|_F \| \mathbf{U}^s \|_{\mathrm{op}} \rangle_{\mathbf{x}} + \langle \| \bar{\mathbf{J}}^s_{\gamma,1} \|_{\mathrm{op}} \| \mathbf{U}^s - \bar{\mathbf{U}}^s_\gamma \|_F \rangle_{\mathbf{x}} \\ & + \langle \| \mathbf{J}_2^s - \bar{\mathbf{J}}^s_{\gamma,2} \|_F \| \mathbf{W}^s \|_F \rangle_{\mathbf{x}} + \langle \| \bar{\mathbf{J}}^s_{\gamma,2} \|_F \| \mathbf{W}^s - \bar{\mathbf{W}}^s_\gamma \|_F \rangle_{\mathbf{x}} \Big] \mathrm{d}s \\ & + \int_{\lfloor t \rfloor}^t \Big[\langle \| \mathbf{J}_1^s \|_F \| \mathbf{U}^s \|_{\mathrm{op}} \rangle_{\mathbf{x}} + \langle \| \mathbf{J}_2^s \|_F \| \mathbf{W}^s \|_F \rangle_{\mathbf{x}} \Big] \mathrm{d}s \end{aligned}$$

Let $C, C' > 0$ be constants depending on T but not γ, and let $\iota(\gamma), \iota'(\gamma)$ be constants depending also on γ and satisfying $\iota(\gamma), \iota'(\gamma) \to 0$ as $\gamma \to 0$, all changing from instance to instance. By Lemma 5.6, with $\mathbf{x} = (\boldsymbol{\theta}, \widehat{\alpha})$, we have

$$\begin{gathered} \| \mathbf{U}^s \|_{\mathrm{op}} \leq C, \quad \sqrt{d} \| \mathbf{W}^s \|_F \leq C, \quad \| \mathbf{J}_1^s \|_{\mathrm{op}}, \| \bar{\mathbf{J}}^s_{\gamma,1} \|_{\mathrm{op}} \leq C, \quad \| \mathbf{J}_2^s \|_F, \| \bar{\mathbf{J}}^s_{\gamma,2} \|_F \leq C\sqrt{d}, \\ \langle \| \mathbf{J}_1^s - \bar{\mathbf{J}}^s_{\gamma,1} \|_F \rangle_{\mathbf{x}} \leq \iota(\gamma)(\| \boldsymbol{\theta} \| + \sqrt{d} \| \widehat{\alpha} \| + \sqrt{d}), \quad \langle \| \mathbf{J}_2^s - \bar{\mathbf{J}}^s_{\gamma,2} \|_F \rangle_{\mathbf{x}} \leq \iota(\gamma)(\| \boldsymbol{\theta} \| + \sqrt{d} \| \widehat{\alpha} \| + \sqrt{d}), \end{gathered}$$

hence

$$\langle \| \mathbf{U}^t - \bar{\mathbf{U}}^t_\gamma \|_F \rangle_{\mathbf{x}} \leq C \int_0^t \big(\langle \| \mathbf{U}^s - \bar{\mathbf{U}}^s_\gamma \|_F \rangle_{\mathbf{x}} + \sqrt{d} \langle \| \mathbf{W}^s - \bar{\mathbf{W}}^s_\gamma \|_F \rangle_{\mathbf{x}} \big) \mathrm{d}s + \iota(\gamma)(\| \boldsymbol{\theta} \| + \sqrt{d} \| \widehat{\alpha} \| + \sqrt{d}). \tag{174}$$

Next we have

$$\begin{aligned} \langle \| \mathbf{W}^t - \bar{\mathbf{W}}^t_\gamma \|_F \rangle_{\mathbf{x}} \leq \int_0^{\lfloor t \rfloor} \Big[& \langle \| \mathbf{J}_3^s - \bar{\mathbf{J}}^s_{\gamma,3} \|_F \| \mathbf{U}^s \|_{\mathrm{op}} \rangle_{\mathbf{x}} + \langle \| \bar{\mathbf{J}}^s_{\gamma,3} \|_F \| \mathbf{U}^s - \bar{\mathbf{U}}^s_\gamma \|_F \rangle_{\mathbf{x}} \\ & + \langle \| \mathbf{J}_4^s - \bar{\mathbf{J}}^s_{\gamma,4} \|_F \| \mathbf{W}^s \|_F \rangle_{\mathbf{x}} + \langle \| \bar{\mathbf{J}}^s_{\gamma,4} \|_F \| \mathbf{W}^s - \bar{\mathbf{W}}^s_\gamma \|_F \rangle_{\mathbf{x}} \Big] \mathrm{d}s \\ & + \int_{\lfloor t \rfloor}^t \Big[\langle \| \mathbf{J}_3^s \|_F \| \mathbf{U}^s \|_{\mathrm{op}} \rangle_{\mathbf{x}} + \langle \| \mathbf{J}_4^s \|_F \| \mathbf{W}^s \|_F \rangle_{\mathbf{x}} \Big] \mathrm{d}s. \end{aligned}$$

By Lemma 5.6, we have also

$$\| \mathbf{J}_3^s \|_F, \| \bar{\mathbf{J}}^s_{\gamma,3} \|_F \leq C/\sqrt{d}, \quad \| \mathbf{J}_4^s \|_F, \| \bar{\mathbf{J}}^s_{\gamma,4} \|_F \leq C,$$

$$d\langle\|\mathbf{J}_3^s-\bar{\mathbf{J}}_{\gamma,3}^s\|_F\rangle_{\mathbf{x}}\le\iota(\gamma)(\|\boldsymbol{\theta}\|+\sqrt{d}\|\widehat{\alpha}\|+\sqrt{d}),\quad \sqrt{d}\|\mathbf{J}_4^s-\bar{\mathbf{J}}_{\gamma,4}^s\|_F\le\iota(\gamma)(\|\boldsymbol{\theta}\|+\sqrt{d}\|\widehat{\alpha}\|+\sqrt{d}),$$

which implies that

$$\sqrt{d}\langle\|\mathbf{W}^t-\bar{\mathbf{W}}_\gamma^t\|_F\rangle_{\mathbf{x}}\le C\int_0^t(\langle\|\mathbf{U}^s-\bar{\mathbf{U}}^s\|_F\rangle_{\mathbf{x}}+\sqrt{d}\langle\|\mathbf{W}^s-\bar{\mathbf{W}}^s\|_F\rangle_{\mathbf{x}})\mathrm{d}s+\iota(\gamma)\Big(\frac{\|\boldsymbol{\theta}^0\|}{\sqrt{d}}+\|\widehat{\alpha}^0\|+1\Big).\quad(175)$$

Combining (174) and (175) yields

$$\begin{aligned}&\langle\|\mathbf{U}^t-\bar{\mathbf{U}}_\gamma^t\|_F+\sqrt{d}\|\mathbf{W}^t-\bar{\mathbf{W}}_\gamma^t\|_F\rangle_{\mathbf{x}}\\&\le C\int_0^t(\langle\|\mathbf{U}^s-\bar{\mathbf{U}}_\gamma^s\|_F+\sqrt{d}\|\mathbf{W}^s-\bar{\mathbf{W}}_\gamma^s\|_F\rangle_{\mathbf{x}})\mathrm{d}s+\iota(\gamma)(\|\boldsymbol{\theta}\|+\sqrt{d}\|\widehat{\alpha}\|+\sqrt{d}),\end{aligned}$$

so Gronwall's lemma gives $\sup_{t\in[0,T]}\langle\|\mathbf{U}^t-\bar{\mathbf{U}}_\gamma^t\|_F\rangle_{\mathbf{x}}+\sqrt{d}\langle\|\mathbf{W}^t-\bar{\mathbf{W}}_\gamma^t\|_F\rangle_{\mathbf{x}}\le\iota(\gamma)(\|\boldsymbol{\theta}\|+\sqrt{d}\|\widehat{\alpha}\|+\sqrt{d})$.

Hence the bound (173) reads, for $\mathbf{x}=(\boldsymbol{\theta},\widehat{\alpha})$,

$$\Big|\sum_{j=1}^d\partial_jP_{t-s}e_j(\mathbf{x})-\partial_jP^\gamma_{[t]-[s]-1}e_j(\mathbf{x})\Big|\le\iota(\gamma)(\sqrt{d}\|\boldsymbol{\theta}\|+d\|\widehat{\alpha}\|+d).\quad(176)$$

Applying this with $\mathbf{x}=\mathbf{x}^s=(\boldsymbol{\theta}^s,\widehat{\alpha}^s)$, this implies that

$$(I)\le\iota(\gamma)(\sqrt{d}\langle\|\boldsymbol{\theta}^s\|\rangle_{\mathbf{x}^0}+d\langle\|\widehat{\alpha}^s\|\rangle_{\mathbf{x}^0}+d)\le\iota(\gamma)d,$$

the last step using the bound (168) and conditions for $(\boldsymbol{\theta}^0,\widehat{\alpha}^0)$ on the event (170).

Bound of (II)**.** Let $\mathsf{f}(\mathbf{x})=\sum_{j=1}^d\partial_jP^\gamma_{[t]-[s]}e_j(\mathbf{x})$. We first establish a Lipschitz bound for f: Let $\{\bar{\mathbf{x}}_\gamma^t,\bar{\mathbf{V}}_\gamma^t\}_{t\in\mathbb{Z}_+}$ and $\{\tilde{\mathbf{x}}_\gamma^t,\tilde{\mathbf{V}}_\gamma^t\}_{t\in\mathbb{Z}_+}$ be defined by (161) with initializations $\mathbf{x}=(\boldsymbol{\theta},\widehat{\alpha})$ and $\tilde{\mathbf{x}}=(\tilde{\boldsymbol{\theta}},\tilde{\widehat{\alpha}})$ respectively, coupled by the same Brownian motion. We write $\langle\cdot\rangle$ for the average over this Brownian motion, and denote by $\tilde{\mathbf{U}}_\gamma^t,\tilde{\mathbf{W}}_\gamma^t$ and $\tilde{\mathbf{J}}_{\gamma,1}^t,\tilde{\mathbf{J}}_{\gamma,2}^t,\tilde{\mathbf{J}}_{\gamma,3}^t,\tilde{\mathbf{J}}_{\gamma,4}^t$ the blocks of $\tilde{\mathbf{V}}_\gamma^t$ and $\mathrm{d}u(\tilde{\mathbf{x}}_\gamma^t)$. Then, using $\mathsf{f}(\mathbf{x})=\langle\operatorname{Tr}\bar{\mathbf{U}}_\gamma^\tau\rangle$ with $\tau=\lfloor t\rfloor-\lfloor s\rfloor$ as established above,

$$|\mathsf{f}(\mathbf{x})-\mathsf{f}(\tilde{\mathbf{x}})|\le|\langle\operatorname{Tr}\bar{\mathbf{U}}_\gamma^\tau-\operatorname{Tr}\tilde{\mathbf{U}}_\gamma^\tau\rangle|\le\sqrt{d}\langle\|\bar{\mathbf{U}}_\gamma^\tau-\tilde{\mathbf{U}}_\gamma^\tau\|_F\rangle.\quad(177)$$

We apply a similar argument as in term (I), noting that

$$\begin{aligned}\|\bar{\mathbf{U}}_\gamma^{t+\gamma}-\tilde{\mathbf{U}}_\gamma^{t+\gamma}\|_F&\le\gamma\|\bar{\mathbf{J}}_{\gamma,1}^t-\tilde{\mathbf{J}}_{\gamma,1}^t\|_F\|\bar{\mathbf{U}}_\gamma^t\|_{\mathrm{op}}+(1+\gamma\|\tilde{\mathbf{J}}_{\gamma,1}^t\|_{\mathrm{op}})\|\bar{\mathbf{U}}_\gamma^t-\tilde{\mathbf{U}}_\gamma^t\|_F\\&\quad+\gamma\|\bar{\mathbf{J}}_{\gamma,2}^t-\tilde{\mathbf{J}}_{\gamma,2}^t\|_F\|\bar{\mathbf{W}}_\gamma^t\|_F+\gamma\|\tilde{\mathbf{J}}_{\gamma,2}^t\|_F\|\bar{\mathbf{W}}_\gamma^t-\tilde{\mathbf{W}}_\gamma^t\|_F,\\\|\bar{\mathbf{W}}_\gamma^{t+\gamma}-\tilde{\mathbf{W}}_\gamma^{t+\gamma}\|_F&\le\gamma\|\bar{\mathbf{J}}_{\gamma,3}^t-\tilde{\mathbf{J}}_{\gamma,3}^t\|_F\|\bar{\mathbf{U}}_\gamma^t\|_{\mathrm{op}}+\gamma\|\tilde{\mathbf{J}}_{\gamma,3}^t\|_F\|\bar{\mathbf{U}}_\gamma^t-\tilde{\mathbf{U}}_\gamma^t\|_F\\&\quad+\gamma\|\bar{\mathbf{J}}_{\gamma,4}^t-\tilde{\mathbf{J}}_{\gamma,4}^t\|_F\|\bar{\mathbf{W}}_\gamma^t\|_F+(1+\gamma\|\tilde{\mathbf{J}}_{\gamma,4}^t\|_F)\|\bar{\mathbf{W}}_\gamma^t-\tilde{\mathbf{W}}_\gamma^t\|_F.\end{aligned}$$

By Lemma 5.6, we have $\|\bar{\mathbf{U}}_\gamma^t\|_{\mathrm{op}},\sqrt{d}\|\bar{\mathbf{W}}_\gamma^t\|_F\le C$, and $\|\tilde{\mathbf{J}}_{\gamma,1}^t\|_{\mathrm{op}},\frac{\|\tilde{\mathbf{J}}_{\gamma,2}^t\|_F}{\sqrt{d}},\sqrt{d}\|\tilde{\mathbf{J}}_{\gamma,3}^t\|_F,\|\tilde{\mathbf{J}}_{\gamma,4}^t\|_F\le C$. Furthermore similar arguments to (166) in Lemma 5.6 show that

$$\begin{aligned}&\langle\|\bar{\mathbf{J}}_{\gamma,1}^t-\tilde{\mathbf{J}}_{\gamma,1}^t\|_F\rangle,\langle\|\bar{\mathbf{J}}_{\gamma,2}^t-\tilde{\mathbf{J}}_{\gamma,2}^t\|_F\rangle,d\langle\|\bar{\mathbf{J}}_{\gamma,3}^t-\tilde{\mathbf{J}}_{\gamma,3}^t\|_F\rangle,\sqrt{d}\langle\|\bar{\mathbf{J}}_{\gamma,4}^t-\tilde{\mathbf{J}}_{\gamma,4}^t\|_F\rangle\\&\le C\Big\langle\|\bar{\boldsymbol{\theta}}^t-\tilde{\boldsymbol{\theta}}^t\|+\sqrt{d}\|\bar{\widehat{\alpha}}^t-\tilde{\widehat{\alpha}}^t\|\Big\rangle\le C'\Big(\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\|+\sqrt{d}\|\widehat{\alpha}-\tilde{\widehat{\alpha}}\|\Big),\end{aligned}$$

the quantities in the last expression denoting the differences in initial conditions. Hence

$$\begin{aligned}&\langle\|\tilde{\mathbf{U}}_\gamma^{t+\gamma}-\bar{\mathbf{U}}_\gamma^{t+\gamma}\|_F+\sqrt{d}\|\bar{\mathbf{W}}_\gamma^{t+\gamma}-\tilde{\mathbf{W}}_\gamma^{t+\gamma}\|_F\rangle\\&\le(1+C\gamma)\langle\|\bar{\mathbf{U}}_\gamma^t-\tilde{\mathbf{U}}_\gamma^t\|_F+\sqrt{d}\|\bar{\mathbf{W}}_\gamma^t-\tilde{\mathbf{W}}_\gamma^t\|_F\rangle+C\gamma\Big(\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\|+\sqrt{d}\|\widehat{\alpha}-\tilde{\widehat{\alpha}}\|\Big).\end{aligned}$$

Iterating this bound gives $\langle\|\bar{\mathbf{U}}_\gamma^\tau-\tilde{\mathbf{U}}_\gamma^\tau\|_F\rangle\le C(\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\|+\sqrt{d}\|\widehat{\alpha}-\tilde{\widehat{\alpha}}\|)$, which applied to (177) yields our desired Lipschitz bound

$$|\mathsf{f}(\mathbf{x})-\mathsf{f}(\tilde{\mathbf{x}})|\le C\sqrt{d}(\|\boldsymbol{\theta}-\tilde{\boldsymbol{\theta}}\|+\sqrt{d}\|\widehat{\alpha}-\tilde{\widehat{\alpha}}\|).$$

Then, writing $\mathbf{x}^0 = (\boldsymbol{\theta}^0, \widehat{\alpha}^0)$ for the original initial conditions,

$$(II) = \left|(P_s - P^\gamma_{[s]+1})\mathsf{f}(\mathbf{x}^0)\right| = \left|\langle \mathsf{f}(\mathbf{x}^s)\rangle_{\mathbf{x}^0} - \langle \mathsf{f}(\bar{\mathbf{x}}_\gamma^{\lfloor s\rfloor+\gamma})\rangle_{\mathbf{x}^0}\right| \leq C\sqrt{d}\langle \|\boldsymbol{\theta}^s - \bar{\boldsymbol{\theta}}_\gamma^{\lfloor s\rfloor+\gamma}\| + \sqrt{d}\|\widehat{\alpha}^s - \bar{\widehat{\alpha}}_\gamma^{\lfloor s\rfloor+\gamma}\|\rangle_{\mathbf{x}^0}$$

where we couple $\{(\boldsymbol{\theta}^t, \widehat{\alpha}^t)\}_{t\geq 0}$ and $\{\bar{\boldsymbol{\theta}}^t_\gamma, \bar{\widehat{\alpha}}^t_\gamma\}_{t\geq 0}$ by the same Brownian motion. Bounding

$$\langle \|\boldsymbol{\theta}^s - \bar{\boldsymbol{\theta}}_\gamma^{\lfloor s\rfloor+\gamma}\|\rangle_{\mathbf{x}^0} \leq \langle \|\boldsymbol{\theta}^s - \bar{\boldsymbol{\theta}}_\gamma^{s}\|\rangle_{\mathbf{x}^0} + \langle \|\bar{\boldsymbol{\theta}}_\gamma^s - \bar{\boldsymbol{\theta}}_\gamma^{\lfloor s\rfloor+\gamma}\|\rangle_{\mathbf{x}^0}$$

and similarly for $\widehat{\alpha}$, and then applying (168) and (169), we obtain on the event (170) that

$$\langle \|\boldsymbol{\theta}^s - \bar{\boldsymbol{\theta}}_\gamma^{\lfloor s\rfloor+\gamma}\| + \sqrt{d}\|\widehat{\alpha}^s - \bar{\widehat{\alpha}}_\gamma^{\lfloor s\rfloor+\gamma}\|\rangle_{\mathbf{x}^0} \leq \iota(\gamma)(\|\boldsymbol{\theta}^0\| + \sqrt{d}\|\widehat{\alpha}^0\| + \sqrt{d}) \leq C\iota(\gamma)\sqrt{d}.$$

Hence also

$$(II) \leq \iota(\gamma)d.$$

The proof of (171) is completed by combining the bounds for (I) and (II).

Discretization of $\mathbf{R}_\eta$. Let P^γ_t and P_t be the discrete and continuous Markov semigroups defined above. Let $x_i(\boldsymbol{\theta}, \widehat{\alpha}) = \mathbf{e}_i^\top \mathbf{X}\boldsymbol{\theta}$. Introduce the matrix $\mathbf{E} \in \mathbb{R}^{d\times(d+K)}$ defined by $\mathbf{E}(\boldsymbol{\theta}, \widehat{\alpha}) = \boldsymbol{\theta}$, so that this reads $x_i(\mathbf{x}) = \mathbf{e}_i^\top \mathbf{X}\mathbf{E}\mathbf{x}$ for $\mathbf{x} = (\boldsymbol{\theta}, \widehat{\alpha})$. Then for any $s, t \in \mathbb{Z}_+$ with $s < t$, Proposition A.4 gives

$$\partial_\varepsilon|_{\varepsilon=0}\langle \eta_i^{t,[s,i],\varepsilon}\rangle_{\mathbf{x}} = \gamma P^\gamma_{s+1}\mathbf{e}_i^\top \mathbf{X}\mathbf{E}\nabla P^\gamma_{t-s-1}x_i(\mathbf{x}).$$

Let us introduce the shorthand $P^\gamma_t(\mathbf{x}) = \langle \mathbf{x}^t_\gamma\rangle_{\mathbf{x}}$ as a map $P^\gamma_t : \mathbb{R}^{d+K} \to \mathbb{R}^{d+K}$, so $P^\gamma_t x_i(\mathbf{x}) = \mathbf{e}_i^\top \mathbf{X}\mathbf{E}P^\gamma_t(\mathbf{x})$. Denote also $\mathrm{d}P^\gamma_t(\cdot) : \mathbb{R}^{d+K} \to \mathbb{R}^{(d+K)\times(d+K)}$ as the derivative of this map $\mathbf{x} \mapsto P^\gamma_t(\mathbf{x})$. Then the above may be written as

$$\partial_\varepsilon|_{\varepsilon=0}\langle \eta_i^{t,[s,i],\varepsilon}\rangle_{\mathbf{x}} = \gamma P^\gamma_{s+1}\Big(\mathbf{e}_i^\top \mathbf{X}\mathbf{E}\mathrm{d}P^\gamma_{t-s-1}(\cdot)^\top \mathbf{E}^\top \mathbf{X}^\top \mathbf{e}_i\Big)(\mathbf{x}),$$

implying that

$$\gamma^{-1}\operatorname{Tr}\mathbf{R}^\gamma_\eta(t,s) = \delta\beta^2\sum_{i=1}^n P^\gamma_{s+1}\Big(\mathbf{e}_i^\top \mathbf{X}\mathbf{E}\mathrm{d}P^\gamma_{t-s-1}(\cdot)^\top \mathbf{E}^\top \mathbf{X}^\top \mathbf{e}_i\Big)(\mathbf{x}^0) = \delta\beta^2 P^\gamma_{s+1}\operatorname{Tr}\Big[\mathrm{d}P^\gamma_{t-s-1}(\cdot)\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E}\Big](\mathbf{x}^0).$$

By Proposition A.1, we have analogously for any $s, t \in \mathbb{R}_+$ with $s \leq t$ that

$$\operatorname{Tr}\mathbf{R}_\eta(t,s) = \delta\beta^2 P_s \operatorname{Tr}\big[\mathrm{d}P_{t-s}(\cdot)\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E}\big](\mathbf{x}^0).$$

Hence for all $s, t \in \mathbb{R}_+$ with $s \leq t$,

$$\begin{aligned}&\Big|\operatorname{Tr}\mathbf{R}_\eta(t,s) - \gamma^{-1}\operatorname{Tr}\mathbf{R}^\gamma_\eta([t]+1,[s])\Big| \\ &= \delta\beta^2\Big[\underbrace{\Big|P_s\operatorname{Tr}\Big[\Big(\mathrm{d}P_{t-s}(\cdot) - \mathrm{d}P^\gamma_{[t]-[s]}(\cdot)\Big)\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E}\Big](\mathbf{x}^0)\Big|}_{(I)} + \underbrace{\Big|(P_s - P^\gamma_{[s]+1})\operatorname{Tr}\Big[\mathrm{d}P^\gamma_{[t]-[s]}(\cdot)\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E}\Big](\mathbf{x}^0)\Big|}_{(II)}\Big].\end{aligned}$$

Bound of (I). Note that by Proposition A.2(c), $\operatorname{Tr}\mathrm{d}P_{t-s}(\mathbf{x})\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E} = \langle \operatorname{Tr}\mathbf{V}^{t-s}\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E}\rangle_{\mathbf{x}} = \langle \operatorname{Tr}\mathbf{U}^{t-s}\cdot \mathbf{X}^\top \mathbf{X}\rangle_{\mathbf{x}}$, where $\{\mathbf{x}^t, \mathbf{V}^t\}_{t\geq 0}$ follow the dynamics (160) and $\mathbf{U}^t$ as before is the upper-left block of $\mathbf{V}^t$. Similarly, Lemma A.5 yields that $\operatorname{Tr}\mathrm{d}P^\gamma_{[t]-[s]}(\mathbf{x})\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E} = \langle \operatorname{Tr}\bar{\mathbf{V}}_\gamma^{\lfloor t\rfloor-\lfloor s\rfloor}\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E}\rangle_{\mathbf{x}} = \langle \operatorname{Tr}\bar{\mathbf{U}}_\gamma^{\lfloor t\rfloor-\lfloor s\rfloor}\mathbf{X}^\top \mathbf{X}\rangle_{\mathbf{x}}$, where $\{\bar{\mathbf{x}}^t_\gamma, \bar{\mathbf{V}}^t_\gamma\}_{t\geq 0}$ follow (161). Hence, with $\mathbf{x} = (\boldsymbol{\theta}, \widehat{\alpha})$,

$$\begin{aligned}\Big|\operatorname{Tr}\Big(\mathrm{d}P_{t-s}(\mathbf{x}) - \mathrm{d}P^\gamma_{[t]-[s]}(\mathbf{x})\Big)\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E}\Big| &= \langle \operatorname{Tr}(\mathbf{U}^{t-s} - \bar{\mathbf{U}}_\gamma^{\lfloor t\rfloor-\lfloor s\rfloor})\mathbf{X}^\top \mathbf{X}\rangle_{\mathbf{x}} \\ &\leq \sqrt{d}\|\mathbf{X}\|^2_{\mathrm{op}}\langle \|\mathbf{U}^{t-s} - \bar{\mathbf{U}}_\gamma^{\lfloor t\rfloor-\lfloor s\rfloor}\|_F\rangle_{\mathbf{x}} \\ &\leq \iota(\gamma)(\sqrt{d}\|\boldsymbol{\theta}\| + d\|\widehat{\alpha}\| + d)\end{aligned}$$

using the preceding bounds leading to (176). Then applying this with $\mathbf{x} = \mathbf{x}^s$ shows $(I) \leq \iota(\gamma)d$.

Bound of (II). Let $\mathsf{f}(\mathbf{x}) = \operatorname{Tr} \mathrm{d}P^{\gamma}_{[t]-[s]}(\mathbf{x})\mathbf{E}^\top \mathbf{X}^\top \mathbf{X}\mathbf{E} = \operatorname{Tr}\langle \bar{\mathbf{U}}^{\tau}_{\gamma}\rangle_{\mathbf{x}} \mathbf{X}^\top \mathbf{X}$, where $\tau = \lfloor t \rfloor - \lfloor s \rfloor$. By the same arguments as above,

$$|\mathsf{f}(\mathbf{x}) - \mathsf{f}(\tilde{\mathbf{x}})| \leq C\sqrt{d}(\|\boldsymbol{\theta} - \tilde{\boldsymbol{\theta}}\| + \sqrt{d}\|\widehat{\alpha} - \tilde{\widehat{\alpha}}\|),$$

leading to $(II) \leq \iota(\gamma)d$. Combining these bounds for (I) and (II) shows (172). □

We now conclude the proof of Theorem 2.8.

Proof of Theorem 2.8. The claims for $d^{-1} \operatorname{Tr} \mathbf{C}_\theta(t,s)$, $d^{-1} \operatorname{Tr} \mathbf{C}_\theta(t,*)$, and $n^{-1} \operatorname{Tr} \mathbf{C}_\eta(t,s)$ follow immediately from the definitions of these quantities, Corollary 2.6 applied with $f_\theta(\theta^s, \theta^t) = \theta^s\theta^t$, $f_\theta(\theta^*, \theta^t) = \theta^*\theta^t$, $f_\eta(\eta^*, \varepsilon, \eta^s, \eta^t) = \delta\beta^2(\eta^s - \eta^* - \varepsilon)(\eta^t - \eta^* - \varepsilon)$, and an application of the dominated convergence theorem to take expectations over $\{\mathbf{b}^t\}_{t\in[0,T]}$ in the almost-sure convergence statements of Corollary 2.6.

For the claim for $d^{-1} \operatorname{Tr} \mathbf{R}_\theta(t,s)$, for any $s,t \in [0,T]$ with $s \leq t$, by Lemma 5.7, almost surely

$$\limsup_{n,d\to\infty} \left| \frac{1}{d} \operatorname{Tr} \mathbf{R}_\theta(t,s) - \frac{1}{\gamma} \cdot \frac{1}{d} \operatorname{Tr} \mathbf{R}^{\gamma}_{\theta}([t]+1,[s]) \right| \leq \iota(\gamma).$$

By Lemma 5.1 and the identification (99) of Lemma 4.3, almost surely

$$\lim_{n,d\to\infty} \frac{1}{\gamma} \cdot \frac{1}{d} \operatorname{Tr} \mathbf{R}^{\gamma}_{\theta}([t]+1,[s]) = \frac{1}{\gamma} R^{\gamma}_{\theta}([t]+1,[s]) = \bar{R}^{\gamma}_{\theta}(t+\gamma, s).$$

The bound (104) implies uniform convergence of $\bar{R}^{\gamma}_{\theta}(t,s)$ to $R^{\gamma}_{\theta}(t,s)$ as $\gamma \to 0$, and $R^{\gamma}_{\theta}(t,s)$ is continuous in s,t by Theorem 2.4 and the definition of the space $\mathcal{S}^{\text{cont}}$. Thus

$$\lim_{\gamma\to 0} |\bar{R}^{\gamma}_{\theta}(t+\gamma,s) - R_\theta(t,s)| = 0.$$

Then, taking the limit $n,d \to \infty$ followed by $\gamma \to 0$ shows almost surely

$$\lim_{n,d\to\infty} \left| \frac{1}{d} \operatorname{Tr} \mathbf{R}_\theta(t,s) - R_\theta(t,s) \right| = 0.$$

The proof of the claim for $n^{-1} \operatorname{Tr} \mathbf{R}_\eta(t,s)$ is the same. □

A Existence of linear response functions

A.1 Continuous dynamics

Fix any dimension $m \geq 1$, and consider the function classes

$$\mathcal{A} = \Big\{ f : \mathbb{R}^m \to \mathbb{R} \text{ twice continuously-differentiable} : \nabla f(\mathbf{x}), \nabla^2 f(\mathbf{x}) \text{ are globally bounded} \Big\},$$
$$\mathcal{B} = \Big\{ f : \mathbb{R}^m \to \mathbb{R}^m \text{ twice continuously-differentiable} :$$
$$\nabla f_i(\mathbf{x}), \nabla^2 f_i(\mathbf{x}) \text{ are globally bounded and Hölder-continuous for each } i = 1,\ldots,m \Big\}.$$

We consider a general stochastic diffusion over $\mathbf{x}^t \in \mathbb{R}^m$ given by

$$\mathrm{d}\mathbf{x}^t = u(\mathbf{x}^t)\mathrm{d}t + \sqrt{2}\,\mathbf{M}\,\mathrm{d}\mathbf{b}^t \tag{178}$$

where $\mathbf{b}^t \in \mathbb{R}^m$ is a standard Brownian motion, $u(\cdot)$ a Lipschitz drift function, and $\mathbf{M} \in \mathbb{R}^{m\times m}$ a deterministic diffusion coefficient matrix. We note that the joint evolution of $\mathbf{x}^t = (\boldsymbol{\theta}^t, \widehat{\alpha}^t)$ in (4–5) is of this form, with $m = d + K$ and with $u(\cdot)$ and $\mathbf{M}$ as defined in (159). The conditions of Theorem 2.8 ensure that this drift function $u(\cdot)$ satisfies $u \in \mathcal{B}$.

We prove in this section the following result:

Proposition A.1. *Suppose $u \in \mathcal{B}$, and let $\{\mathbf{x}^t\}_{t\geq 0}$ be the solution of (178) with initial condition $\mathbf{x}^0 = \mathbf{x}$. For any $a \in \mathcal{A}$, $b \in \mathcal{B}$, and $\mathbf{x} \in \mathbb{R}^m$, define*

$$R(t,s) = P_s\big(b^\top \nabla P_{t-s} a\big)(\mathbf{x}) \tag{179}$$

where $P_t f(\mathbf{x}) = \mathbb{E}[f(\mathbf{x}^t) \mid \mathbf{x}^0 = \mathbf{x}]$. Then $\{R(t,s)\}_{0\leq s\leq t}$ is the unique continuous function for which the following holds:

Let $h : [0,\infty) \to \mathbb{R}$ be any continuous bounded function, and for each $\varepsilon > 0$ let $\{\mathbf{x}^{t,\varepsilon}\}_{t\geq 0}$ be the solution of the perturbed dynamics

$$\mathrm{d}\mathbf{x}^{t,\varepsilon} = \Big(u(\mathbf{x}^{t,\varepsilon}) + \varepsilon h(t) b(\mathbf{x}^{t,\varepsilon})\Big)\mathrm{d}t + \sqrt{2}\,\mathbf{M}\,\mathrm{d}\mathbf{b}^t \tag{180}$$

with the same initial condition $\mathbf{x}^{0,\varepsilon} = \mathbf{x}$. Then for any $t > 0$,

$$\lim_{\varepsilon\to 0} \frac{1}{\varepsilon}\Big(\mathbb{E}[a(\mathbf{x}^{t,\varepsilon}) \mid \mathbf{x}^{0,\varepsilon} = \mathbf{x}] - \mathbb{E}[a(\mathbf{x}^t) \mid \mathbf{x}^0 = \mathbf{x}]\Big) = \int_0^t R(t,s)h(s)\mathrm{d}s.$$

Statements similar to Proposition A.1 have been established in [55, 56]. Our setting here is somewhat non-standard, in that $\mathbf{M}$ may be rank-degenerate, so the PDE describing the law of $\{\mathbf{x}^t\}_{t\geq 0}$ is not uniformly elliptic. We show Proposition A.1 in two steps, first deriving regularity estimates for the Markov semigroup $\{P_t\}_{t\geq 0}$ in such settings using the results of [63], and then applying the proof idea of [56, Theorem 3.9] with these regularity estimates in place of the Schauder estimates derived therein from uniform ellipticity.

We will write

$$P_t f(\mathbf{x}) = \langle f(\mathbf{x}^t)\rangle_{\mathbf{x}^0=\mathbf{x}} = \mathbb{E}[f(\mathbf{x}^t) \mid \mathbf{x}^0 = \mathbf{x}] \tag{181}$$

for the Markov semigroup associated to (178). When the initial condition $\mathbf{x}^0 = \mathbf{x}$ is clear from context, we will abbreviate $\langle f(\mathbf{x}^t)\rangle = \langle f(\mathbf{x}^t)\rangle_{\mathbf{x}^0=\mathbf{x}}$. We denote the infinitesimal generator L of this semigroup by

$$\mathrm{L}f(\mathbf{x}) = u(\mathbf{x})^\top \nabla f(\mathbf{x}) + \operatorname{Tr} \mathbf{M}\mathbf{M}^\top \nabla^2 f(\mathbf{x}). \tag{182}$$

Throughout this section, constants $C, C', c > 0$ may depend on the dimension m and the functions u, a, b.

Proposition A.2. *Suppose the assumptions of Proposition A.1 hold. Let $u_i : \mathbb{R}^m \to \mathbb{R}$ be the i^{th} coordinate of u, and let $\partial_j u_i$ and $\partial_j\partial_k u_i$ be its first-order and second-order partial derivatives.*

(a) For each $\mathbf{x} \in \mathbb{R}^m$, the diffusion (178) has a unique solution $\{\mathbf{x}^t\}_{t\geq 0}$ with initial condition $\mathbf{x}^0 = \mathbf{x}$. Furthermore, there exists a modification $\mathbf{x}^t(\mathbf{x})$ of this solution for each initial condition $\mathbf{x}^0 = \mathbf{x}$ such that $\mathbf{x}^t(\mathbf{x})$ is jointly continuous in $(t,\mathbf{x})$ and twice continuously-differentiable in $\mathbf{x}$.

(b) For every $i = 1,\ldots,m$, let $x_i^t(\mathbf{x})$ be the i^{th} coordinate of $\mathbf{x}^t(\mathbf{x})$, and let $\mathbf{v}_i^t(\mathbf{x}) = \nabla x_i^t(\mathbf{x}) \in \mathbb{R}^m$ and $\mathbf{H}_i^t(\mathbf{x}) = \nabla^2 x_i^t(\mathbf{x}) \in \mathbb{R}^{m\times m}$ be its gradient and Hessian in $\mathbf{x}$. Then $(\mathbf{v}_i^t(\mathbf{x}), \mathbf{H}_i^t(\mathbf{x}))$ are solutions to the first and second variation processes

$$\begin{cases} \mathrm{d}\mathbf{v}_i^t = \sum_{j=1}^m \partial_j u_i(\mathbf{x}^t(\mathbf{x}))\cdot \mathbf{v}_j^t\,\mathrm{d}t \\ \mathrm{d}\mathbf{H}_i^t = \big(\sum_{j,k=1}^m \partial_j\partial_k u_i(\mathbf{x}^t(\mathbf{x}))\cdot \mathbf{v}_j^t\mathbf{v}_k^{t\top} + \sum_{j=1}^m \partial_j u_i(\mathbf{x}^t(\mathbf{x}))\cdot \mathbf{H}_j^t\big)\mathrm{d}t \end{cases} \tag{183}$$

with initial conditions $\mathbf{v}_i^0(\mathbf{x}) = \mathbf{e}_i$ (the i^{th} standard basis vector in $\mathbb{R}^m$) and $\mathbf{H}_i^0(\mathbf{x}) = 0$.

Furthermore $\|\mathbf{v}_i^t(\mathbf{x})\|_2, \|\mathbf{H}_i^t(\mathbf{x})\|_{\mathrm{op}} \leq e^{Ct}$ for some $C > 0$ and all $\mathbf{x} \in \mathbb{R}^m$ and $t \geq 0$.

(c) For any $f \in \mathcal{A}$, the map $(t,\mathbf{x}) \mapsto P_t f(\mathbf{x})$ is continuously-differentiable in t and twice continuously-differentiable in $\mathbf{x}$, and furthermore $\nabla P_t f(\mathbf{x}), \nabla^2 P_t f(\mathbf{x})$ are uniformly bounded over $t \in [0,T]$ and $\mathbf{x} \in \mathbb{R}^m$ for any fixed $T > 0$. For any $t \geq 0$ and initial condition $\mathbf{x}^0 = \mathbf{x}$, letting $(\mathbf{x}^t, \mathbf{v}_i^t, \mathbf{H}_i^t) \equiv (\mathbf{x}^t(\mathbf{x}), \mathbf{v}_i^t(\mathbf{x}), \mathbf{H}_i^t(\mathbf{x}))$ be as defined in parts (a) and (b), we have

$$\begin{aligned} \nabla P_t f(\mathbf{x}) &= \Big\langle \sum_{j=1}^m \partial_j f(\mathbf{x}^t)\mathbf{v}_j^t \Big\rangle \\ \nabla^2 P_t f(\mathbf{x}) &= \Big\langle \sum_{j,k=1}^m \partial_j\partial_k f(\mathbf{x}^t)\mathbf{v}_j^t\mathbf{v}_k^{t\top} + \sum_{j=1}^m \partial_j f(\mathbf{x}^t)\mathbf{H}_j^t \Big\rangle \end{aligned} \tag{184}$$

and

$$\partial_t P_t f(\mathbf{x}) = P_t \mathrm{L} f(\mathbf{x}) = \mathrm{L} P_t f(\mathbf{x}). \tag{185}$$

Proof. Since the coordinates of $u \in \mathcal{B}$ are Lipschitz with bounded and Hölder-continuous first and second derivatives, part (a) follows directly from [63, Theorems II.1.2, II.3.3].

For part (b), since $u \in \mathcal{B}$ has bounded and Hölder-continuous first derivative, [63, Theorem II.3.1] shows that $\mathbf{x}^t(\mathbf{x})$ has derivative $\mathbf{V}^t(\mathbf{x}) = \mathrm{d}_{\mathbf{x}}\mathbf{x}^t \in \mathbb{R}^{m\times m}$ solving the first-variation equation

$$\mathrm{d}\mathbf{V}^t = [\mathrm{d}u(\mathbf{x}^t)]\mathbf{V}^t\mathrm{d}t, \qquad \mathbf{V}^0 = \mathbf{I}.$$

Noting that $\mathbf{v}_i^t = \nabla x_i^t(\mathbf{x})$ is (the transpose of) the i^{th} row of $\mathbf{V}^t$, this gives the first equation of (183) with initial condition $\mathbf{v}_i^0 = \mathbf{e}_i$. Next, consider the joint diffusion

$$\mathrm{d}(\mathbf{x}^t, \mathbf{V}^t) = P(\mathbf{x}^t, \mathbf{V}^t)\mathrm{d}t + \sqrt{2}(\mathbf{M}\,\mathrm{d}\mathbf{b}^t, 0), \qquad P(\mathbf{x}, \mathbf{V}) = \Big(u(\mathbf{x}), [\mathrm{d}u(\mathbf{x})]\mathbf{V}\Big).$$

The condition $u \in \mathcal{B}$ implies also that $P(\mathbf{x}, \mathbf{V})$ has bounded and Hölder-continuous first derivative $\mathrm{d}P(\mathbf{x}, \mathbf{V})$, which we identify as a square matrix of dimension $(m + m^2) \times (m + m^2)$ under the vectorization of $\mathbf{V}$. Then [63, Theorem II.3.1] applied again shows that $(\mathbf{x}^t(\mathbf{x}), \mathbf{V}^t(\mathbf{x}))$ has derivative $\mathbf{U}^t = \mathrm{d}_{(\mathbf{x},\mathbf{V})}(\mathbf{x}^t, \mathbf{V}^t) \in \mathbb{R}^{(m+m^2)\times(m+m^2)}$ solving the second-variation equation

$$\mathrm{d}\mathbf{U}^t = [\mathrm{d}P(\mathbf{x}^t, \mathbf{V}^t)]\mathbf{U}^t\mathrm{d}t, \qquad \mathbf{U}^0 = \mathbf{I}. \tag{186}$$

Noting that $\mathbf{H}_i^t = \nabla^2 x_i^t(\mathbf{x})$ is the block of $\mathbf{U}^t$ corresponding to $\mathrm{d}_{\mathbf{x}}\mathbf{v}_i^t$, and that the block corresponding to $\mathrm{d}_{\mathbf{x}}\mathbf{x}^t$ is $\mathbf{V}^t$, one may check that the restriction of (186) to the $\mathrm{d}_{\mathbf{x}}\mathbf{v}_i^t$ block gives exactly the second equation of (183) with initialization $\mathbf{H}_i^0 = 0$. If $C > 0$ is an upper bound for $\sup_{\mathbf{x}\in\mathbb{R}^m} \|\mathrm{d}u(\mathbf{x})\|_{\text{op}}$ and $\sup_{\mathbf{x}\in\mathbb{R}^m} \|\mathrm{d}P(\mathbf{x}, \mathbf{V})\|_{\text{op}}$, then integrating these equations gives $\|\mathbf{V}^t\|_{\text{op}} \le e^{Ct}\|\mathbf{V}^0\|_{\text{op}} = e^{Ct}$ and $\|\mathbf{U}^t\|_{\text{op}} \le e^{Ct}\|\mathbf{U}^0\|_{\text{op}} = e^{Ct}$, which implies the bounds for $\mathbf{v}_i^t$ and $\mathbf{H}_i^t$.

For part (c), consider any $f \in \mathcal{A}$. Applying (b) and the chain rule,

$$\begin{aligned}\nabla_{\mathbf{x}} f(\mathbf{x}^t(\mathbf{x})) &= \sum_{j=1}^m \partial_j f(\mathbf{x}^t)\mathbf{v}_i^t \\ \nabla^2_{\mathbf{x}} f(\mathbf{x}^t(\mathbf{x})) &= \sum_{j,k=1}^m \partial_j\partial_k f(\mathbf{x}^t)\mathbf{v}_j^t\mathbf{v}_k^{t\top} + \sum_{j=1}^m \partial_j f(\mathbf{x}^t)\mathbf{H}_j^t.\end{aligned} \tag{187}$$

By parts (a–b) and the condition $f \in \mathcal{A}$, for any $T > 0$, the right sides of (187) are uniformly bounded and continuous in $(t, \mathbf{x})$ over $t \in [0, T]$. Then dominated convergence implies that $P_t f(\mathbf{x})$ is twice continuously-differentiable in $\mathbf{x}$, that $\nabla P_t f(\mathbf{x}) = \nabla_{\mathbf{x}}\langle f(\mathbf{x}^t)\rangle_{\mathbf{x}^0=\mathbf{x}} = \langle \nabla_{\mathbf{x}} f(\mathbf{x}^t(\mathbf{x}))\rangle$ and $\nabla^2 P_t f(\mathbf{x}) = \nabla^2_{\mathbf{x}}\langle f(\mathbf{x}^t)\rangle_{\mathbf{x}^0=\mathbf{x}} = \langle \nabla^2_{\mathbf{x}} f(\mathbf{x}^t(\mathbf{x}))\rangle$, and that these are also uniformly bounded and continuous over $t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^m$.

For the derivative in t, by Itô's formula

$$f(\mathbf{x}^t) = f(\mathbf{x}) + \int_0^t \mathrm{L}f(\mathbf{x}^s)\mathrm{d}s + \int_0^t \nabla f(\mathbf{x}^s)^\top \sqrt{2}\,\mathbf{M}\,\mathrm{d}\mathbf{b}^s$$

where L is the generator defined in (182). Since $\nabla f(\mathbf{x}^s)$ is bounded over $s \in [0, t]$ and $\mathbf{x}^s$ is adapted to the filtration of $\{\mathbf{b}^s\}$, the last term is a martingale, so taking expectations gives

$$P_t f(\mathbf{x}) = \langle f(\mathbf{x}^t)\rangle = f(\mathbf{x}) + \int_0^t \langle \mathrm{L}f(\mathbf{x}^s)\rangle \mathrm{d}s.$$

Hence, differentiating in t, for any $t > 0$ we have

$$\partial_t P_t f(\mathbf{x}) = \langle \mathrm{L}f(\mathbf{x}^t)\rangle = P_t \mathrm{L}f(\mathbf{x}). \tag{188}$$

By Jensen's inequality, for any $s, t \ge 0$, we have

$$\langle (P_s\mathrm{L}f(\mathbf{x}^t))^2\rangle \le \langle \mathrm{L}f(\mathbf{x}^{t+s})^2\rangle = \langle (u(\mathbf{x}^{t+s})^\top \nabla f(\mathbf{x}^{t+s}) + \operatorname{Tr}\mathbf{M}\mathbf{M}^\top\nabla^2 f(\mathbf{x}^{t+s}))^2\rangle \le C(1 + \langle \|\mathbf{x}^{t+s}\|_2^2\rangle),$$

the last inequality holding for some $C > 0$ by boundedness of $\nabla f, \nabla^2 f$ and the Lipschitz continuity of u. Then [63, Theorem II.2.1] implies that $P_s\mathrm{L}f(\mathbf{x}^t(\mathbf{x}))$ is uniformly bounded in L^2 over compact domains of $s, t \ge 0$

and of the initial condition $\mathbf{x} \in \mathbb{R}^m$, and hence is also uniformly integrable over these domains. This uniform integrability for $s = 0$ and dominated convergence shows that $\langle \mathrm{L}f(\mathbf{x}^t)\rangle$ in (188) is continuous in $(t, \mathbf{x})$, and hence $P_t f$ is continuously-differentiable in t. Taking the limit $t \to 0$ in (188), also $\mathrm{L}f(\mathbf{x}) = \lim_{t\to 0} \partial_t P_t f(\mathbf{x})$. Then applying this with $P_t f \in \mathcal{A}$ in place of f,

$$\mathrm{L}P_t f(\mathbf{x}) = \lim_{s\to 0} \partial_s P_{t+s} f(\mathbf{x}) = \lim_{s\to 0} \partial_s \langle P_s f(\mathbf{x}^t)\rangle \stackrel{(*)}{=} \langle \mathrm{L}f(\mathbf{x}^t)\rangle = P_t \mathrm{L}f(\mathbf{x}).$$

Here, to justify $(*)$, we note that $\partial_s P_s f(\mathbf{x}^t) = P_s \mathrm{L}f(\mathbf{x}^t)$ by (188), so $(*)$ follows from uniform integrability of this quantity and dominated convergence to take the limit $\lim_{s\to 0} \partial_s \langle P_s f(\mathbf{x}^t)\rangle = \lim_{s\to 0}\langle P_s \mathrm{L}f(\mathbf{x}^t)\rangle = \langle \mathrm{L}f(\mathbf{x}^t)\rangle$. Combining with (188), this shows all claims about $\partial_t P_t f$ in part (c). □

Now consider the perturbed dynamics (180) for any $\varepsilon > 0$. Let us denote the perturbed drift as

$$u^\varepsilon(t, \mathbf{x}) = u(\mathbf{x}) + \varepsilon h(t) b(\mathbf{x}).$$

For any $t \geq s \geq 0$, we define its (time inhomogeneous) Markov semigroup and infinitesimal generator

$$P^\varepsilon_{s,t} f(\mathbf{x}) = \langle f(\mathbf{x}^t)\rangle_{\mathbf{x}^s = \mathbf{x}} = \mathbb{E}[f(\mathbf{x}^t) \mid \mathbf{x}^s = \mathbf{x}], \qquad \mathrm{L}^\varepsilon_t f(\mathbf{x}) = u^\varepsilon(t, \mathbf{x})^\top \nabla f(\mathbf{x}) + \operatorname{Tr} \mathbf{M}\mathbf{M}^\top \nabla^2 f(\mathbf{x}).$$

The following extends the semigroup regularity estimates of Proposition A.2 to this perturbed process.

Proposition A.3. *Suppose the assumptions of Proposition A.1 hold. Then for any $f \in \mathcal{A}$, the map $(s, t, \mathbf{x}) \mapsto P^\varepsilon_{s,t} f(\mathbf{x})$ is continuously-differentiable in (s, t) and twice continuously-differentiable in $\mathbf{x}$, and furthermore $\nabla P^\varepsilon_{s,t} f(\mathbf{x}), \nabla^2 P^\varepsilon_{s,t} f(\mathbf{x})$ are uniformly bounded over $s, t \in [0, T]$ and $\mathbf{x} \in \mathbb{R}^m$ for any fixed $T > 0$. We have*

$$\partial_t P^\varepsilon_{s,t} f(\mathbf{x}) = P^\varepsilon_{s,t} \mathrm{L}^\varepsilon_t f(\mathbf{x}), \qquad \partial_s P^\varepsilon_{s,t} f(\mathbf{x}) = -\mathrm{L}^\varepsilon_s P^\varepsilon_{s,t} f(\mathbf{x}). \tag{189}$$

Proof. We omit the superscript ε and write $\mathbf{x}^t \equiv \mathbf{x}^{t,\varepsilon}$. The same arguments as in Proposition A.2 using [63, Theorems II.1.2, II.3.1, II.3.3] show, for each $s \geq 0$ and $\mathbf{x} \in \mathbb{R}^m$, there exists a modification $\{\mathbf{x}^t(s, \mathbf{x})\}_{t\geq s}$ of the solution to (180) with initial condition $\mathbf{x}^s = \mathbf{x}$, such that $\mathbf{x}^t(s, \mathbf{x})$ is jointly continuous in $(s, t, \mathbf{x})$ and twice continuously-differentiable in $\mathbf{x}$. Each component $x^t_i(s, \mathbf{x})$ of this solution has gradient $\mathbf{v}^t_i = \nabla_\mathbf{x} x^t_i(s, \mathbf{x})$ and Hessian $\mathbf{H}^t_i = \nabla^2_\mathbf{x} x^t_i(s, \mathbf{x})$ solving

$$\begin{cases} \mathrm{d}\mathbf{v}^t_i = \sum_{j=1}^m \partial_j u^\varepsilon_i(t, \mathbf{x}^t(s, \mathbf{x})) \cdot \mathbf{v}^t_j \, \mathrm{d}t \\ \mathrm{d}\mathbf{H}^t_i = \big(\sum_{j,k=1}^m \partial_j \partial_k u^\varepsilon_i(t, \mathbf{x}^t(s, \mathbf{x})) \cdot \mathbf{v}^t_j {\mathbf{v}^t_k}^\top + \sum_{j=1}^m \partial_j u^\varepsilon_i(t, \mathbf{x}^t(s, \mathbf{x})) \cdot \mathbf{H}^t_j\big) \mathrm{d}t \end{cases}$$

with initial conditions $\mathbf{v}^s_i(s, \mathbf{x}) = \mathbf{e}_i$ and $\mathbf{H}^s_i(s, \mathbf{x}) = 0$. Furthermore, $\|\mathbf{v}^t_i(s, \mathbf{x})\|_2, \|\mathbf{H}^t_i(s, \mathbf{x})\|_{\mathrm{op}} \leq e^{C(t-s)}$ for some $C > 0$ and all $\mathbf{x} \in \mathbb{R}^m$ and $t \geq s \geq 0$.

Then for any $f \in \mathcal{A}$, the same dominated convergence argument as in Proposition A.2 shows that $P^\varepsilon_{s,t} f(\mathbf{x})$ is twice continuously-differentiable in $\mathbf{x}$, where its first and second derivatives are uniformly bounded and continuous in $(s, t, \mathbf{x})$ over $s, t \in [0, T]$ and may be computed by differentiating in $\mathbf{x}$ under the integral. The same argument as in Proposition A.2 using Itô's formula shows also that $P^\varepsilon_{s,t} f(\mathbf{x})$ is continuously-differentiable in t, with

$$\partial_t P^\varepsilon_{s,t} f(\mathbf{x}) = P^\varepsilon_{s,t} \mathrm{L}^\varepsilon_t f(\mathbf{x}) = \langle \mathrm{L}^\varepsilon_t f(\mathbf{x}^t)\rangle_{\mathbf{x}^s = \mathbf{x}}.$$

For the derivative in s, we have by Itô's formula for any $h > 0$,

$$P^\varepsilon_{s-h,s} f(\mathbf{x}) = \langle f(\mathbf{x}^s)\rangle_{\mathbf{x}^{s-h} = \mathbf{x}} = f(\mathbf{x}) + \int_{s-h}^s \langle \mathrm{L}^\varepsilon_r f(\mathbf{x}^r)\rangle_{\mathbf{x}^{s-h} = \mathbf{x}} \, \mathrm{d}r.$$

The same argument as in Proposition A.2 shows that $\mathrm{L}^\varepsilon_t f(\mathbf{x}^t(s, \mathbf{x}))$ is uniformly integrable over compact domains of $t \geq s \geq 0$ and of $\mathbf{x} \in \mathbb{R}^m$, so by dominated convergence we have $\lim_{h\downarrow 0,\, r\uparrow s} \langle \mathrm{L}^\varepsilon_r f(\mathbf{x}^r)\rangle_{\mathbf{x}^{s-h} = \mathbf{x}} = \mathrm{L}^\varepsilon_s(\mathbf{x})$. So taking the limit $h \to 0$ above and rearranging shows

$$\mathrm{L}^\varepsilon_s f(\mathbf{x}) = \lim_{h\downarrow 0} \frac{P^\varepsilon_{s-h,s} f(\mathbf{x}) - f(\mathbf{x})}{h}. \tag{190}$$

Then for any $s \leq t$, applying this to $P^\varepsilon_{s,t} f \in \mathcal{A}$ in place of f gives

$$\lim_{h\downarrow 0} \frac{P^\varepsilon_{s,t}f(\mathbf{x}) - P^\varepsilon_{s-h,t}f(\mathbf{x})}{h} = \lim_{h\downarrow 0} \frac{P^\varepsilon_{s,t}f(\mathbf{x}) - P^\varepsilon_{s-h,s}(P^\varepsilon_{s,t}f)(\mathbf{x})}{h} = -\mathrm{L}^\varepsilon_s P^\varepsilon_{s,t} f(\mathbf{x}),$$

i.e. $P^\varepsilon_{s,t}f(\mathbf{x})$ is left-differentiable in s. Here $-\mathrm{L}^\varepsilon_s P^\varepsilon_{s,t}f(\mathbf{x}) = -u^\varepsilon(s,\mathbf{x})^\top \nabla P^\varepsilon_{s,t}f(\mathbf{x}) - \operatorname{Tr}\mathbf{M}\mathbf{M}^\top \nabla^2 P^\varepsilon_{s,t}f(\mathbf{x})$ is continuous in $(s,t,\mathbf{x})$ by the continuity of $\nabla P^\varepsilon_{s,t}f$ and $\nabla^2 P^\varepsilon_{s,t}f$ argued above. Then $P^\varepsilon_{s,t}f(\mathbf{x})$ is also continuously-differentiable in s with $\partial_s P^\varepsilon_{s,t}f(\mathbf{x}) = -\mathrm{L}^\varepsilon_s P^\varepsilon_{s,t}f(\mathbf{x})$. □

Proof of Proposition A.1. Let $\{\mathbf{x}^t\}_{t\geq 0}$ and $\{\mathbf{x}^{t,\varepsilon}\}_{t\geq 0}$ be the solutions to the unperturbed and perturbed diffusions. Let $\{P_t\}$ and L be the semigroup and infinitesimal generator for $\{\mathbf{x}^t\}_{t\geq 0}$, and let $\{P^\varepsilon_{s,t}\}$ and L^ε_t be those for $\{\mathbf{x}^{t,\varepsilon}\}_{t\geq 0}$. We write ∂_s, ∂_t for the derivatives in s,t and reserve $\nabla f(t,\mathbf{x})$ for the gradient of f in its second argument $\mathbf{x}$.

For any $t > s$ and $r \in [s,t]$, define $f^\varepsilon(r,\mathbf{x}) = P^\varepsilon_{r,t}a(\mathbf{x})$. Then by Itô's formula applied to the unperturbed process $\{\mathbf{x}^t\}_{t\geq 0}$,

$$f^\varepsilon(t,\mathbf{x}^t) = f^\varepsilon(s,\mathbf{x}^s) + \int_s^t (\partial_r + \mathrm{L}) f^\varepsilon(r,\mathbf{x}^r)\mathrm{d}r + \int_s^t \nabla f^\varepsilon(r,\mathbf{x}^r)^\top \sqrt{2}\,\mathbf{M}\mathrm{d}\mathbf{b}^r.$$

Proposition A.3 shows $P^\varepsilon_{r,t}a \in \mathcal{A}$, so $\nabla f^\varepsilon(r,\mathbf{x}^r)$ is uniformly bounded and the last term is a martingale. Then, taking expectations under the initial condition $\mathbf{x}^s = \mathbf{x}$ and applying (189),

$$\begin{aligned}
\langle a(\mathbf{x}^t)\rangle_{\mathbf{x}^s=\mathbf{x}} = \langle f^\varepsilon(t,\mathbf{x}^t)\rangle_{\mathbf{x}^s=\mathbf{x}} &= f^\varepsilon(s,\mathbf{x}) + \int_s^t \langle (\partial_r + \mathrm{L}) f^\varepsilon(r,\mathbf{x}^r)\rangle_{\mathbf{x}^s=\mathbf{x}}\mathrm{d}r \\
&= P^\varepsilon_{s,t}a(\mathbf{x}) + \int_s^t \langle (-\mathrm{L}^\varepsilon_r + \mathrm{L}) P^\varepsilon_{r,t}a(\mathbf{x}^r)\rangle_{\mathbf{x}^s=\mathbf{x}}\mathrm{d}r \\
&= P^\varepsilon_{s,t}a(\mathbf{x}) - \int_s^t \varepsilon h(r)\langle (b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x}^r)\rangle_{\mathbf{x}^s=\mathbf{x}}\mathrm{d}r \\
&= P^\varepsilon_{s,t}a(\mathbf{x}) - \varepsilon \int_s^t h(r) P_{r-s}(b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x})\mathrm{d}r.
\end{aligned}$$

Applying this also with $\varepsilon = 0$ and $P^0_{s,t} = P_{t-s}$ and taking the difference, we obtain the identity

$$P^\varepsilon_{s,t}a(\mathbf{x}) - P_{t-s}a(\mathbf{x}) = \varepsilon \int_s^t h(r) P_{r-s}(b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x})\mathrm{d}r. \tag{191}$$

From the definition of $P_t f(\mathbf{x})$ and form of $\nabla P_t f(\mathbf{x})$ in (184), we have

$$P_{r-s}(b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x}) = \left\langle (b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x}^{r-s})\right\rangle_{\mathbf{x}_0=\mathbf{x}}, \tag{192}$$

$$\nabla P_{r-s}(b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x}) = \left\langle \sum_{i=1}^m \partial_{x_i}[b^\top \nabla P^\varepsilon_{r,t}a](\mathbf{x}^{r-s})\mathbf{v}_i^{r-s}\right\rangle_{\mathbf{x}_0=\mathbf{x}}. \tag{193}$$

Since $b \in \mathcal{B}$ is Lipschitz by assumption, and $P^\varepsilon_{r,t}a \in \mathcal{A}$ by Proposition A.3, we have

$$|(b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x}^{r-s})|,\ |\partial_{x_i}[b^\top \nabla P^\varepsilon_{r,t}a](\mathbf{x}^{r-s})| \leq C(1 + \|\mathbf{x}^{r-s}\|_2)$$

for some $C > 0$. Then these quantities are uniformly integrable over bounded domains of $s \leq r \leq t$ and $\mathbf{x}$, by [63, Theorem II.2.1]. Furthermore $\|\mathbf{v}_i^{r-s}\|_2$ is bounded by Proposition A.2(b), so the integrands on the right sides of both (192–193) are also uniformly integrable over these domains. Then applying dominated convergence, we may differentiate (191) in $\mathbf{x}$ under the integral to obtain

$$\nabla P^\varepsilon_{s,t}a(\mathbf{x}) - \nabla P_{t-s}a(\mathbf{x}) = \varepsilon \int_s^t h(r) \nabla P_{r-s}(b^\top \nabla P^\varepsilon_{r,t}a)(\mathbf{x})\mathrm{d}r,$$

and take the limit $\varepsilon \to 0$ to get $\nabla P^{\varepsilon}_{s,t} a(\mathbf{x}) \to \nabla P_{t-s} a(\mathbf{x})$. Applying this with $s = r$ to the right side of (191), and taking the limit $\varepsilon \to 0$ in (191) using uniform integrability of (192), we arrive at

$$\lim_{\varepsilon \to 0} \frac{P^{\varepsilon}_{s,t} a(\mathbf{x}) - P_{t-s} a(\mathbf{x})}{\varepsilon} = \int_s^t h(r) P_{r-s}(b^\top \nabla P_{t-r} a)(\mathbf{x}) \mathrm{d}r.$$

For $s = 0$, this means

$$\lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\Big(\langle a(\mathbf{x}^{t,\varepsilon})\rangle - \langle a(\mathbf{x}^t)\rangle\Big) = \int_0^t h(r) P_r(b^\top \nabla P_{t-r} a)(\mathbf{x}) \mathrm{d}r,$$

verifying that (21) holds with response function $R(t,s)$ given by (179). Continuity of this function $R(t,s)$ in (s,t) follows from the above uniform integrability statements, together with continuity of $t \mapsto \nabla P_t(\mathbf{x})$ in t as shown in Proposition A.2.

For uniqueness, observe that if $\tilde{R}(t,s)$ is any continuous function different from $R(t,s)$, then they must differ on a subset of (s,t) of positive Lebesgue measure. Then there exists a continuous bounded function $h : [0,\infty) \to \mathbb{R}$ such that $\int_0^t R(t,s)h(s)\mathrm{d}s \neq \int_0^t \tilde{R}(t,s)h(s)\mathrm{d}s$, implying that $\tilde{R}$ cannot satisfy (21). Thus this response function $R(t,s)$ is unique. □

A.2 Discrete dynamics

We record (elementary) analogues of the preceding results for discrete dynamics

$$\mathbf{x}^{t+1} = \mathbf{x}^t + u(\mathbf{x}^t) + \sqrt{2}\,\mathbf{M}(\mathbf{b}^{t+1} - \mathbf{b}^t) \tag{194}$$

where $\{\mathbf{b}^t\}_{t \in \mathbb{Z}_+}$ is a Gaussian process with $\mathbf{b}^0 = 0$ and independent increments $\mathbf{b}^{t+1} - \mathbf{b}^t \sim \mathcal{N}(0, \gamma \mathbf{I})$, for some $\gamma > 0$. The following is an analogue of Proposition A.1.

Proposition A.4. *Suppose $u : \mathbb{R}^m \to \mathbb{R}^m$ is Lipschitz, and let $\{\mathbf{x}^t\}_{t \in \mathbb{Z}_+}$ be the solution of (194) with initial condition $\mathbf{x}^0 = \mathbf{x}$. For any Lipschitz functions $a : \mathbb{R}^m \to \mathbb{R}$ and $b : \mathbb{R}^m \to \mathbb{R}^m$, define*

$$R(t,s) = P_s\big(b^\top P(\nabla P_{t-s-1} a)\big)(\mathbf{x})$$

where $P_t f(\mathbf{x}) = \mathbb{E}[f(\mathbf{x}^t) \mid \mathbf{x}^0 = \mathbf{x}]$. Then for any $s, t \in \mathbb{Z}_+$ with $s < t$,

$$R(t,s) = \lim_{\varepsilon \to 0} \frac{1}{\varepsilon}\Big(\mathbb{E}[a(\mathbf{x}^{t,\varepsilon}) \mid \mathbf{x}^{0,\varepsilon} = \mathbf{x}] - \mathbb{E}[a(\mathbf{x}^t) \mid \mathbf{x}^0 = \mathbf{x}]\Big)$$

where $\{\mathbf{x}^{t,\varepsilon}\}_{t \in \mathbb{Z}_+}$ is the solution of the perturbed dynamics

$$\mathbf{x}^{t+1,\varepsilon} = \mathbf{x}^{t,\varepsilon} + u(\mathbf{x}^{t,\varepsilon}) + \varepsilon b(\mathbf{x}^{t,\varepsilon})\mathbf{1}_{s=t} + \sqrt{2}\,\mathbf{M}(\mathbf{b}^{t+1} - \mathbf{b}^t)$$

with the same initial condition $\mathbf{x}^{0,\varepsilon} = \mathbf{x}$.

Proof. Write as shorthand $P = P_1$. If f is L-Lipschitz, then (coupling the processes with initializations $\mathbf{x}, \mathbf{y}$ by the same $\{\mathbf{b}^t\}$)

$$|Pf(\mathbf{x}) - Pf(\mathbf{y})| = \Big|\mathbb{E}[f(\mathbf{x} + u(\mathbf{x}) - \sqrt{2}\mathbf{M}\,\mathbf{b}^1)] - \mathbb{E}[f(\mathbf{y} + u(\mathbf{y}) - \sqrt{2}\mathbf{M}\,\mathbf{b}^1)]\Big| \leq L(1 + L_u)\|\mathbf{x} - \mathbf{y}\|$$

where L_u is the Lipschitz constant of u. Hence Pf is Lipschitz, so $P_t f$ is Lipschitz for all $t \geq 0$.

Let P^{ε}_t be the Markov semigroup for the dynamics

$$\mathbf{x}^{t+1} = \mathbf{x}^t + u(\mathbf{x}^t) + \varepsilon b(\mathbf{x}^t) + \sqrt{2}\,\mathbf{M}(\mathbf{b}^{t+1} - \mathbf{b}^t),$$

and write as shorthand $P^{\varepsilon} = P^{\varepsilon}_1$. Then by definition,

$$\mathbb{E}[a(\mathbf{x}^{t,\varepsilon}) \mid \mathbf{x}^{0,\varepsilon} = \mathbf{x}] = P_s P^{\varepsilon} P_{t-s-1}(\mathbf{x}),$$

so

$$\lim_{\varepsilon\to 0}\frac{1}{\varepsilon}\Big(\mathbb{E}[a(\mathbf{x}^{t,\varepsilon}) \mid \mathbf{x}^{0,\varepsilon}=\mathbf{x}] - \mathbb{E}[a(\mathbf{x}^{t}) \mid \mathbf{x}^{0}=\mathbf{x}]\Big) = \partial_\varepsilon|_{\varepsilon=0} P_s P^\varepsilon P_{t-s-1} a(\mathbf{x}). \tag{195}$$

Note that for any L-Lipschitz function f, we have

$$\partial_\varepsilon P^\varepsilon f(\mathbf{x}) = \partial_\varepsilon \mathbb{E}[f(\mathbf{x}+u(\mathbf{x})+\varepsilon b(\mathbf{x})+\mathbf{b}^1)] = b(\mathbf{x})^\top \mathbb{E}[\nabla f(\mathbf{x}+u(\mathbf{x})+\varepsilon b(\mathbf{x})+\mathbf{b}^1)] \tag{196}$$

where the derivative may be taken under the expectation by dominated convergence. In particular,

$$\partial_\varepsilon|_{\varepsilon=0} P^\varepsilon f(\mathbf{x}) = b(\mathbf{x})^\top \mathbb{E}[\nabla f(\mathbf{x}+u(\mathbf{x})+\mathbf{b}^1)] = b(\mathbf{x})^\top P(\nabla f)(\mathbf{x}).$$

The derivative (196) is also bounded for all $\varepsilon \geq 0$ by $L\|b(\mathbf{x})\|$, which is integrable under P_s since b is Lipschitz. Then again by dominated convergence,

$$\partial_\varepsilon|_{\varepsilon=0} P_s P^\varepsilon P_{t-s-1} a(\mathbf{x}) = P_s \partial_\varepsilon|_{\varepsilon=0} P^\varepsilon P_{t-s-1} a(\mathbf{x}) = P_s(b^\top P(\nabla P_{t-s-1} a))(\mathbf{x}),$$

and the result follows from applying this to (195). □

The following is an analogue of the first statement of (184).

Lemma A.5. *Let $\{\mathbf{x}^t\}_{t\in\mathbb{Z}_+}$ be the solution to (194) where $u(\cdot)$ is Lipschitz, and consider the first variation processes*

$$\mathbf{v}_i^{t+1} = \mathbf{v}_i^t + \sum_{j=1}^m \partial_j u_i(\mathbf{x}^t)\cdot \mathbf{v}_j^t$$

with initializations $\mathbf{v}_i^0 = \mathbf{e}_i$. Denote $P_t f(\mathbf{x}) = \mathbb{E}[f(\mathbf{x}^t) \mid \mathbf{x}^0 = \mathbf{x}] = \langle f(\mathbf{x}^t)\rangle$. Then for any Lipschitz function $f : \mathbb{R}^{d+K} \to \mathbb{R}$,

$$\nabla P_t f(\mathbf{x}) = \Big\langle \sum_{j=1}^m \partial_j f(\mathbf{x}^t)\mathbf{v}_j^t \Big\rangle.$$

Proof. Stacking $\mathbf{V}^t = [\mathbf{v}_1^t, \ldots, \mathbf{v}_m^t]^\top \in \mathbb{R}^{m\times m}$ with initial condition $\mathbf{V}^0 = \mathbf{I}_m$, the evolution of $\mathbf{V}^t$ is

$$\mathbf{V}^{t+1} = [\mathbf{I} + \mathrm{d}u(\mathbf{x}^t)]\mathbf{V}^t$$

where $\mathrm{d}u$ is the derivative of $u(\cdot)$. Writing $\mathbf{x}^t(\mathbf{x})$ for the dependence of $\mathbf{x}^t$ on the initial condition $\mathbf{x}^0 = \mathbf{x}$, and writing $\mathrm{d}\mathbf{x}^t(\mathbf{x})$ for its derivative in $\mathbf{x}$, by the chain rule we have $\mathrm{d}\mathbf{x}^{t+1}(\mathbf{x}) = [\mathbf{I} + \mathrm{d}u(\mathbf{x}^t)]\mathrm{d}\mathbf{x}^t(\mathbf{x})$, with initial condition $\mathrm{d}\mathbf{x}^0(\mathbf{x}) = \mathbf{I}$. Thus $(\mathbf{V}^t)^\top = \mathrm{d}\mathbf{x}^t(\mathbf{x})$ for all $t \geq 0$, so

$$\nabla_{\mathbf{x}} f(\mathbf{x}^t(\mathbf{x})) = [\mathrm{d}\mathbf{x}^t(\mathbf{x})]^\top \nabla f(\mathbf{x}^t) = \sum_{j=1}^m \partial_j f(\mathbf{x}^t)\mathbf{v}_j^t.$$

By dominated convergence we have $\nabla P_t f(\mathbf{x}) = \nabla_{\mathbf{x}}\langle f(\mathbf{x}^t(\mathbf{x}))\rangle = \langle \nabla_{\mathbf{x}} f(\mathbf{x}^t(\mathbf{x}))\rangle$, and the result follows. □

References

[1] Mark Kac. Foundations of kinetic theory. In *Proceedings of The third Berkeley symposium on mathematical statistics and probability*, volume 3, pages 171–197, 1956.

[2] Henry P McKean Jr. A class of Markov processes associated with nonlinear parabolic equations. *Proceedings of the National Academy of Sciences*, 56(6):1907–1911, 1966.

[3] Jürgen Gärtner. On the McKean-Vlasov limit for interacting diffusions. *Mathematische Nachrichten*, 137(1):197–248, 1988.

[4] Alain-Sol Sznitman. Topics in propagation of chaos. *Ecole d'été de probabilités de Saint-Flour XIX—1989*, 1464:165–251, 1991.

[5] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. I. Models and methods. *arXiv preprint arXiv:2203.00446*, 2022.

[6] Louis-Pierre Chaintron and Antoine Diez. Propagation of chaos: a review of models, methods and applications. II. Applications. *arXiv preprint arXiv:2106.14812*, 2021.

[7] Daniel Lacker. Hierarchies, entropy, and quantitative propagation of chaos for mean field diffusions. *Probability and Mathematical Physics*, 4(2):377–432, 2023.

[8] Daniel Lacker and Luc Le Flem. Sharp uniform-in-time propagation of chaos. *Probability Theory and Related Fields*, 187(1-2):443–480, 2023.

[9] Haim Sompolinsky and Annette Zippelius. Dynamic theory of the spin-glass phase. *Physical Review Letters*, 47(5):359, 1981.

[10] Haim Sompolinsky and Annette Zippelius. Relaxational dynamics of the Edwards-Anderson model and the mean-field theory of spin-glasses. *Physical Review B*, 25(11):6860, 1982.

[11] Andrea Crisanti, Heinz Horner, and H J Sommers. The spherical p-spin interaction spin-glass model: the dynamics. *Zeitschrift für Physik B Condensed Matter*, 92:257–271, 1993.

[12] Leticia F Cugliandolo and Jorge Kurchan. Analytical solution of the off-equilibrium dynamics of a long-range spin-glass model. *Physical Review Letters*, 71(1):173, 1993.

[13] Leticia F Cugliandolo and Jorge Kurchan. On the out-of-equilibrium relaxation of the sherrington-kirkpatrick model. *Journal of Physics A: Mathematical and General*, 27(17):5749, 1994.

[14] Marc Mézard, Giorgio Parisi, and Miguel Angel Virasoro. *Spin glass theory and beyond*, volume 9 of *World Scientific Lecture Notes in Physics*. World Scientific Publishing Co., Inc., Teaneck, NJ, 1987.

[15] Elisabeth Agoritsas, Giulio Biroli, Pierfrancesco Urbani, and Francesco Zamponi. Out-of-equilibrium dynamical mean-field equations for the perceptron model. *J. Phys. A*, 51(8):085002, 36, 2018.

[16] C De Dominicis. Dynamics as a substitute for replicas in systems with quenched random impurities. *Physical Review B*, 18(9):4913, 1978.

[17] Jorge Kurchan. Supersymmetry in spin glass dynamics. *Journal de Physique I*, 2(7):1333–1352, 1992.

[18] Leticia F Cugliandolo. Dynamics of glassy systems. *arXiv preprint cond-mat/0210312*, 2002.

[19] Stefano Sarao Mannelli, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborova. Passed & spurious: Descent algorithms and local minima in spiked matrix-tensor models. In *International Conference on Machine Learning*, pages 4333–4342, 2019.

[20] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, and Lenka Zdeborová. Who is afraid of big bad minima? analysis of gradient-flow in spiked matrix-tensor models. *Advances in neural information processing systems*, 32, 2019.

[21] 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(1):011057, 2020.

[22] Stefano Sarao Mannelli and Pierfrancesco Urbani. Analytical study of momentum-based acceleration methods in paradigmatic high-dimensional non-convex problems. *Advances in Neural Information Processing Systems*, 34:187–199, 2021.

[23] Tengyuan Liang, Subhabrata Sen, and Pragya Sur. High-dimensional asymptotics of Langevin dynamics in spiked matrix models. *Inf. Inference*, 12(4):iaad042,33, 2023.

[24] Francesca Mignacco, Pierfrancesco Urbani, and Lenka Zdeborová. Stochasticity helps to navigate rough landscapes: comparing gradient-descent-based algorithms in the phase retrieval problem. *Machine Learning: Science and Technology*, 2(3):035029, 2021.

[25] Stefano Sarao Mannelli, Giulio Biroli, Chiara Cammarota, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Complex dynamics in simple neural networks: Understanding gradient flow in phase retrieval. *Advances in Neural Information Processing Systems*, 33:3265–3274, 2020.

[26] Qiyang Han and Xiaocong Xu. Gradient descent inference in empirical risk minimization. *arXiv preprint arXiv:2412.09498*, 2024.

[27] Francesca Mignacco, Florent Krzakala, Pierfrancesco Urbani, and Lenka Zdeborová. Dynamical mean-field theory for stochastic gradient descent in Gaussian mixture classification. *J. Stat. Mech. Theory Exp.*, (12):Paper No. 124008, 23, 2021.

[28] Francesca Mignacco and Pierfrancesco Urbani. The effective noise of stochastic gradient descent. *Journal of Statistical Mechanics: Theory and Experiment*, 2022(8):083405, 2022.

[29] Blake Bordelon and Cengiz Pehlevan. Self-consistent dynamical field theory of kernel evolution in wide neural networks. *Advances in Neural Information Processing Systems*, 35:32240–32256, 2022.

[30] Yatin Dandi, Emanuele Troiani, Luca Arnaboldi, Luca Pesce, Lenka Zdeborová, and Florent Krzakala. The benefits of reusing batches for gradient descent in two-layer networks: Breaking the curse of information and leap exponents. *arXiv preprint arXiv:2402.03220*, 2024.

[31] Blake Bordelon, Hamza Chaudhry, and Cengiz Pehlevan. Infinite limits of multi-head transformer dynamics. *Advances in Neural Information Processing Systems*, 37:35824–35878, 2024.

[32] Blake Bordelon, Alexander Atanasov, and Cengiz Pehlevan. A dynamical model of neural scaling laws. *arXiv preprint arXiv:2402.01092*, 2024.

[33] Andrea Montanari and Pierfrancesco Urbani. Dynamical decoupling of generalization and overfitting in large two-layer networks. *arXiv preprint arXiv:2502.21269*, 2025.

[34] G. Ben Arous and A. Guionnet. Large deviations for Langevin spin glass dynamics. *Probab. Theory Related Fields*, 102(4):455–509, 1995.

[35] M. Grunwald. Sanov results for Glauber spin-glass dynamics. *Probab. Theory Related Fields*, 106(2):187–232, 1996.

[36] G. Ben Arous and A. Guionnet. Symmetric Langevin spin glass dynamics. *Ann. Probab.*, 25(3):1367–1422, 1997.

[37] A. Guionnet. Averaged and quenched propagation of chaos for spin glass dynamics. *Probab. Theory Related Fields*, 109(2):183–215, 1997.

[38] G. Ben Arous, A. Dembo, and A. Guionnet. Aging of spherical spin glasses. *Probab. Theory Related Fields*, 120(1):1–67, 2001.

[39] Gérard Ben Arous, Amir Dembo, and Alice Guionnet. Cugliandolo-Kurchan equations for dynamics of spin-glasses. *Probab. Theory Related Fields*, 136(4):619–660, 2006.

[40] Amir Dembo and Reza Gheissari. Diffusions interacting through a random matrix: universality via stochastic Taylor expansion. *Probab. Theory Related Fields*, 180(3-4):1057–1097, 2021.

[41] Amir Dembo, Eyal Lubetzky, and Ofer Zeitouni. Universality for Langevin-like spin glass dynamics. *Ann. Appl. Probab.*, 31(6):2864–2880, 2021.

[42] Michael Celentano, Chen Cheng, and Andrea Montanari. The high-dimensional asymptotics of first order methods with random data. *arXiv preprint arXiv:2112.07572*, 2021.

[43] Erwin Bolthausen. An iterative construction of solutions of the TAP equations for the Sherrington-Kirkpatrick model. *Comm. Math. Phys.*, 325(1):333–366, 2014.

[44] Mohsen Bayati and Andrea Montanari. The dynamics of message passing on dense graphs, with applications to compressed sensing. *IEEE Trans. Inform. Theory*, 57(2):764–785, 2011.

[45] Adel Javanmard and Andrea Montanari. State evolution for general approximate message passing algorithms, with applications to spatial coupling. *Inf. Inference*, 2(2):115–144, 2013.

[46] Cédric Gerbelot, Emanuele Troiani, Francesca Mignacco, Florent Krzakala, and Lenka Zdeborová. Rigorous dynamical mean-field theory for stochastic gradient descent methods. *SIAM J. Math. Data Sci.*, 6(2):400–427, 2024.

[47] Qiyang Han. Entrywise dynamics and universality of general first order methods. *arXiv preprint arXiv:2406.19061*, 2024.

[48] Jeffrey P Spence, Nasa Sinnott-Armstrong, Themistocles L Assimes, and Jonathan K Pritchard. A flexible modeling and inference framework for estimating variant effect sizes from GWAS summary statistics. *BioRxiv*, pages 2022–04, 2022.

[49] Sumit Mukherjee, Bodhisattva Sen, and Subhabrata Sen. A mean field approach to empirical Bayes estimation in high-dimensional linear regression. *arXiv preprint arXiv:2309.16843*, 2023.

[50] Youngseok Kim, Wei Wang, Peter Carbonetto, and Matthew Stephens. A flexible empirical Bayes approach to multiple linear regression and connections with penalized regression. *Journal of Machine Learning Research*, 25(185):1–59, 2024.

[51] Zhou Fan, Leying Guan, Yandi Shen, and Yihong Wu. Gradient flows for empirical Bayes in high-dimensional linear models. *arXiv preprint arXiv:2312.12708*, 2023.

[52] Zhou Fan, Justin Ko, Bruno Loureiro, Yue M. Lu, and Yandi Shen. Dynamical mean-field analysis of adaptive Langevin diffusions: Replica-symmetric fixed point and empirical Bayes. 2025.

[53] Rep Kubo. The fluctuation-dissipation theorem. *Reports on progress in physics*, 29(1):255, 1966.

[54] Cédric Villani. *Optimal transport: old and new*, volume 338. Springer, 2008.

[55] Amir Dembo and Jean-Dominique Deuschel. Markovian perturbation, response and fluctuation dissipation theorem. *Annales de l'IHP Probabilités et statistiques*, 46(3):822–852, 2010.

[56] Xian Chen and Chen Jia. Mathematical foundation of nonequilibrium fluctuation–dissipation theorems for inhomogeneous diffusion processes with unbounded coefficients. *Stochastic Processes and their Applications*, 130(1):171–202, 2020.

[57] Jean-François Le Gall. *Brownian motion, martingales, and stochastic calculus*, volume 274 of *Graduate Texts in Mathematics*. Springer, 2016.

[58] L Chris G Rogers and David Williams. *Diffusions, Markov processes, and martingales: Itô calculus*, volume 2. Cambridge university press, 2000.

[59] Hermann Brunner. *Collocation methods for Volterra integral and related functional differential equations*, volume 15. Cambridge university press, 2004.

[60] Tianhao Wang, Xinyi Zhong, and Zhou Fan. Universality of approximate message passing algorithms and tensor networks. *The Annals of Applied Probability*, 34(4):3943–3994, 2024.

[61] Roman Vershynin. *High-dimensional probability: An introduction with applications in data science.* Cambridge university press, 2018.

[62] Ramon Van Handel. Probability in high dimension. *Lecture Notes (Princeton University)*, 2014.

[63] H Kunita. Stochastic differential equations and stochastic flows of diffeomorphisms. *École d'Été de Probabilités de Saint-Flour XII-1982*, pages 143–303, 1984.