



POLITECNICO DI TORINO
CORSO DI LAUREA MAGISTRALE IN
PHYSICS OF COMPLEX SYSTEMS (FISICA DEI SISTEMI COMPLESSI)
A.A. 2024/2025
SESSIONE DI LAUREA LUGLIO 2025

Dynamics of Gradient Descent in High-dimensional Non-convex Canyon Landscapes

A THESIS AUTHORED BY

Dafne Prado Bandeira
s313434

BASED ON AN INTERNSHIP AT
CNRS, CEA, Institut de Physique Théorique,
F-91191, Gif-sur-Yvette, France

AND SUPERVISED BY

INTERNAL SUPERVISOR

Alessandro Pelizzola^{1,2,3}

EXTERNAL SUPERVISOR

Pierfrancesco Urbani⁴

¹ DISAT, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy

² INFN, Sezione di Torino, via Pietro Giuria 1, I-10125 Torino, Italy

³ CNR-ISC c/o DISAT, Politecnico di Torino, Corso Duca degli Abruzzi 24, I-10129 Torino, Italy

⁴ Université Paris-Saclay, CNRS, CEA, Institut de Physique Théorique, F-91191, Gif-sur-Yvette, France

Dynamics of Gradient Descent in High-dimensional Non-convex Canyon Landscapes

Dafne Prado Bandeira

Abstract

Continuous constraint satisfaction problems (CCSPs) describe systems with continuous degrees of freedom subject to random constraints, and they provide a unifying framework for studying high-dimensional optimization problems such as confluent biological tissues and artificial neural networks. In this work, we investigate a mean-field model that captures the essential features of non-convex energy landscapes in the satisfiable (SAT) phase. We extend the model by introducing correlated replicas to probe how similar initial configurations evolve under gradient descent dynamics. Using dynamical mean-field theory (DMFT), we analyze the evolution of inter-replica correlations in the overparametrized regime and find that they relax to non-zero asymptotic values, demonstrating that gradient descent retains memory of initial overlaps. The scaling behavior of these correlations exhibits a Lyapunov-like sensitivity to perturbations. This study contributes to the theoretical understanding of high-dimensional optimization and offers a tractable setting for investigating the geometry of solution manifolds relevant to machine learning.

Contents

1	Introduction	1
2	Model	4
2.1	Generalization of the non-linear constraint	5
2.2	Dynamical equations	5
3	Methods	7
3.1	Dynamical mean field theory (DMFT)	7
3.2	Numerical Integration	12
4	Results	16
5	Conclusions	19

Chapter 1

Introduction

Continuous constraint satisfaction problems (CCSPs) involve finding solutions for systems with continuous degrees of freedom that are subject to a set of constraints. These systems are considered overparametrized when the number of degrees of freedom exceeds the number of constraints. From a statistical physics perspective, a particularly interesting scenario arises when the constraints are randomly selected from a specific ensemble. However, most of the literature focuses on discrete, rather than continuous, degrees of freedom [1], [2].

In such scenarios, large systems typically exhibit a phase transition as the density of constraints relative to the number of degrees of freedom increases. This transition occurs from an overparametrized, satisfiable (SAT) phase—where feasible configurations exist that satisfy all constraints—to an underparametrized, unsatisfiable (UNSAT) phase, where the minimum number of unsatisfied constraints is greater than zero.

This type of phase transition is reminiscent of the jamming transition of spheres. In this transition, densely packed spheres move from a fluid-like state to a rigid, solid-like state as the packing density increases, exhibiting universal features [3]–[5]. Using the random perceptron problem [6] as a prototype for continuous constraint satisfaction problems that exhibit jamming-like transitions, previous work [7] has suggested that the non-trivial criticality and universality of the SAT/UNSAT transition are associated with non-convexity in the free energy landscape. In more abstract terms, studying this universality requires characterizing overparametrized non-convex loss landscapes under random constraints. This exploration is conducted using gradient descent (GD) algorithms, which iteratively update the system by moving in the direction of the steepest descent of the loss function. Physically, this corresponds to minimizing a free energy landscape, similar to Langevin dynamics at zero temperature. However, the specific mechanisms and pathways through which GD navigates these landscapes remain largely unexplored. Recent progress has been made in this direction using dynamical mean-field theory (DMFT), which has revealed hidden dynamical time scales in gradient descent for constraint satisfaction problems [8]. These insights further motivate the study of simplified solvable models, where the interplay between geometry and dynamics can be analyzed in detail.

Understanding the abstract nature of CCSPs allows for insights from diverse contexts. One notable application of CCSPs is in the study of confluent tissues [9]–[11]. Confluent tissues are biological tissues where cells are tightly packed, with no gaps or overlaps,

forming a tessellated surface. These tissues are typically modeled as polygon-filled surfaces, with the cells represented as polygons that strive to achieve a target volume and area. The most relevant models are the Vertex [12] and Voronoi [13] models, where the degrees of freedom are, respectively, the vertices of the polygons and their centers. These tissues exhibit various states of aggregation, such as liquid-like, solid-like, and glassy-like states, and have been shown to undergo phase transitions similar to those found in jamming models [14]. These transitions are significant in processes like morphogenesis [15] and the spread of metastatic cancer [16], [17]. In Voronoi models, the rigidity transition from fluid to solid is attributed to changes in cell shape, with the packing fraction acting as a control parameter. Added stress due to deviations from the target cell shape occurs because the surface constraint may make it impossible to satisfy the shape requirements for all cells.

In 3D, it has been observed that when the order parameter is sufficiently large, gradient descent successfully identifies configurations where the cells achieve their target volume and surface area [18]. From a constraint satisfaction perspective, this translates into a zero-energy configuration, corresponding to the ground state of the Hamiltonian. Furthermore, analysis of the Hessian spectrum of the Hamiltonian indicates an extensive number of zero-energy modes. Consequently, the minima identified by gradient descent are connected along a zero-energy manifold of configurations, resembling a canyon in phase space. If the order parameter is lowered below a critical point, this canyon landscape disappears, and gradient descent leads to configurations trapped in local minima at finite energy, behaving like glass. This establishes a direct analogy between the rigidity transition in confluent tissues and the SAT/UNSAT transition in CCSPs.

This work examines the dynamics of a mean-field abstract model that undergoes a phase transition similar to the rigidity transition in confluent tissues. Specifically, the model describes a satisfiability transition for continuous degrees of freedom subjected to random non-linear equality constraints. Previous studies have suggested that the satisfiability threshold determined by gradient descent dynamics may coincide with the thermodynamic threshold and that zero-temperature replica symmetry breaking does not affect GD dynamics [11]. Although this model does not derive the rigidity transition from first principles, it offers the advantage of being analytically solvable.

A similar landscape also appears in artificial neural networks (ANNs). Essentially, ANNs are used for performing non-linear regression tasks, where training involves minimizing a loss function within a high-dimensional landscape. The zero-training loss manifold, or “canyon”, arises from the constraints on the neural weights imposed by the data points. This minimization is carried out using gradient descent. The high dimensionality is due to both the complexity and quantity of the training data and the large number of parameters to be optimized. Typically, this scenario is overparametrized, with more parameters than data points. Since the cost function depends on the data, the landscape is expected to be rugged. As a result, the gradient descent dynamics are highly complex and do not guarantee finding a global minimum, often leading to entrapment in local minima.

In this thesis, we extend the mean-field framework by explicitly introducing correlated replicas of the system. This construction allows us to track how initially similar configurations evolve under gradient descent. Our analysis, based on DMFT and numerical integration, shows that replicas retain a finite memory of their initial overlap: cross-correlation functions relax to non-zero plateaus, while equal-time correlations exhibit an almost linear scaling with the initial overlap. These findings demonstrate that

gradient descent dynamics preserve structured memory of initial conditions, indicating that relaxation in high-dimensional landscapes is far from featureless.

The remainder of this thesis is organized as follows. In Chapter 2, we introduce the mean-field model and its replica extension. Chapter 3 presents the derivation of the DMFT equations using the Martin-Siggia-Rose-Janssen-De Dominicis (MSRJD) formalism and discusses the numerical discretization scheme employed for their integration. Our results for correlation and response functions, including the analysis of cross-replica dynamics, are reported in Chapter 4. Finally, in Chapter 5, we summarize our findings and outline possible future extensions of this work.

Chapter 2

Model

The set of degrees of freedom will be taken to be n copies parametrized by $\gamma = 0, \dots, n-1$ of an N -dimensional vector $\vec{x}^{(\gamma)} = [x_1^{(\gamma)}, x_2^{(\gamma)}, \dots, x_N^{(\gamma)}]$ with a fixed modulus $|\vec{x}^{(\gamma)}|^2 = N$, which defines a compact hyper-spherical phase space. For simplicity of notation, we denote $\vec{x}$ the vector that includes all these degrees of freedom, therefore $\vec{x} = \{\vec{x}^{(\gamma)}\}_{\gamma=0}^{n-1}$. The random constraints will be added by the simplest non-linear term possible, which is a two-body interaction term. Each replica experiences the same disorder through interaction matrices J^μ that add constraints as defined by Eq. (2.1). The matrix is symmetric $J_{ij}^\mu = J_{ji}^\mu$ and has random Gaussian entries with zero mean and unity variance.

$$r_\mu(\vec{x}^{(\gamma)}) = \frac{1}{N} \sum_{i < j}^N J_{ij}^\mu x_i^{(\gamma)} x_j^{(\gamma)} \quad (2.1)$$

Eq. (2.1) delineates a series of $M = \varepsilon N$ non-linear functions, yielding an extensive random variable. The equality constraints are imposed by holding Eq. (2.1) constant at $r_\mu(\vec{x}^{(\gamma)}) = p_0$ for all replicas. Our focus lies within the over-parametrized regime, characterized by $\varepsilon < 1$, signifying more degrees of freedom than constraints. This behaviour persists as a function of p_0 while keeping ε fixed. Specifically, for small p_0 , the model resides in the satisfiable (SAT/over-parametrized) phase, whereas for large p_0 , it transitions to the unsatisfiable (UNSAT/under-parametrized) phase.

The value to be optimized is then the square-loss cost function, which may also be interpreted as the Hamiltonian of our system.

$$H[\vec{x}] = \frac{1}{2} \sum_{\gamma=0}^{n-1} \sum_{\mu=1}^M (r_\mu(\vec{x}^{(\gamma)}) - p_0)^2 \quad (2.2)$$

Therefore, the cost function depends on a non-linear random parameter ($r_\mu(\vec{x}^{(\gamma)})$), the p_0 value, and the number of constraints M parametrized by the value of ε . Qualitatively, for a fixed value of ε and sufficiently small p_0 , there exists a set of solutions $\vec{x}$ for which the Hamiltonian is zero. This corresponds to a perfectly satisfiable problem ($r_\mu(\vec{x}^{(\gamma)}) = p_0, \forall \mu, \gamma$), which is achievable for small p_0 because r_μ is a Gaussian random variable with a zero mean (being the convolution of Gaussians with zero mean). Because of the Gaussian nature of the model, the small values of p_0 are typical realizations (SAT). In

contrast, large values of p_0 are large deviations of the zero energy Hamiltonian (UNSAT).

The thermodynamic properties of this problem for a fixed value of ε and a single replica in the zero temperature limit were studied in reference [9]. The study reveals the existence of two regions separated by a critical value p_J . In one region, with probability one, the typical configuration of the Gibbs measure defined by Eq. (2.2) has zero energy, corresponding to the SAT phase. In contrast, the other region is characterized by a single minimum solution (UNSAT phase). Additionally, within the SAT region, there are two phases separated by another critical value p_G . This transition in the SAT region occurs when one-step replica symmetry breaking solutions ($p < p_G$) become unstable at the critical value p_G and undergo a Gardner transition into a full replica symmetry breaking phase ($p_G < p < p_J$) [19]. This type of transition belongs to the same universality class as other non-convex continuous constraint satisfaction problems [4].

This study aims to explore the characteristics of out-of-equilibrium algorithms and juxtapose them with the thermodynamic framework. Specifically, we concentrate on the gradient descent dynamics as described by equations derived from dynamical mean-field theory (DMFT), as previously investigated in [11] for a single replica. Our focus lies in seeking zero-energy solutions across the multiple replicas of the degrees of freedom. The rationale behind employing multiple replicas is to gain a deeper insight into the canyon landscape and the significance of correlations between the degrees of freedom.

2.1 Generalization of the non-linear constraint

The model can be generalized by considering that the variance of the distribution of Eq. (2.1) can be expressed as a generalized function $G(q)$, as shown in Eq. (2.3). With this, $h_\mu(\vec{x})$ is a general Gaussian random variable that correlates the degrees of freedom.

$$\overline{r_\mu(\vec{x})r_\nu(\vec{y})} = \delta_{\mu,\nu}G\left(\frac{\vec{x}\cdot\vec{y}}{N}\right) \quad (2.3)$$

$$r_\mu(\vec{x}) = \sum_{\tau} a_{\tau} N^{-\tau/2} \sum_{i_1 < \dots < i_N} J_{i_1, \dots, i_{\tau}}^{\mu} x_{i_1} \dots x_{i_{\tau}} \quad (2.4)$$

In Eq. (2.4), $J_{i_1, \dots, i_{\tau}}^{\mu}$ is a rank τ tensor that contracts with τ terms of $\vec{x}$. These terms are unitary Gaussian variables with zero mean. Consequently, a general function $G(q)$ can be expanded as a Taylor series in powers of q , with coefficients a_{τ} .

2.2 Dynamical equations

In statistical mechanics, the configuration of a dynamical system is represented by a point in phase space that evolves over time. This point can become trapped in the numerous wells (local minima) of the free energy landscape or relax to the ground state. Equilibrium properties can be determined using classical statistical mechanics methods, such as the cavity approach and the replica method. However, out-of-equilibrium properties are best described by envisioning the system's movement through this landscape according to Langevin dynamics. In the model described in this chapter, the Langevin equation minimizes the Hamiltonian for all degrees of freedom, which corresponds to the physical

description of gradient descent dynamics. This process is constrained on the hyper-sphere at all times by the dynamical Lagrange multiplier $\mu^{(\gamma)}(t)$.

$$\dot{x}_i^{(\gamma)}(t) = -\mu^{(\gamma)}(t)x_i^{(\gamma)}(t) - \frac{\partial H}{\partial x_i^{(\gamma)}(t)} \quad (2.5)$$

The initial conditions are uniformly selected at random within the constrained spherical space, resulting in a Gaussian distribution as the flat measure. This set of Nn differential equations can be solved using the mean-field character of the model [20]. To study the dynamics, we focus on the time-dependence of the correlation (Eq. (2.6)) and response (Eq. (2.7)) functions for different replicas, which can be found through dynamical mean field theory (DMFT).

$$C^{(\alpha,\beta)}(t_a, t_b) = \left\langle \frac{\vec{x}^{(\alpha)}(t_a) \cdot \vec{x}^{(\beta)}(t_b)}{N} \right\rangle \quad (2.6)$$

$$R^{(\alpha,\beta)}(t_a, t_b) = \frac{\delta \langle \vec{x}^{(\alpha)}(t_a) \rangle}{\delta h^{(\beta)}(t_b)} \Big|_{h^{(\beta)}(t_b)=0} \quad (2.7)$$

where the angled brackets $\langle \cdot \rangle$ represent average over disorder and initial condition, while $h^{(\beta)}(t_b)$ represents an instantaneous linear perturbation applied to the replica parametrized by β at time t_b . The response function in Eq. (2.7) indicates how this linear perturbation affects the trajectory of replica α at the instant t_a . According to causality, the response is zero if $t_b > t_a$. Additionally, one should expect that, for $\alpha \neq \beta$, $R^{(\alpha,\beta)}(t_a, t_b) = 0$ since the response for different replicas should be independent for all times.

Chapter 3

Methods

3.1 Dynamical mean field theory (DMFT)

The dynamic nature of the system provides a way to identify phase transitions in disordered systems without the need for replica calculations [21]. Employing the mean-field approach, we derive a set of equations for dynamic correlation and response functions, particularly in scenarios with a large number of degrees of freedom ($N \rightarrow \infty$), using Dynamical Mean Field Theory (DMFT) [20]. These DMFT equations are obtained through path integrals, known as the Martin-Siggia-Rose-Jenssen-De Dominicis (MSRJD) approach [22]. To simplify our calculations, we introduce Grassmann algebra, generating a supersymmetric (SUSY) field. Within this framework, the symmetries of the action functional implicitly encode the fluctuation-dissipation relations and time homogeneity properties of the system [23].

The dynamical generating functional is given by forcing the satisfaction of Eq. (2.5) in the entire path and averaging over disorder, and initial conditions. This formalism gives the MSRJD identity (Eq. (3.1)) due to the causality of the dynamics.

$$\begin{aligned}
 Z_{\text{dyn}} &= 1 \\
 &= \left\langle \int \prod_{\alpha=0}^{n-1} \mathcal{D}\vec{x}^{(\alpha)}(t) \prod_{i=0}^{N-1} \prod_{\alpha=0}^{n-1} \delta \left(-\dot{x}_i^{(\alpha)}(t) - \mu^{(\alpha)}(t)x_i^{(\alpha)}(t) - \frac{\partial H}{\partial x_i^{(\alpha)}(t)} \right) \right\rangle \\
 &= \left\langle \int \prod_{\alpha=0}^{n-1} \mathcal{D}\vec{x}^{(\alpha)}(t) \mathcal{D}\vec{\hat{x}}^{(\alpha)}(t) \exp \left[i \sum_{\alpha=0}^{n-1} \int dt \vec{\hat{x}}^{(\alpha)}(t) \cdot \left(-\dot{\vec{x}}^{(\alpha)}(t) - \mu^{(\alpha)}(t)\vec{x}^{(\alpha)}(t) - \frac{\partial H}{\partial \vec{x}^{(\alpha)}(t)} \right) \right] \right\rangle
 \end{aligned} \tag{3.1}$$

where we have included the conjugate field $\vec{\hat{x}}$ to rewrite the delta function restriction. It can be shown that the dynamical partition function can be written as

$$Z_{\text{dyn}} = \left\langle \int \mathcal{D}\vec{x}(t) \mathcal{D}\vec{\hat{x}}(t) \exp [S_K(\vec{x}(t)) + S_{\text{int}}(\vec{\hat{x}}(t))] \right\rangle. \tag{3.2}$$

where we defined $\mathcal{D}\vec{x}(t) \mathcal{D}\vec{\hat{x}}(t) = \prod_{\alpha=0}^{n-1} \mathcal{D}\vec{x}^{(\alpha)}(t) \mathcal{D}\vec{\hat{x}}^{(\alpha)}(t)$ and, as before, $\vec{x}(t) = \{\vec{x}^{(\alpha)}(t)\}_{\alpha=0}^{n-1}$

for simplicity of notation. In Eq. (3.2), the action is separated in kinetic and interaction parts, S_K and S_{int} respectively, which are explicitly given as

$$S_K(\vec{x}(t)) = -i \sum_{\alpha=0}^{n-1} \int dt \vec{x}^{(\alpha)}(t) [\partial_t + \mu^{(\alpha)}(t)] \vec{x}^{(\alpha)}(t) \quad (3.3)$$

$$S_{int}(\vec{x}(t)) = -i \sum_{\alpha=0}^{n-1} \int dt \vec{x}^{(\alpha)}(t) \frac{\partial H}{\partial \vec{x}^{(\alpha)}(t)}. \quad (3.4)$$

Notice that only the interaction action depends on disorder through the Hamiltonian. One can then introduce the super-symmetric (SUSY) formalism [23] to incorporate the variables $\vec{\tilde{x}}$ and $\vec{x}$ in a single variable $\vec{X}$ by extending the time coordinate to include a Grassmann coordinate θ as $t_a \rightarrow a = (t_a, \theta_a)$.

$$\vec{X}^{(\alpha)}(a) = \vec{x}^{(\alpha)}(t_a) + i\theta_a \vec{\tilde{x}}^{(\alpha)}(t_a) \quad (3.5)$$

Due to the inherent properties of Grassmann algebra [24], any function involving Grassmann variables must be linear in these terms. This is a consequence of the anticommuting algebra of the Grassmann variables, which causes integration and derivation to be equivalent operations concerning the Grassman variable.

It can be shown that the interaction action can be rewritten as a function of the Grassmann variable. For simplicity of notation we denote $\vec{X} = \{\vec{X}^{(\alpha)}\}_{\alpha=0}^{n-1}$.

$$S_{int}(\vec{X}) = - \sum_{\alpha=0}^{n-1} \int da H[\vec{X}^{(\alpha)}(a)] \quad (3.6)$$

where $H[\vec{X}]$ is defined by Eq. (2.2). Also, the kinetic action can be written as a function of a kinetic kernel $K^{(\alpha,\beta)}(a, b)$.

$$S_K(\vec{X}) = -\frac{1}{2} \sum_{\alpha,\beta=0}^{n-1} \int da db \vec{X}^{(\alpha)}(a) K^{(\alpha,\beta)}(a, b) \vec{X}^{(\beta)}(b) \quad (3.7)$$

The kinetic kernel can be explicitly expressed as $K^{(\alpha,\beta)}(a, b) = 2\delta_{\alpha,\beta}\delta(t_a - t_b)\theta_b[\partial_{t_b} + \mu^{(\beta)}(t_b)]$. The disorder average, denoted by $\mathbb{E}$, specifically applies to the term influenced by disorder within the interaction matrix. Eq. (3.8) calculates this average for the pertinent term within S_{int} , up to a multiplicative constant.

$$\begin{aligned} \mathbb{E} \exp \left(\sum_{\mu=1}^M \sum_{\alpha=0}^{n-1} \sum_{i < j}^N i \int da \frac{\hat{r}_\mu^{(\alpha)}(a)}{N} J_{ij}^\mu X_i^{(\alpha)}(a) X_j^{(\alpha)}(a) \right) \\ = \prod_{\mu=1}^M \prod_{i,j=1}^N \exp \left(-\frac{1}{4} \int da db \sum_{\alpha,\beta=0}^{n-1} \frac{\hat{r}_\mu^{(\alpha)}(a) X_i^{(\alpha)}(a) X_j^{(\alpha)}(a)}{N} \frac{\hat{r}_\mu^{(\beta)}(b) X_i^{(\beta)}(b) X_j^{(\beta)}(b)}{N} \right) \end{aligned} \quad (3.8)$$

This can be simplified using the dynamical overlap matrix defined by Eq. (3.9) and a function $G(Q)$.

$$Q^{(\alpha,\beta)}(a,b) = \frac{\vec{X}^{(\alpha)}(a) \cdot \vec{X}^{(\beta)}(b)}{N} \quad (3.9)$$

$$G(Q^{(\alpha,\beta)}(a,b)) \doteq G^{(\alpha,\beta)}(a,b) = \frac{[Q^{(\alpha,\beta)}(a,b)]^2}{2} \quad (3.10)$$

with this change of variables, the average over the disorder of the interaction action can be rewritten as

$$\begin{aligned} \mathbb{E} \exp(S_{\text{int}}(\vec{X})) &= \int \mathcal{D}\vec{r} \mathcal{D}\vec{\hat{r}} \prod_{\mu=1}^M \prod_{\alpha=0}^{n-1} \exp \left[\int da \left(-\frac{1}{2} (r_{\mu}^{(\alpha)}(a) - p_0)^2 + i \hat{r}_{\mu}^{(\alpha)}(a) r_{\mu}^{(\alpha)}(a) \right) \right. \\ &\quad \left. - \frac{1}{2} \int dadb \sum_{\beta=0}^{n-1} \hat{r}_{\mu}^{(\alpha)}(a) \hat{r}_{\mu}^{(\beta)}(b) G(Q^{(\alpha,\beta)}(a,b)) \right] \\ &= \left[\sqrt{\det(G+I)} \exp \left(\frac{p_0^2}{2} \sum_{\alpha,\beta=0}^{n-1} \int dadb (G^{-1}(Q^{(\alpha,\beta)}(a,b)) + I)^{-1} \right) \right]^{\varepsilon N} \\ &\doteq \mathcal{Z}^{\varepsilon N}. \end{aligned} \quad (3.11)$$

where I represents the identity operator, and the determinant appears from successive Gaussian integrals. The conjugate field $\vec{\hat{r}}$ is introduced to obtain a functional integral representation of the delta function. Notice that the kinetic action in Eq. (3.7) can be expressed as a function of Q , making it extensive with N , similar to the interaction action defined by Eq. (3.11). Using this result, the dynamical partition function can be computed for large N using a saddle-node approximation.

$$Z_{dyn} = \int \mathcal{D}Q \exp(N S_{dyn}[Q]) \xrightarrow{N \rightarrow \infty} \exp(N S_{dyn}[Q^*]) \quad (3.12)$$

$$S_{dyn} = \frac{1}{2} \ln(\det(Q)) + \varepsilon \ln \mathcal{Z} - \frac{1}{2} \sum_{\alpha,\beta=0}^{n-1} \int dadb K^{(\alpha,\beta)}(a,b) Q^{(\alpha,\beta)}(a,b) \quad (3.13)$$

and the saddle point Q^* satisfies for a given element $Q^{*(\alpha,\gamma)}(a,c)$

$$\begin{aligned} \frac{\delta S_{dyn}[Q]}{\delta Q^{(\alpha,\gamma)}(a,c)} \Big|_{Q^*} &= 0 \\ \frac{-K^{(\alpha,\gamma)}(a,c)}{2} + \frac{Q^{-1(\alpha,\gamma)}(a,c)}{2} + \varepsilon \frac{\delta \ln \mathcal{Z}}{\delta Q^{(\alpha,\gamma)}(a,c)} \Big|_{Q^*} &= 0. \end{aligned} \quad (3.14)$$

With the saddle-point approximation, the average of Q over the Gibbs measure of Z_{dyn} is given by the saddle-point itself.

$$\begin{aligned}
\langle Q^{(\alpha,\beta)}(a,b) \rangle_{Z_{dyn}} &= Q^{*(\alpha,\beta)}(a,b) \\
&= \frac{\vec{x}^{(\alpha)}(a) \cdot \vec{x}^{(\beta)}(b)}{N} + \theta_a \frac{i\vec{x}^{(\alpha)}(a) \cdot \vec{\hat{x}}^{(\beta)}(b)}{N} \\
&\quad + \theta_b \frac{i\vec{\hat{x}}^{(\alpha)}(a) \cdot \vec{x}^{(\beta)}(b)}{N} - \theta_a \theta_b \frac{\vec{\hat{x}}^{(\alpha)}(a) \cdot \vec{\hat{x}}^{(\beta)}(b)}{N} \\
&= C^{(\alpha,\beta)}(t_a, t_b) + \theta_a R^{(\beta,\alpha)}(t_b, t_a) + \theta_b R^{(\alpha,\beta)}(t_a, t_b)
\end{aligned} \tag{3.15}$$

In the second equality, we notice that the term proportional to $\theta_a \theta_b$ is zero because the Lagrange multipliers $\vec{\hat{x}}$ that ensure the Langevin equation cannot correlate. Additionally, in the last equality, we recognize the correlation defined in Eq. (2.6). Moreover, it can be shown that in this setting the reaction can be written as Eq. (3.16), as defined explicitly in reference [20].

$$R^{(\alpha,\beta)}(t_a, t_b) = \frac{1}{N} \sum_{i=0}^{N-1} \frac{\delta \langle x_i^{(\alpha)}(t_a) \rangle}{\delta h^{(\beta)}(t_b)} \Big|_{h^{(\beta)}(t_b)=0} \tag{3.16}$$

From this point forward, we adopt the simplified notation $R^{(\alpha,\beta)}(t_a, t_b) = R_{ab}$, $C^{(\alpha,\beta)}(t_a, t_b) = C_{ab} = C_{ba}$, and $Q^{(\alpha,\beta)}(a, b) = Q_{ab} = Q_{ba}$.

Through the relation of Q_{ab} with the correlations expressed in Eq. (3.15), the saddle-point given by Eq. (3.14) can be interpreted as a Schwinger–Dyson equation [24], which describes the relationship between correlation functions in quantum field theory. This relationship can be expressed more conveniently by multiplying it by Q_{cb} , integrating over c and summing over γ [25], as demonstrated in Eq. (3.17) through an explicit derivation of the $\mathcal{Z}$ term.

$$\begin{aligned}
\sum_{\gamma} \int dc K_{ac} Q_{cb} &= (\theta_a + \theta_b) \delta_{\alpha,\beta} \delta(t_a - t_b) + \varepsilon \sum_{\gamma} \int dc A_{ca} G'(Q_{ac}) Q_{cb} \\
&\quad + \varepsilon p_0^2 \sum_{\epsilon, \kappa, \gamma} \int de dk dc A_{ea} G'(Q_{ac}) A_{ck} Q_{cb}
\end{aligned} \tag{3.17}$$

$$A_{ab} \doteq (I + G)^{-1}(Q_{ab}) = C^{A(\alpha,\beta)}(t_a, t_b) + \theta_a R^{A(\beta,\alpha)}(t_b, t_a) + \theta_b R^{A(\alpha,\beta)}(t_a, t_b) \tag{3.18}$$

This expansion is possible due to the Grassmann properties. In what follows, we use the simplified notation $R^{A(\alpha,\beta)}(t_a, t_b) = R_{ab}^A$ and $C^{A(\alpha,\beta)}(t_a, t_b) = C_{ab}^A = C_{ba}^A$. It can be shown that no term proportional to $\theta_a \theta_b$ survives, similar to the dynamical overlap matrix. By introducing the definition of the kinetic kernel and unfolding the supersymmetric algebra of the functions A , G' and Q into Eq. (3.17) one obtains the dynamical equations for the correlation and reaction functions for $t_a \geq t_b$.

$$\begin{aligned}
\partial_{t_a} C_{ab} = & -\mu^{(\alpha)}(t_a)C_{ab} - \varepsilon \sum_{\gamma} \int_0^{t_b} dt_c R_{bc} C_{ac}^A G'(C_{ac}) \\
& - \varepsilon \sum_{\gamma} \int_0^{t_a} dt_c [C_{bc} R_{ac}^A G'(C_{ac}) + R_{ac} G''(C_{ac}) C_{ac}^A C_{cb}] \\
& + \varepsilon p_0^2 \sum_{\epsilon, \kappa, \gamma} \int_0^{t_a} dt_e \int_0^{t_b} dt_c \int_0^{t_c} dt_k R_{ae}^A R_{ck}^A G'(C_{ac}) R_{bc} \\
& + \varepsilon p_0^2 \sum_{\epsilon, \kappa, \gamma} \int_0^{t_a} dt_e \int_0^{t_a} dt_c \int_0^{t_c} dt_k R_{ae}^A R_{ck}^A R_{ac} G''(C_{ac}) C_{bc}
\end{aligned} \tag{3.19}$$

$$\begin{aligned}
\partial_{t_a} R_{ab} = & -\mu^{(\alpha)}(t_a)R_{ab} + \delta_{\alpha, \beta} \delta(t_a - t_b) - \varepsilon \sum_{\gamma} \int_{t_b}^{t_a} dt_c [R_{ac}^A G'(C_{ac}) R_{cb} + C_{ac}^A G''(C_{ac}) R_{cb} R_{ac}] \\
& + \varepsilon p_0^2 \sum_{\epsilon, \kappa, \gamma} \int_0^{t_a} dt_e \int_{t_b}^{t_a} dt_c \int_0^{t_c} dt_k R_{ae}^A R_{ck}^A G''(C_{ac}) R_{ac} R_{cb}
\end{aligned} \tag{3.20}$$

The equation for the Lagrange multiplier is found by imposing $C^{(\alpha, \alpha)}(t, t) = 1 \rightarrow \partial_t C^{(\alpha, \alpha)}(t, t) = 0$.

$$\begin{aligned}
\mu^{(\tau)}(t) = & \epsilon p_0^2 \sum_{\epsilon, \kappa, \gamma} \int_0^t dt_e \int_0^t dt_c \int_0^{t_c} dt_k R^{A(\tau, \epsilon)}(t, t_e) R^{A(\gamma, \kappa)}(t_c, t_k) R^{(\tau, \gamma)}(t, t_c) \\
& \times \left[G'(C^{(\tau, \gamma)}(t, t_c)) + G''(C^{(\tau, \gamma)}(t, t_c)) C^{(\tau, \gamma)}(t, t_c) \right] \\
& - \varepsilon \sum_{\gamma} \int_0^t dt_c \left[R^{(\tau, \gamma)}(t, t_c) C^{A(\tau, \gamma)}(t, t_c) G'(C^{(\tau, \gamma)}(t, t_c)) \right. \\
& \quad \left. + R^{A(\tau, \gamma)}(t, t_c) G'(C^{(\tau, \gamma)}(t, t_c)) C^{(\tau, \gamma)}(t, t_c) \right. \\
& \quad \left. + R^{(\tau, \gamma)}(t, t_c) C^{A(\tau, \gamma)}(t, t_c) G''(C^{(\tau, \gamma)}(t, t_c)) C^{(\tau, \gamma)}(t, t_c) \right]
\end{aligned} \tag{3.21}$$

Or, setting $G(q) = q^2/2$, it simplifies to

$$\begin{aligned}
\mu^{(\tau)}(t) = & 2\epsilon p_0^2 \sum_{\epsilon, \kappa, \gamma} \int_0^t dt_e \int_0^t dt_c \int_0^{t_c} dt_k R^{A(\tau, \epsilon)}(t, t_e) R^{A(\gamma, \kappa)}(t_c, t_k) R^{(\tau, \gamma)}(t, t_c) C^{(\tau, \gamma)}(t, t_c) \\
& - \varepsilon \sum_{\gamma} \int_0^t dt_c [2R^{(\tau, \gamma)}(t, t_c) C^{A(\tau, \gamma)}(t, t_c) C^{(\tau, \gamma)}(t, t_c) \\
& + R^{A(\tau, \gamma)}(t, t_c) (C^{(\tau, \gamma)}(t, t_c))^2].
\end{aligned} \tag{3.22}$$

To close the system of equations represented by Eq. (3.19) and Eq. (3.20), we need to find explicitly the expressions for C^A and R^A . This is achieved by unfolding the definitions of A and A^{-1} and utilizing the delta relation

$$\sum_{\gamma=0}^{n-1} \int dc A_{ac} A_{cb}^{-1} = \theta_b \delta_{\alpha,\beta} \delta(t_a - t_b). \quad (3.23)$$

One can then conclude

$$\begin{bmatrix} 0 \\ \delta_{\alpha,\beta} \delta(t_a - t_b) \end{bmatrix} = \sum_{\gamma=0}^{n-1} \int dt_c \mathcal{M}_{bc} \begin{bmatrix} C_{ac}^A \\ R_{ac}^A \end{bmatrix}, \quad (3.24)$$

$$\mathcal{M}_{bc} = \begin{bmatrix} \delta_{\gamma,\beta} \delta(t_c - t_b) + R_{bc} G'(C_{bc}) & G(C_{bc}) \\ 0 & \delta_{\gamma,\beta} \delta(t_c - t_b) + R_{cb} G'(C_{bc}) \end{bmatrix}. \quad (3.25)$$

Using these results, we can numerically integrate the equations for the response and correlation functions by inverting the operator in Eq. (3.25).

3.2 Numerical Integration

In this section, we introduce a numerical method for solving the DMFT equations by discretizing time into small intervals. The Lagrange multiplier μ only accounts for the linear order of the spherical constraint in a discrete system; hence, the equations are solved approximately. We employ a discrete time step dt , representing two-point functions as matrices and one-point functions as vectors. Additionally, we must establish a parametrization to manage the replica and the time indexes simultaneously.

Let T denote the number of steps, and n the number of replicas. A two-point function in this interval is described by an $nT \times nT$ matrix, while one-point functions are described by a vector with nT elements. The parametrization is organized in blocks of the same time for all replicas, resembling the organization of a tensor product in the form $(t_a, t_b) \otimes (\alpha, \beta)$ or $(t_a) \otimes (\alpha)$. The time discretization takes $t_a = \tilde{a} dt$, where $\tilde{a} \in [0, T-1]$.

$$\begin{aligned} A^{(\alpha,\beta)}(t_a, t_b) &= A^{(\alpha,\beta)}(\tilde{a}dt, \tilde{b}dt) = A_{\tilde{a}n+\alpha, \tilde{b}n+\beta} \\ v^{(\alpha)}(t_a) &= v^{(\alpha)}(\tilde{a}dt) = v_{\tilde{a}n+\alpha}, \end{aligned} \quad (3.26)$$

where A and v represent generic two-point and one-point functions or operators, respectively. Since $\tilde{a}$ is always paired with α , we simplify the notation as

$$\begin{aligned} A_{a,b} &\doteq A_{\tilde{a}n+\alpha, \tilde{b}n+\beta} \\ v_a &\doteq v_{\tilde{a}n+\alpha}. \end{aligned} \quad (3.27)$$

And the original indexes can be recovered by

$$\begin{aligned} A_{a,b} &= A^{(a\%n, b\%n)}(dt[a//n], dt[b//n]) \\ v_a &= v^{(a\%n)}(dt[a//n]), \end{aligned} \quad (3.28)$$

where $\%$ denotes the modulus and $//$ represents the truncating integer division operators.

It is important to note that with this parametrization, it is not necessary for $R_{ab}R_{ba} = 0$ because the diagonal blocks for equal times are symmetric with respect to the replicas, unlike the previous lower triangular arrangement. However, as explained before, we expect the response for different replicas to be null, resulting in diagonal matrices for each pair of times. Therefore, we can maintain the lower triangular property of the matrix R . The parametrization also does not affect the symmetric property of the matrix C .

The numerical integration of equations Eq. (3.19) and Eq. (3.20) will be carried out using a simple Euler scheme. The process involves discretizing time, which transforms all integrals into sums.

$$\begin{aligned}
C_{(a+n),b} - C_{a,b} = dt & \left[-\mu_a C_{a,b} - \varepsilon dt \sum_{c=0}^{(\tilde{b}+1)n-1} R_{b,c} C_{a,c}^A G'(C_{a,c}) \right. \\
& - \varepsilon dt \sum_{c=0}^{(\tilde{a}+1)n-1} [C_{b,c} R_{a,c}^A G'(C_{a,c}) + R_{a,c} G''(C_{a,c}) C_{a,c}^A C_{c,b}] \\
& + \varepsilon p_0^2 dt^3 \sum_{e=0}^{(\tilde{a}+1)n-1} R_{a,e}^A \sum_{c=0}^{(\tilde{b}+1)n-1} G'(C_{a,c}) R_{b,c} \sum_{k=0}^{(\tilde{c}+1)n-1} R_{c,k}^A \\
& \left. + \varepsilon p_0^2 dt^3 \sum_{e=0}^{(\tilde{a}+1)n-1} R_{a,e}^A \sum_{c=0}^{(\tilde{a}+1)n-1} R_{a,c} G''(C_{a,c}) C_{b,c} \sum_{k=0}^{(\tilde{c}+1)n-1} R_{c,k}^A \right] \quad (3.29)
\end{aligned}$$

$$\begin{aligned}
R_{(a+n),b} - R_{a,b} = -\delta_{a,b} + dt & \left[-\mu_a R_{a,b} - \varepsilon dt \sum_{c=\tilde{b}n}^{(\tilde{a}+1)n-1} [R_{a,c}^A G'(C_{a,c}) R_{c,b} \right. \\
& \left. + C_{a,c}^A G''(C_{a,c}) R_{c,b} R_{a,c}] \right. \\
& \left. + \varepsilon p_0^2 dt^3 \sum_{e=0}^{(\tilde{a}+1)n-1} R_{a,e}^A \sum_{c=\tilde{b}n}^{(\tilde{a}+1)n-1} G''(C_{a,c}) R_{a,c} R_{b,c} \sum_{k=0}^{(\tilde{c}+1)n-1} R_{c,k}^A \right] \quad (3.30)
\end{aligned}$$

As in the continuous case, we have that $t_a \geq t_b$. However, one must be careful in the discretization of equal times to maintain causality. Therefore, still using the Euler scheme, one finds

$$C^{(\alpha,\beta)}(dt(\tilde{a}+1), dt(\tilde{a}+1)) = C^{(\alpha,\beta)}(dt(\tilde{a}+1), dt\tilde{a}) + C^{(\beta,\alpha)}(dt(\tilde{a}+1), dt\tilde{a}) - C^{(\beta,\alpha)}(dt\tilde{a}, dt\tilde{a}) \quad (3.31)$$

which can be written in matrix indexes as

$$C_{(a+n,b+n)} = C_{(a+n,b)} + C_{(b+n,a)} - C_{(b,a)} \quad : \quad b/n = a/n \quad (3.32)$$

These equations are solved if we give explicit values for μ_a , C^A , and R^A . The Lagrange multiplier approximately imposes the spherical constraint by fixing the diagonal elements of matrix C to one,

$$\begin{aligned}
\mu_a = & -\varepsilon dt \sum_{c=0}^{(\bar{a}+1)n-1} (R_{a,c} C_{a,c}^A G'(C_{a,c}) + C_{a,c} R_{a,c}^A G'(C_{a,c}) + R_{a,c} G''(C_{a,c}) C_{a,c}^A C_{a,c}) \\
& + \varepsilon p_0^2 dt^3 \sum_{e=0}^{(\bar{a}+1)n-1} R_{a,e}^A \sum_{c=0}^{(\bar{a}+1)n-1} R_{a,c} (G'(C_{a,c}) + G''(C_{a,c}) C_{a,c}) \sum_{k=0}^{(\bar{c}+1)n-1} R_{c,k}^A.
\end{aligned} \tag{3.33}$$

Finally, one needs to express the discrete version of Eq. (3.24) in a convenient way to be able to perform the inversion.

$$\begin{aligned}
\begin{bmatrix} 0 \\ \frac{\delta_{a,b}}{dt} \end{bmatrix} &= dt \sum_{\gamma=0}^{n-1} \sum_{\tilde{c}=0}^{T-1} \mathcal{M}_{bc} \begin{bmatrix} C_{ac}^A \\ R_{ac}^A \end{bmatrix} \\
&= dt \sum_{c=0}^{nT-1} \mathcal{M}_{bc} \begin{bmatrix} C_{ac}^A \\ R_{ac}^A \end{bmatrix},
\end{aligned} \tag{3.34}$$

In the last equality, the unique transformation described by Eq. (3.27) enables this simplification. Let us multiply the matrix $\mathcal{M}_{bc}$ by dt and index its entries with superscripts $[\mathcal{M}_{bc}]_{00} = \mathcal{M}_{bc}^1$, $[\mathcal{M}_{bc}]_{01} = \mathcal{M}_{bc}^2$, $[\mathcal{M}_{bc}]_{10} = \mathcal{M}_{bc}^3$, and $[\mathcal{M}_{bc}]_{11} = \mathcal{M}_{bc}^4$. By considering all possible values of $b \in [0, nT-1]$, the left-hand side of Eq. (3.34) becomes a vector of size $2nT$, with a single non-zero value for each given value of a . Meanwhile, the right-hand side can be expressed as a simple matrix-vector multiplication.

$$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \frac{\delta_{a,0}}{dt} \\ \vdots \\ \frac{\delta_{a,(nT-1)}}{dt} \end{bmatrix} = \begin{bmatrix} \begin{bmatrix} \mathcal{M}_{00}^1 & \cdots & \mathcal{M}_{0,(nT-1)}^1 \\ \vdots & \ddots & \vdots \\ \mathcal{M}_{(nT-1),0}^1 & \cdots & \mathcal{M}_{(nT-1),(nT-1)}^1 \end{bmatrix} & \begin{bmatrix} \mathcal{M}_{00}^2 & \cdots & \mathcal{M}_{0,(nT-1)}^2 \\ \vdots & \ddots & \vdots \\ \mathcal{M}_{(nT-1),0}^2 & \cdots & \mathcal{M}_{(nT-1),(nT-1)}^2 \end{bmatrix} \\ \begin{bmatrix} \mathcal{M}_{00}^3 & \cdots & \mathcal{M}_{0,(nT-1)}^3 \\ \vdots & \ddots & \vdots \\ \mathcal{M}_{(nT-1),0}^3 & \cdots & \mathcal{M}_{(nT-1),(nT-1)}^3 \end{bmatrix} & \begin{bmatrix} \mathcal{M}_{00}^4 & \cdots & \mathcal{M}_{0,(nT-1)}^4 \\ \vdots & \ddots & \vdots \\ \mathcal{M}_{(nT-1),0}^4 & \cdots & \mathcal{M}_{(nT-1),(nT-1)}^4 \end{bmatrix} \end{bmatrix} \begin{bmatrix} C_{a,0}^A \\ \vdots \\ C_{a,(nT-1)}^A \\ R_{a,0}^A \\ \vdots \\ R_{a,(nT-1)}^A \end{bmatrix}$$

By observing the individual elements, we can identify some simplifications. Specifically, when $b > c$, we have $\mathcal{M}_{bc}^1 = \mathcal{M}_{cb}^1$ and $\mathcal{M}_{bc}^1 = \mathcal{M}_{bc}^4 = 0$. Additionally, $\mathcal{M}_{bc}^3 = 0$ is symmetric and $\mathcal{M}_{bc}^2$ for all values of b and c . Therefore, we can transform this matrix to be upper triangular, which is advantageous for the numerical computation of the inversion. This is because the system can then be computed recursively. With this transformation, we can rewrite Eq. (3.34) as $\vec{w}(a) = \Gamma \vec{v}(a)$ to obtain C^A and R^A for any given value of a .

$$w_i(a) = \sum_{j=0}^{2nT-1} \Gamma_{ij} v_j(a) \tag{3.35}$$

$$v_j(a) = \begin{cases} C_{a,nT-j-1}^A & j < nT \\ R_{a,j \% nT}^A & j \geq nT \end{cases} \tag{3.36}$$

$$w_i(a) = \begin{cases} 0 & i < nT \\ \frac{\delta_{a,i\%nT}}{dt} & i \geq nT \end{cases} \quad (3.37)$$

$$\Gamma_{ij} = \begin{cases} \delta_{i,j} + dt R_{nT-1-i\%nT, nT-1-j\%nT} G'(C_{nT-1-i\%nT, nT-1-j\%nT}) & i \leq j < nT \\ \delta_{i,j} + dt R_{i\%nT, j\%nT} G'(C_{i\%nT, j\%nT}) & j \geq i \geq nT \\ dt G(C_{nT-1-i\%nT, j\%nT}) & i < nT, j \geq nT \\ 0 & i > j \end{cases} \quad (3.38)$$

In the context of solving the equations, the use of this particular structure guarantees causality. This means that the values for a specific time can be obtained without relying on future realizations. For each value of a , the corresponding row vectors in the matrices C^A and R^A can be extracted and used as inputs in (3.29) and Eq. (3.30) to perform the integration for future times.

Chapter 4

Results

To explore the dynamical evolution of replica configurations, we initialize n replicas with a fixed mutual correlation q . The initial correlation matrix contains ones on the diagonal (normalization) and q in all off-diagonal entries, ensuring that replicas start from nearby points in the high-dimensional phase space. This setup is analogous to initializing different neural network trainings from correlated weights, thereby allowing us to probe how memory of shared structure persists under gradient-based dynamics.

The prescribed initial correlation is implemented by constructing each replica $\vec{x}^{(\gamma)}$ on the N -dimensional hypersphere (with $|\vec{x}^{(\gamma)}|^2 = N$) as a combination of a shared direction $\vec{\xi}$ and replica-specific orthogonal fluctuations $\vec{\eta}^{(\gamma)}$:

$$\vec{x}^{(\gamma)} = \sqrt{N} \left(\sqrt{q} \vec{\xi} + \sqrt{1-q} \vec{\eta}^{(\gamma)} \right). \quad (4.1)$$

Here, $\vec{\xi}$ is a fixed unit vector common to all replicas, while the $\vec{\eta}^{(\gamma)}$ are unit vectors orthogonal to $\vec{\xi}$ and mutually orthogonal across replicas. This construction guarantees that the scalar product between distinct replicas satisfies $\vec{x}^{(\alpha)} \cdot \vec{x}^{(\beta)} / N = q$, thereby realizing the desired overlap.

As a consistency check, we first consider the limiting case $q = 0$, where replicas are initialized independently. In this regime, the dynamics reduce to those of a single replica: intra-replica correlations remain finite, while inter-replica correlations vanish. This expectation is confirmed in Fig. 4.1, where both the correlation function and the Lagrange multiplier μ reproduce the single-replica behavior reported in Ref. [11].

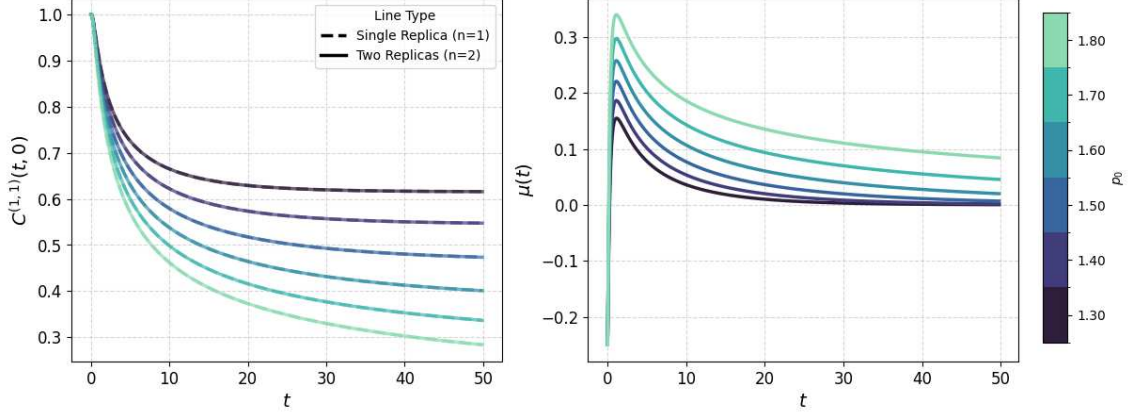


Figure 4.1: Same-replica correlation function (left) and Lagrange multiplier μ (right) for different values of p_0 , with uncorrelated initial conditions ($q = 0$). Results match the single-replica case [11], as expected from the single-replica analysis.

We now turn to finite initial correlations. Using the numerical integration scheme outlined in Sec. 3.2, we compute cross-replica observables for various values of q , with parameters fixed at $N = 500$, $\Delta t = 0.025$, and $p_0 = 1.4$. Two quantities are particularly informative: the cross-correlation at unequal times, $C^{(1,2)}(t, 0)$, and at equal times, $C^{(1,2)}(t, t)$. The former probes how much memory a replica retains of another's initial condition, while the latter reflects their degree of dynamical alignment at time t .

Figure 4.2 shows that $C^{(1,2)}(t, 0)$ decays from its initial value but saturates at a non-zero plateau. This indicates that replicas retain partial memory of their shared initialization even after long times.

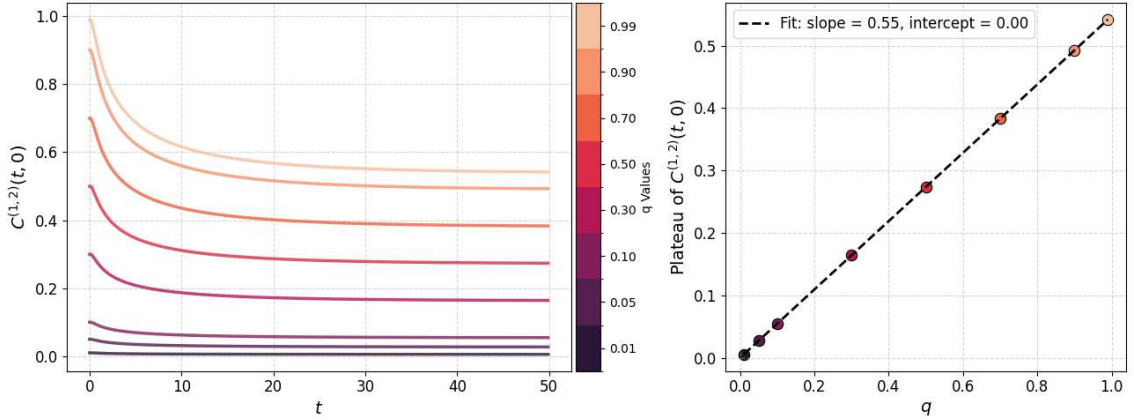


Figure 4.2: Right: Time evolution of $C^{(1,2)}(t, 0)$ for various initial correlations q . Left: Long-time plateau values as a function of q . All results obtained for $p_0 = 1.4$.

A similar picture emerges for $C^{(1,2)}(t, t)$, which also reaches a non-zero plateau at long times (Fig. 4.3). Plotting the plateau values against q in log-log scale reveals an exponent of approximately 1.04, suggesting an almost linear scaling with a mild deviation from linearity. This near-linear response can be interpreted as a marginal Lyapunov-like

sensitivity: replicas remain strongly correlated, in contrast to chaotic systems where small perturbations are exponentially amplified. The persistence of alignment indicates that the dynamics explore the landscape in a structured way, with long-time memory of initial overlaps. A natural extension of this work would be to investigate the energy barriers between solutions along geodesic pathways, as recently studied in neural network models [26], [27]. The simplicity of the present model makes such an analysis accessible both through simulations and potentially through an analytical treatment.

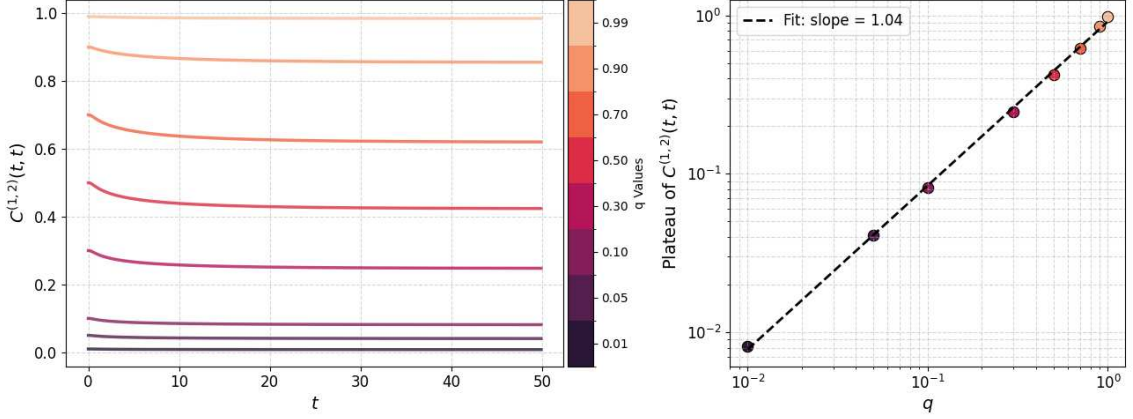


Figure 4.3: Right: Evolution of $C^{(1,2)}(t, t)$ for different initial correlations q . Left: Plateau values plotted in log-log scale vs. q , revealing a scaling exponent ≈ 1.04 .

In summary, while the dynamics drive each replica toward relaxation, cross-replica observables retain a lasting imprint of the initial configuration. The persistence and scaling of these correlations show that memory of initial conditions is not completely lost. These findings highlight that the relaxation dynamics preserve structured correlations across replicas, providing a controlled setting to study how gradient descent explores complex non-convex landscapes.

Chapter 5

Conclusions

In this work, we revisited a simplified mean-field model originally proposed to describe the rigidity transition in biological tissues and applied it as a framework to explore high-dimensional energy landscapes typical of continuous constraint satisfaction problems (CC-SPs). We derived the dynamical mean-field theory (DMFT) equations using the Martin-Siggia-Rose-Janssen-De Dominicis formalism and implemented a discretization scheme to solve them numerically. This allowed us to compute time-dependent correlation and response functions, offering an exact description of gradient descent (GD) dynamics in the overparametrized regime.

A central contribution of this study is the introduction of correlated replicas, which enabled us to probe how initially similar configurations evolve under GD. We found that replicas retain a finite memory of their initial overlap throughout the relaxation dynamics. Specifically, the cross-correlation function $C^{(1,2)}(t, 0)$ converges to a plateau that increases linearly with the initial correlation q . In contrast, the equal-time correlation $C^{(1,2)}(t, t)$ displays an approximately linear scaling with q , suggesting a Lyapunov-like sensitivity to initial conditions. These results indicate that the relaxation dynamics of the system do not fully erase the memory of the initial configurations, highlighting the non-trivial structure of the landscape and the dynamical stability of the canyon manifold.

Several promising directions emerge from this work. First, the replicated model could provide direct access to energy barriers between solutions, for example, by computing barrier heights along geodesic paths. The simplicity of the present model makes such an extension accessible both through simulations and, potentially, via analytical treatment. Second, the replica framework naturally suggests a route to study stochastic gradient descent (SGD), where mini-batches could be modeled as subsets of replicas. This perspective may clarify the origin of the empirical efficiency of SGD and its superiority over standard GD in navigating high-dimensional non-convex landscapes.

In summary, the Canyon model provides an analytically tractable and numerically solvable framework to investigate complex energy landscapes and the dynamics of optimization algorithms. Our findings demonstrate that gradient descent dynamics preserve a finite memory of initial correlations, underscoring the structured nature of relaxation in high dimensions. Beyond its immediate results, this work provides a foundation for studying energy barriers between solutions, stochastic dynamics, and the geometry of solution manifolds, with implications for both statistical physics and machine learning.

Bibliography

- [1] M. Mézard, G. Parisi, and R. Zecchina, “Analytic and algorithmic solution of random satisfiability problems,” *Science*, vol. 297, no. 5582, pp. 812–815, 2002.
- [2] R. Monasson, R. Zecchina, S. Kirkpatrick, B. Selman, and L. Troyansky, “2+ p-sat: Relation of typical-case complexity to the nature of the phase transition,” *Random Structures & Algorithms*, vol. 15, no. 3-4, pp. 414–435, 1999.
- [3] C. S. O’hern, L. E. Silbert, A. J. Liu, and S. R. Nagel, “Jamming at zero temperature and zero applied stress: The epitome of disorder,” *Physical Review E*, vol. 68, no. 1, p. 011 306, 2003.
- [4] S. Franz, G. Parisi, M. Sevelev, P. Urbani, and F. Zamponi, “Universality of the sat-unsat (jamming) threshold in non-convex continuous constraint satisfaction problems,” *SciPost Physics*, vol. 2, no. 3, p. 019, 2017.
- [5] P. Charbonneau, J. Kurchan, G. Parisi, P. Urbani, and F. Zamponi, “Glass and jamming transitions: From exact results to finite-dimensional descriptions,” *Annual Review of Condensed Matter Physics*, vol. 8, pp. 265–288, 2017.
- [6] E. Gardner and B. Derrida, “Optimal storage properties of neural network models,” *Journal of Physics A: Mathematical and general*, vol. 21, no. 1, p. 271, 1988.
- [7] S. Franz and G. Parisi, “The simplest model of jamming,” *Journal of Physics A: Mathematical and Theoretical*, vol. 49, no. 14, p. 145 001, 2016.
- [8] A. Montanari and P. Urbani, “Dynamical decoupling of generalization and overfitting in large two-layer networks,” *arXiv preprint arXiv:2502.21269*, 2025.
- [9] P. Urbani, “A continuous constraint satisfaction problem for the rigidity transition in confluent tissues,” *Journal of Physics A: Mathematical and Theoretical*, vol. 56, no. 11, p. 115 003, 2023.
- [10] P. Urbani, “Statistical physics of complex systems: Glasses, spin glasses, continuous constraint satisfaction problems, high-dimensional inference and neural networks,” *arXiv preprint arXiv:2405.06384*, 2024.
- [11] P. J. Kamali and P. Urbani, “Dynamical mean field theory for models of confluent tissues and beyond,” *arXiv preprint arXiv:2306.06420*, 2023.
- [12] A. G. Fletcher, M. Osterfield, R. E. Baker, and S. Y. Shvartsman, “Vertex models of epithelial morphogenesis,” *Biophysical journal*, vol. 106, no. 11, pp. 2291–2304, 2014.

- [13] D. Bi, X. Yang, M. C. Marchetti, and M. L. Manning, “Motility-driven glass and jamming transitions in biological tissues,” *Physical Review X*, vol. 6, no. 2, p. 021 011, 2016.
- [14] P.-F. Lenne and V. Trivedi, “Sculpting tissues by phase transitions,” *Nature Communications*, vol. 13, no. 1, p. 664, 2022.
- [15] C. Guillot and T. Lecuit, “Mechanics of epithelial tissue homeostasis and morphogenesis,” *Science*, vol. 340, no. 6137, pp. 1185–1189, 2013.
- [16] S. Sinha, A. N. Malmi-Kakkada, X. Li, H. S. Samanta, and D. Thirumalai, “Spatially heterogeneous dynamics of cells in a growing tumor spheroid: Comparison between theory and experiments,” *Soft matter*, vol. 16, no. 22, pp. 5294–5304, 2020.
- [17] E. Blauth, H. Kubitschke, P. Gottheil, S. Grosser, and J. A. Käs, “Jamming in embryogenesis and cancer progression,” *Frontiers in Physics*, vol. 9, p. 666 709, 2021.
- [18] M. Merkel and M. L. Manning, “A geometrically controlled rigidity transition in a model for confluent 3d tissues,” *New Journal of Physics*, vol. 20, no. 2, p. 022 002, 2018.
- [19] E. Gardner, “Spin glasses with p-spin interactions,” *Nuclear Physics B*, vol. 257, pp. 747–765, 1985.
- [20] L. F. Cugliandolo, “Dynamics of glassy systems,” *arXiv preprint cond-mat/0210312*, 2002.
- [21] C. De Dominicis, “Dynamics as a substitute for replicas in systems with quenched random impurities,” *Physical Review B*, vol. 18, no. 9, p. 4913, 1978.
- [22] P. C. Martin, E. Siggia, and H. Rose, “Statistical dynamics of classical systems,” *Physical Review A*, vol. 8, no. 1, p. 423, 1973.
- [23] J. Kurchan, “Supersymmetry in spin glass dynamics,” *Journal de Physique I*, vol. 2, no. 7, pp. 1333–1352, 1992.
- [24] J. Zinn-Justin, *Quantum field theory and critical phenomena*. Oxford university press, 2021, vol. 171.
- [25] F. Mignacco, “Statistical physics insights on the dynamics and generalisation of artificial neural networks,” Ph.D. dissertation, Université Paris-Saclay, 2022.
- [26] F. Draxler, K. Veschgini, M. Salmhofer, and F. Hamprecht, “Essentially no barriers in neural network energy landscape,” in *Proceedings of the 35th International Conference on Machine Learning*, J. Dy and A. Krause, Eds., ser. Proceedings of Machine Learning Research, vol. 80, PMLR, Oct. 2018, pp. 1309–1318. [Online]. Available: <https://proceedings.mlr.press/v80/draxler18a.html>.
- [27] B. L. Annesi, C. Lauditi, C. Lucibello, *et al.*, “Star-shaped space of solutions of the spherical negative perceptron,” *Physical Review Letters*, vol. 131, no. 22, p. 227 301, 2023.