A Hamilton-Jacobi approach to the stochastic block model

by

Tomas Dominguez Chiozza

A thesis submitted in conformity with the requirements
for the degree of Doctor of Philosophy

Graduate Department of Mathematics
University of Toronto

A Hamilton-Jacobi approach to the stochastic block model

Tomas Dominguez Chiozza
Doctor of Philosophy

Graduate Department of Mathematics
University of Toronto
2024

Abstract

This thesis addresses the problem of recovering the community structure in the stochastic block model with two communities. The stochastic block model is a random graph model with planted clusters widely employed as the canonical model to study clustering and community detection. The focus is on the fundamental limits of community detection, quantified by the asymptotic mutual information between the observed network and the actual community structure. This mutual information is studied using the Hamilton-Jacobi approach, pioneered by Jean-Christophe Mourrat.

The first contribution of this thesis is a detailed description of the Hamilton-Jacobi approach, and its application to computing the limit of the mutual information in the dense stochastic block model, where the average degree of a node diverges with the total number of nodes. The main novelty is a well-posedness theory for Hamilton-Jacobi equations on positive half-space that leverages the monotonicity of the non-linearity to circumvent the imposition of an artificial boundary condition as previously done in the literature.

The second contribution of this thesis is a novel well-posedness theory for an infinite-dimensional Hamilton-Jacobi equation posed on the set of non-negative measures and with a monotonic non-linearity. Such an infinite-dimensional Hamilton-Jacobi equation appears naturally when applying the Hamilton-Jacobi approach to the sparse stochastic block model, where the total number of nodes diverges while the average degree of a node remains bounded. The solution to the infinite-dimensional Hamilton-Jacobi equation is defined as the limit of the solutions to an approximating family of finite-dimensional Hamilton-Jacobi equations on positive half-space. In the special setting of a convex non-linearity, a Hopf-Lax variational representation of the solution is also established.

The third contribution of this thesis is a conjecture for the limit of the mutual information in the sparse stochastic block model, and a proof that this conjectured limit provides a lower bound for the asymptotic mutual information. In the case when links across communities are more likely than links within communities, the asymptotic mutual information is known to be given by a variational formula. It is also shown that the conjectured limit coincides with this formula in this case.

Acknowledgements

I must first extend my sincere gratitude to my advisor, Dmitry Panchenko, whose guidance, expertise, and unwavering support have been instrumental in completing this thesis. Dmitry played a crucial role in my growth as a mathematician and was always able to point me in the right direction when I faced challenges. Moreover, his encouragement and trust allowed me to pursue two enriching internships during my doctoral journey, experiences that have significantly influenced my professional path. I will be forever grateful for the freedom, mentorship, and support Dmitry has kindly offered me over these past four years.

I would also like to extend my heartfelt appreciation to Jean-Christophe Mourrat, with whom I have co-authored the two papers and the book upon which this thesis is founded. Jean-Christophe's exceptional ability to provide clear direction and constructive feedback has been instrumental in shaping my academic and personal growth. Over the past three years, he has generously given me more time and attention than I could have ever hoped, guiding me in the exploration of the mathematical unknown. He has been more than a second supervisor, a collaborator, and a mentor; he has been a true friend.

I would also like to thank Jacopo De Simoi, whose part in bringing me to the University of Toronto and guiding my initial steps in this PhD cannot be overstated. Without Jacopo's support and mentorship, I might never have embarked on this doctoral journey. At the University of Toronto, I would like to thank Almut Burchard, Fabio Pusateri, Jeremy Quastel, John Bland, and the rest of the professors I was fortunate enough to learn from. I would also like to thank Jemima Merisca, Sonja Injac, Diana Leonardo, Ashley Armogan, Cherylyn Stina, and all the staff for their hard work.

I want to give special thanks to Raj Patel for his invaluable help and support to make my transition away from academia as smooth as possible. He played a vital role in my professional development, assisting me in securing my first internship and providing me with the guidance, flexibility, and time I needed to secure my second. I am truly grateful for our numerous enriching discussions over the years, and for the genuine friendship that has grown as a result.

I owe a lot to my time at EPFL, which laid the foundation for my academic and professional journey. Benoît, Jesse, Mariana, and Tristan, thank you for the unforgettable memories and for staying in touch despite the distance. André, a special thanks to you for putting up with me, and for being a constant presence and source of support since the day you walked into Gornall's class with your oversized bag all those years ago.

Finally, words cannot express my gratitude to my parents. They have been the rock on which I have built who I am. My achievements are a reflection of their love and encouragement.

Contents

Chapter 1

Introduction

This thesis is centred around the Hamilton-Jacobi approach pioneered by Jean-Christophe Mourrat [50, 83, 84, 85, 86, 87, 88] for studying the behaviour of mean-field systems with disordered interactions. This novel technique is applied to the problem of recovering the community structure in the stochastic block model with two communities. The focus is on the fundamental limits of community detection, analyzed through the lens of the asymptotic mutual information between the observed network and the actual community structure. The asymptotic value of the mutual information has been computed in the dense regime [11, 45, 68], where the average degree of a node diverges with the total number of nodes, and in the sparse disassortative regime [3, 38], where the average degree of a node remains bounded as the total number of nodes diverges and where links across communities are more likely than links within communities. However, its determination has proved more challenging in the assortative sparse regime, where the average degree of a node remains bounded as the total number of nodes diverges and where links within communities are more likely than links across communities. The main contributions of this thesis are a detailed and pedagogical description of the Hamilton-Jacobi approach through its application to computing the limit of the mutual information in the dense regime, a conjecture for the limit of the mutual information in the sparse regime, which is expressed in terms of an infinite-dimensional Hamilton-Jacobi equation posed over a space of probability measures, a well-posedness theory for infinite-dimensional Hamilton-Jacobi equations of this form, a proof that the conjectured limit provides a lower bound for the asymptotic mutual information both in the assortative and disassortative settings, and a proof that the conjectured limit coincides with the known variational formula in the disassortative setting.

1.1 The community detection problem and the stochastic block model

The basic community detection problem consists in partitioning the vertices of a graph into clusters that are more densely connected. A classical real-world example due to Adamic and Glance [6] consists in classifying blogs about US politics into Democrat and Republican leaning by observing only which blog refers to which other blog via a hyperlink. To abstract this problem, one can build a graph of interactions among the blogs, connecting two blogs, or nodes of the graph, via an edge if there is a hyperlink between them — for simplicity, the direction of the hyperlink can be ignored. The classification task then consists in colouring each node blue or red depending on whether it is Democrat or Republican leaning. Intuitively, the graph of interactions should carry meaningful information about the underlying community structure since hyperlinks are more likely between blogs of the same political inclination. A graphical representation of the community detection problem is provided in Figure 1.1.

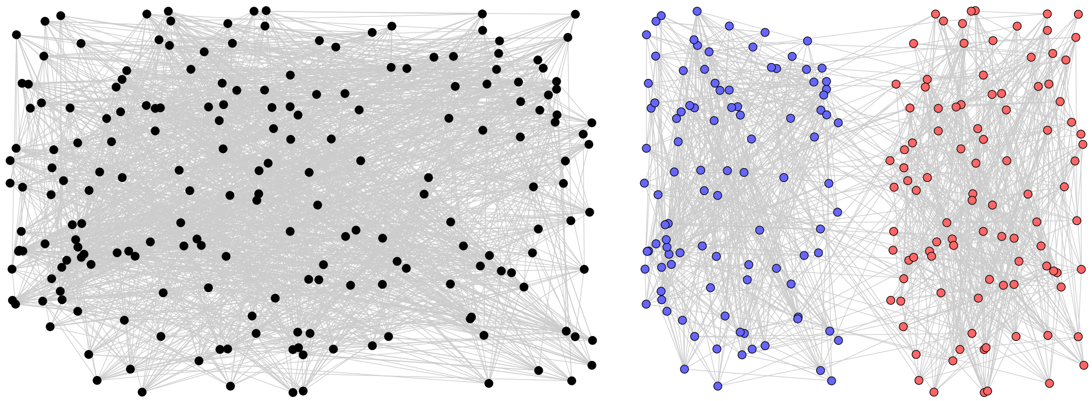

Figure 1.1: The two figures display the same graph of interactions with 100 red nodes and 100 blue nodes, and with each node having on average 10 links with nodes of the same colour and 1 link with a node of a different colour (all edges are independent). On the left figure, the blue and the red nodes have been placed uniformly at random over the entire area, and the colours have been concealed. On the right figure, the blue and the red nodes have been classified by placing them randomly to the left and to the right of the area respectively. In the community detection problem, the left figure is shown to the statistician, whose goal is to infer the colouring of the nodes.

The stochastic block model is the simplest generative model for networks with a community structure. It was first introduced in the machine learning and statistics literature [56, 63, 105, 106] but soon emerged independently in a variety of other scientific disciplines. In the theoretical computer science community, it is often termed the planted partition model [19, 21, 51] while the mathematics literature often refers to it as the inhomogeneous random graph model [18]. Since its introduction, the stochastic block model has become a test bed for clustering and community detection algorithms used in social networks [89], protein-to-protein interaction networks [37], recommendation systems [71], medical prognosis [103], DNA folding [22], image segmentation [102], and natural language processing [10] among others. In this thesis the focus is on the stochastic block model with two communities now described.

Consider N individuals, each belonging to exactly one of two communities. The individuals are encoded as elements of $\{1,\dots,N\}$, and the community structure is represented using a vector

$$\sigma^* := (\sigma_1^*,\dots,\sigma_N^*) \in \Sigma_N := \{-1,+1\}^N. \tag{1.1}$$

It is understood that individuals i and j belong to the same community if and only if $\sigma_i^* = \sigma_j^*$. The labels $(\sigma_i^*)_{i\leqslant N}$ are sampled independently from a Bernoulli distribution P_1 with probability of success $p \in (0,1)$ and expectation $\overline{m}$,

$$p := P_1\{1\} = \mathbb{P}\{\overline{\sigma}_i = 1\} \quad \text{and} \quad \overline{m} := \mathbb{E}\overline{\sigma}_1 = 2p-1. \tag{1.2}$$

The assignment vector σ^* is thus distributed according to the product law

$$\sigma^* \sim P_N^* := (P^*)^{\otimes N}, \tag{1.3}$$

and the expected sizes of the communities are Np and $N(1-p)$. Using the assignment vector σ^*, a random undirected graph $\mathbf{G}_N := (G_{ij})_{i,j\leqslant N}$ with vertex set $\{1,\dots,N\}$ is constructed by stipulating that an edge between node i and node j is present with conditional probability

$$\mathbb{P}\{G_{ij} = 1 \mid \sigma^*\} := \begin{cases} a_N & \text{if } \sigma_i^* = \sigma_j^*, \\ b_N & \text{if } \sigma_i^* \neq \sigma_j^*, \end{cases} \tag{1.4}$$

for some $a_N, b_N \in (0,1)$, independently of all other edges. In other words, the probability that an edge is present between node i and node j depends only on whether or not the individuals i and j belong to the same community. To express (1.4) more succinctly, it is convenient to introduce the average and the gap of a_N and b_N,

$$c_N := \frac{a_N + b_N}{2} \quad \text{and} \quad \Delta_N := \frac{a_N - b_N}{2} \in (-c_N, c_N), \tag{1.5}$$

in such a way that

$$\mathbb{P}\{G_{ij} = 1 \mid \sigma^*\} = c_N + \Delta_N \sigma_i^* \sigma_j^*. \tag{1.6}$$

The data $\mathbf{G}_N = (G_{ij})_{i,j \leqslant N}$ is said to be sampled from the stochastic block model, and the inference task is to reconstruct the community structure σ^* as best as possible given the observation of the network of interactions $\mathbf{G}_N = (G_{ij})_{i,j \leqslant N}$. When $p = 1/2$, the symmetry between the two communities makes it clear that σ^* can at best be recovered up to a change of sign. In the case when $\Delta_N \leqslant 0$, it is more likely for an edge to be present between nodes in different communities, and the model is called *disassortative*. When $\Delta_N > 0$ connections are more likely between individuals in the same community, and the model is termed *assortative*.

In practice, real-world data can be fitted to the stochastic block model. The example of blogs about US politics discussed above would be fitted to an assortative stochastic block model since hyperlinks are more likely between blogs of the same political inclination. The information-theoretic results of this thesis can then be used to determine whether the data is in a regime where the communities can be reliably recovered. This has important algorithmic implications [5, 76, 100, 101]. It is worth noting that the results of this thesis do not address the question of whether the stochastic block model is an adequate model for the data — they merely provide insights on whether the data can be reliably classified assuming it is generated from this model. A survey of community detection is [1].

1.2 Mutual information in the stochastic block model

Recently, the stochastic block model has attracted much renewed attention. On a practical level, it has, for instance, seen extensions allowing for overlapping communities [7] that have proved to be a good fit for real data sets in massive networks [60]. On a theoretical level, the predictions put forth in [44] using deep but non-rigorous statistical physics arguments have been particularly stimulating. The theoretical study of the stochastic block model has seen significant progress in two main directions: exact recovery and detection. The exact recovery task aims to determine the regimes of a_N and b_N, or equivalently of c_N and Δ_N, for which there exists an algorithm that completely recovers the two communities with high probability, up to a global change of sign. A necessary condition for exact recovery is that the random graph $\mathbf{G}_N$ be connected; this makes exact recovery impossible in the sparse regime. The sharp threshold for exact recovery was obtained in [2, 80], where it was shown that in the symmetric dense regime, $p = 1/2$, $a_N = a\log(N)/N$ and $b_N = b\log(N)/N$, exact recovery is possible, and efficiently so, if and only if $\sqrt{a} - \sqrt{b} \geq 2$. On the other hand, the detection task is to construct a partition of the graph $\mathbf{G}_N$ that is positively correlated with the assignment vector σ^* with high probability, possibly up to a global change of sign. The sharp threshold for detection in the sparse regime was obtained in [74, 78, 81], where it was shown that in the symmetric sparse regime, $p = 1/2$, $a_N = a/N$ and $b_N = b/N$, detection is solvable, and efficiently so, if and only if $(a-b)^2 > 2(a+b)$. Notice that detection is much easier in the asymmetric case [23]. Indeed, the expected degree of node i conditional on its community membership is given by

$$\mathbb{E}[\deg(i)|\sigma_i^*] = (N-1)(c_N + \overline{m}\Delta_N \sigma_i^*), \tag{1.7}$$

so meaningful information about the community structure is revealed from the degrees of nodes.

Despite this clear picture regarding the thresholds for exact recovery and detection in the setting of two communities, several questions remain open. In this thesis, the focus is on the problem of quantifying exactly how much information about the communities can be recovered by observing the graph $\mathbf{G}_N$. This is encoded by the mutual information between the assignment vector σ^* and the random graph $\mathbf{G}_N$,

$$I(\mathbf{G}_N;\sigma^*) := \mathbb{E}\log\frac{\mathbb{P}(\mathbf{G}_N,\sigma^*)}{\mathbb{P}(\mathbf{G}_N)\mathbb{P}(\sigma^*)} = \mathbb{E}\int_{\mathbb{R}^N}\log\left(\frac{\mathrm{d}P_{\sigma^*|\mathbf{G}_N}}{\mathrm{d}P_N^*}(\sigma)\right)\mathrm{d}P_{\sigma^*|\mathbf{G}_N}(\sigma). \tag{1.8}$$

Here $P_{\sigma^*|\mathbf{G}_N}$ denotes the conditional law of σ^* given $\mathbf{G}_N$. The mutual information is intimately related to the relative entropy, see Exercise 4.3 in [50], and it is a measure of the "amount of information" obtained about the vector σ^* by observing $\mathbf{G}_N$. The asymptotic value of this mutual information has been computed in the dense regime [11, 45, 68] and in the sparse disassortative regime [3, 38]. Its determination in the assortative sparse regime has proved far more challenging. After the publication of the results of this thesis [49], this problem was resolved in [107] in the symmetric case, $p = 1/2$, building upon the earlier works [4, 66, 79, 82]. The approach developed there does not generalize well to more complex models such as when more than two communities are present [61]. In contrast, this thesis aims to propose a new approach to the analysis of the community detection problem that would be robust to model modifications.

1.2.1 Mutual information in the dense stochastic block model

The dense stochastic block model refers to the stochastic block model in the regime where the average degree of a node diverges with the number of nodes. In this thesis, this will be encoded by the following assumptions on c_N in (1.5) and on the quantity

$$\lambda_N := \frac{N\Delta_N^2}{c_N(1-c_N)}. \tag{1.9}$$

A1 The sequence $(\lambda_N)_{N\geqslant 1}$ converges to some value $\lambda \geq 0$.

A2 The sequence $(Nc_N(1-c_N))_{N\geqslant 1}$ diverges to infinity.

The second of these assumptions implies that the average degree of a node i is

$$\mathbb{E}\deg(i) = (N-1)(c_N + \overline{m}\Delta_N), \tag{1.10}$$

and therefore diverges with N. In this dense setting, a universality property of the mutual information (1.8) makes it possible to understand the information-theoretic properties of the stochastic block model by mapping it to a symmetric rank-one matrix estimation problem. This symmetric rank-one matrix estimation problem has been widely studied, with the most general results obtained in [36] by leveraging the Hamilton-Jacobi approach. Alternative methods that so far have not reached the level of generality of [36] include [12, 13, 16, 46, 52, 68, 69]. To state these results, it will be convenient to fix a standard Gaussian random variable z and introduce the function $\psi : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ defined by

$$\psi(h) := \mathbb{E} \log \int_{\Sigma_1} \exp\left(\sqrt{2h}\sigma z + 2h\sigma\sigma_1^* - h\right) \mathrm{d}P^*(\sigma). \tag{1.11}$$

Mapping the results of [36] back to the dense stochastic block model gives the following variational formula for the limit of the mutual information.

Theorem 1.1. *Under assumptions* (**A1**) *and* (**A2**)*, the limit of the mutual information* (1.8) *in the dense stochastic block model admits the variational representation*

$$\lim_{N\to+\infty} \frac{1}{N} I(\mathbf{G}_N;\sigma^*) = \frac{\lambda}{4} - \sup_{h\geq 0}\left(\psi(h) - \frac{h^2}{\lambda}\right). \tag{1.12}$$

The main difference between the approach taken in this thesis to prove Theorem 1.1 and that proposed in [36] is the solution theory for the Hamilton-Jacobi equation that will describe the limit of the mutual information. Instead of imposing an artificial boundary condition, the results in [33] will be used to ignore the boundary altogether and establish a robust well-posedness theory for Hamilton-Jacobi equations on positive half-space. This well-posedness theory plays an important role in the analysis of the more interesting sparse stochastic block model.

1.2.2 Mutual information in the sparse stochastic block model

The sparse stochastic block model refers to the stochastic block model in the regime where the average degree of a node remains bounded as the number of nodes diverges. In this thesis, this will be encoded by the assumption that c_N and Δ_N in (1.5) are given by

$$c_N := \frac{c}{N} \quad \text{and} \quad \Delta_N := \frac{\Delta}{N} \tag{1.13}$$

for some $c > 0$ and some non-zero $\Delta \in (-c, c)$. The case $\Delta = 0$ is trivial since it corresponds to the case where the graph $\mathbf{G}_N$ and the assignment vector σ^* are independent. The probability (1.6) that an edge is present between node i and node j becomes

$$\mathbb{P}\{G_{ij} = 1 | \sigma^*\} = \frac{c + \Delta\sigma_i^*\sigma_j^*}{N} \tag{1.14}$$

for the family of conditionally independent Bernoulli random variables $\mathbf{G}_N = (G_{ij})_{i,j\leqslant N}$. The expected degree of any node i remains bounded with N,

$$\mathbb{E}\deg(i) = \frac{N-1}{N}\left(c + \Delta\overline{m}^2\right), \tag{1.15}$$

so the stochastic block model is indeed in the sparse regime.

The limit of the mutual information (1.8) in the symmetric disassortative setting, $\Delta \leqslant 0$ and $p = 1/2$, was first established in the seminal paper [38]. A more direct proof in the disassortative setting allowing for arbitrary $p \in (0,1)$ was then obtained in [14] using an interpolation argument and a cavity computation. To state this result concisely, denote by

$$\mathcal{M}_p := \left\{ \mu \in \Pr[-1,1] \mid \int_{-1}^{1} x \, \mathrm{d}\mu = \overline{m} \right\} \tag{1.16}$$

the set of probability measures with mean $\overline{m} = 2p - 1$, and define the functional $\psi : \mathcal{M}_p \to \mathbb{R}$ by

$$\begin{aligned} \psi(\mu) := -c + p\mathbb{E} \log \int_{\Sigma_1} \exp(-\Delta\sigma\overline{m}) \prod_{x \in \Pi_+(\mu)} (c + \Delta\sigma x) \, \mathrm{d}P^*(\sigma) \\ + (1-p)\mathbb{E} \log \int_{\Sigma_1} \exp(-\Delta\sigma\overline{m}) \prod_{x \in \Pi_-(\mu)} (c + \Delta\sigma x) \, \mathrm{d}P^*(\sigma), \end{aligned} \tag{1.17}$$

where $\Pi_\pm(\mu)$ denotes the Poisson point process with mean measure $(c \pm \Delta x) \, \mathrm{d}\mu(x)$ on $[-1,1]$. For the definition and basic properties of a Poisson point process see Chapter 5 in [50]. The limit of the mutual information may be expressed as a variational formula involving the functional $\mathcal{P} : \mathcal{M}_p \to \mathbb{R}$ defined by

$$\mathcal{P}(\mu) := \psi(\mu) + \frac{c}{2} + \frac{\Delta\overline{m}^2}{2} - \frac{1}{2}\mathbb{E}(c + \Delta x_1 x_2) \log(c + \Delta x_1 x_2), \tag{1.18}$$

where x_1 and x_2 are independent samples from the probability measure μ.

Theorem 1.2. *The limit of the mutual information (1.8) in the disassortative sparse stochastic block model with $\Delta \leqslant 0$ admits the variational representation*

$$\lim_{N \to +\infty} \frac{1}{N} I(\mathbf{G}_N; \sigma^*) = \frac{1}{2}\mathbb{E}\left(c + \Delta\sigma_1^*\sigma_2^*\right) \log\left(c + \Delta\sigma_1^*\sigma_2^*\right) - \frac{c}{2} - \frac{\Delta\overline{m}^2}{2} - \sup_{\mu \in \mathcal{M}_p} \mathcal{P}(\mu). \tag{1.19}$$

The limit of the mutual information (1.8) in the assortative setting, $\Delta > 0$, has been explored in the recent works [4, 61, 66, 79, 82, 107]. Noticing that the graph $\mathbf{G}_N$ locally looks like a tree, these works aim to leverage a connection between community detection and a process of broadcasting on trees. For convenience, the latter problem will be briefly described on a regular tree. First, attribute a random ± 1 variable σ^* to the root node. Then, recursively and independently along each edge, "broadcast" it to each child node by flipping the sign of the spin with some fixed probability $\delta \in (0,1)$. A basic question is to determine the mutual information between the spin σ^* attributed to the root node and the spins on all the nodes at a given depth, in the limit of large depth. A fruitful variant of this question consists of adding a "survey" of all nodes by randomly revealing the spins attached to each node independently with some fixed probability ε. If, in the limit of large depth, the knowledge of the spins on all the leaf vertices does not bring meaningful additional information on σ^* on top of surveying compared with surveying alone, then one can relate the mutual information between σ^* and the survey to the mutual information in the community detection problem; in this case, one may speak of "boundary irrelevance". To decide whether boundary irrelevance holds, one can study the evolution of the log-likelihood ratio between the two hypotheses $\sigma^* = \pm 1$ upon revealing the boundary information at a given depth. One can calculate the law of this quantity recursively as the depth varies by iterating a fixed map called the "BP operator". The property of boundary irrelevance essentially corresponds to this BP operator admitting a unique non-trivial fixed point. Building upon earlier works, it was recently established in [107] that this uniqueness property holds in the setting of two balanced communities, $p = 1/2$. As a byproduct, this yields a full identification of the limit mutual information (1.8) in this case. The uniqueness of a non-trivial fixed point to the BP operator has subsequently been shown to be false in general for models with more than two communities [61]. This means that the approach developed in these recent works does not generalize well to more complex models. In contrast, the Hamilton-Jacobi approach proposed in this thesis is designed to be robust to model modifications. The connections between the results of this thesis and the series of works just discussed will be explored in Section 6.5.

1.3 Main contributions of this thesis

This thesis is based on two papers [48, 49] and a book [50] published by the author and Jean-Christophe Mourrat. The paper [49] was published in the Annals of Probability, and it is an analysis of the sparse stochastic block model using the Hamilton-Jacobi approach. The companion paper [48] is in the late review stages in the SIAM Journal on Mathematical Analysis (SIMA), and it develops the well-posedness theory for infinite-dimensional Hamilton-Jacobi equations required to apply the Hamilton-Jacobi approach to the sparse stochastic block model. The book [50] will be published by the European Mathematical Association, and it is an introductory text on the Hamilton-Jacobi approach for studying the behaviour of mean-field systems with disordered interactions. During his doctoral studies, the author also published a paper on the ℓ^p-Gaussian-Grothendieck problem with vector spins [47] in the Electronic Journal of Probability, and two papers on mathematical finance [96, 97], but chose not to include their contents in this thesis.

To state the main contributions of this thesis concisely, it will be convenient to introduce additional notation. For the main result on the dense stochastic block model, introduce the non-linearity $\overline{\mathsf{H}}(p) := p^2 \mathbf{1}\{p \geqslant 0\}$ and its associated finite-dimensional Hamilton-Jacobi equation,

$$\partial_t f(t,h) - \overline{\mathsf{H}}\big(\partial_h f(t,h)\big) = 0 \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}_{>0}. \tag{1.20}$$

For the results on the sparse stochastic block model, denote by $\mathcal{M}_+$ the cone of finite positive measures on $[-1,1]$,

$$\mathcal{M}_+ := \big\{\mu \mid \mu \text{ is a finite non-negative measure on } [-1,1]\big\}, \tag{1.21}$$

and let $g : [-1,1] \to \mathbb{R}$ be the function defined by

$$g(z) := (c+\Delta z)\big(\log(c+\Delta z) - 1\big) = (c+\Delta z)\log(c) + c\sum_{n\geqslant 2} \frac{(-\Delta/c)^n}{n(n-1)} z^n - c. \tag{1.22}$$

Introduce the cone of functions

$$\mathcal{C}_\infty := \left\{G_\mu : [-1,1] \to \mathbb{R} \mid G_\mu(x) := \int_{-1}^1 g(xy)\,\mathrm{d}\mu(y) \text{ for some } \mu \in \mathcal{M}_+\right\}, \tag{1.23}$$

and the non-linearity $\mathsf{C}_\infty : \mathcal{C}_\infty \to \mathbb{R}$ defined on this cone by

$$\mathsf{C}_\infty(G_\mu) := \frac{1}{2}\int_{-1}^1 G_\mu(x)\,\mathrm{d}\mu(x). \tag{1.24}$$

This non-linearity is well-defined by the Fubini-Tonelli theorem (see equations (3.71) - (3.72)). The equation of interest in the sparse stochastic block model will be the infinite-dimensional Hamilton-Jacobi equation

$$\partial_t f(t,\mu) = \mathsf{C}_\infty(D_\mu f(t,\mu)) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathcal{M}_+, \tag{1.25}$$

where $D_\mu f(t,\mu)$ denotes the Gateaux derivative density of the function $f(t,\cdot)$ (see equations (4.4)- (4.5)). The initial condition associated with this Hamilton-Jacobi will be the functional $\psi : \mathcal{M}_+ \to \mathbb{R}$ defined by

$$\psi(\mu) := -\mu[-1,1]c + p\mathbb{E}\log\int_{\Sigma_1} \exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1) \prod_{x\in\Pi_+(\mu)} (c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)$$
$$+ (1-p)\mathbb{E}\log\int_{\Sigma_1} \exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1) \prod_{x\in\Pi_-(\mu)} (c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma), \tag{1.26}$$

where $\Pi_\pm(\mu)$ denotes the Poisson point process with mean measure $(c\pm\Delta x)\,\mathrm{d}\mu(x)$ on $[-1,1]$ and x_1 is sampled from the probability measure $\overline{\mu} := \frac{\mu}{\mu[-1,1]}$. Notice that this initial condition is an extension of the functional (1.17). The six main contributions of the works [48, 49, 50] discussed in this thesis are the following.

(i) A detailed and pedagogical description of the Hamilton-Jacobi approach through its application to proving Theorem 1.1 on the limit of the mutual information in the dense stochastic block model. This is the content of Chapter 2, and its main result is the following.

Theorem 1.3. *Assuming* (**A1**) *and* (**A2**)*, if f denotes the unique viscosity solution to the Hamilton-Jacobi equation* (1.20) *subject to the initial condition* (1.11)*, then the asymptotic value of the mutual information* (1.8) *is*

$$\lim_{N\to+\infty}\frac{1}{N}I(\mathbf{G}_N;\sigma^*) = \frac{\lambda}{4} - f\left(\frac{\lambda}{4},0\right) = \frac{\lambda}{4} - \sup_{h\geq 0}\left(\psi(h) - \frac{h^2}{\lambda}\right). \tag{1.27}$$

(ii) A conjecture for the limit of the mutual information both in the disassortative and assortative regimes. This conjecture is derived in Chapter 3 and expressed in terms of the infinite-dimensional Hamilton-Jacobi equation (1.25). It reads as follows.

Conjecture 1.4. *If f denotes the unique viscosity solution to the infinite-dimensional Hamilton-Jacobi equation* (1.25) *subject to the initial condition* (1.26)*, then the asymptotic value of the mutual information* (1.8) *is*

$$\lim_{N\to+\infty}\frac{1}{N}I(\mathbf{G}_N;\sigma^*) = \frac{1}{2}\mathbb{E}\left(c+\Delta\sigma_1^*\sigma_2^*\right)\log\left(c+\Delta\sigma_1^*\sigma_2^*\right) - \frac{c}{2} - \frac{\Delta\overline{m}^2}{2} - f(1,0). \tag{1.28}$$

(iii) A well-posedness theory for infinite-dimensional Hamilton-Jacobi equations of this form. This is the content of Chapter 4, and it leads to the following well-posedness result.

Theorem 1.5. *The limit* (6.10) *exists, is finite, and is independent of the parameters R and b. This limit is defined to be the solution to the infinite-dimensional Hamilton-Jacobi equation* (1.25)*. Moreover, if the non-linearity in the infinite-dimensional Hamilton-Jacobi equation* (1.25) *is convex in the sense of* (**H5**)*, then the solution to this equation at the point $(t,\mu)\in\mathbb{R}_{\geq 0}\times\mathcal{M}_+$ admits the Hopf-Lax variational representation*

$$f(t,\mu) = \sup_{\nu\in\Pr[-1,1]}\left(\psi(\mu+t\nu) - \frac{t}{2}\int_{-1}^{1} G_\nu(y)\,\mathrm{d}\nu(y)\right). \tag{1.29}$$

(iv) A finitary version of the main multioverlap result in [15]. This is the content of Chapter 5, and the main results are Propositions 5.2 and 5.4.

(v) A proof that the conjectured limit provides a lower bound for the asymptotic mutual information in both the disassortative and assortative regimes. This is the first main result of Chapter 6, and it reads as follows.

Theorem 1.6. *If f denotes the unique viscosity solution to the infinite-dimensional Hamilton-Jacobi equation* (1.25) *subject to the initial condition* (1.26)*, then the asymptotic value of the mutual information* (1.8) *in the sparse stochastic block model satisfies the lower bound*

$$\liminf_{N\to+\infty}\frac{1}{N}I(\mathbf{G}_N;\sigma^*)\geqslant\frac{1}{2}\mathbb{E}\big(c+\Delta\sigma_1^*\sigma_2^*\big)\log\big(c+\Delta\sigma_1^*\sigma_2^*\big)-\frac{c}{2}-\frac{\Delta\overline{m}^2}{2}-f(1,0). \tag{1.30}$$

(vi) A proof that in the disassortative regime, $\Delta\leqslant 0$, the conjectured limit coincides with the variational formula in Theorem 1.2. This is the second main result of Chapter 6, and it reads as follows.

Theorem 1.7. *If f denotes the unique viscosity solution to the infinite-dimensional Hamilton-Jacobi equation* (1.25) *subject to the initial condition* (1.26)*, then the asymptotic value of the mutual information* (1.8) *in the disassortative sparse stochastic block model admits the Hamilton-Jacobi representation*

$$\lim_{N\to+\infty}\frac{1}{N}I(\mathbf{G}_N;\sigma^*)=\frac{1}{2}\mathbb{E}\big(c+\Delta\sigma_1^*\sigma_2^*\big)\log\big(c+\Delta\sigma_1^*\sigma_2^*\big)-\frac{c}{2}-\frac{\Delta\overline{m}^2}{2}-f(1,0). \tag{1.31}$$

Moreover, the function f at the point $(1,0)\in\mathbb{R}_{\geqslant 0}\times\mathcal{M}_+$ admits the variational representation

$$f(1,0)=\sup_{\mu\in\mathcal{M}_p}\mathcal{P}(\mu), \tag{1.32}$$

where $\mathcal{P}:\mathcal{M}_p\to\mathbb{R}$ denotes the functional defined in (1.18)*.*

These results are obtained by leveraging the Hamilton-Jacobi approach developed by Jean-Christophe Mourrat [83, 84, 85, 86, 87, 88], and formalized in the book [50] by the author and Jean-Christophe Mourrat. The observation underlying the Hamilton-Jacobi approach is that, up to an additive constant that can be computed explicitly, the mutual information (1.8) can be interpreted as the free energy of a statistical physics system which, when appropriately enriched, satisfies a Hamilton-Jacobi equation up to an error term that vanishes in the limit of large system size. This leads to the conjecture that, up to an additive constant, the asymptotic mutual information is the unique solution to a Hamilton-Jacobi equation. In the presence of convexity, the Hopf or Hopf-Lax formulas can be used to represent this solution variationally, and therefore obtain variational formulas such as (1.27) or (1.32) for the asymptotic mutual information.

To render the Hamilton-Jacobi approach rigorous, a well-posedness theory for the Hamilton-Jacobi equation must be developed, and the error term associated with the mutual information must be carefully analyzed and controlled. The inability to prove the matching lower bound in Conjecture 1.4 stems from the relatively weak control of this error term. In particular, it is not expected that this error becomes small as N tends to infinity for each individual choice of the equation's parameters. On the other hand, controlling the error after performing a small averaging over the parameters is possible, but does not suffice for the identification of the limit. As will be shown in the context of the dense stochastic block model, this obstacle can be circumvented under modest convexity assumptions through a convenient "selection principle" which allows the identification of the unique solution to a Hamilton-Jacobi equation. Unfortunately, in the context of the sparse stochastic block model, these convexity conditions do not hold [67] so the matching lower bound in Conjecture 1.4 cannot be obtained by appropriately modifying known techniques. Nonetheless, the result in Theorem 1.7 gives some support in favour of Conjecture 1.4 by showing that, in the disassortative regime, it matches the variational formula obtained in [38] and stated in Theorem 1.2. This lack of convexity also implies that no variational formula is to be expected for the assortative sparse stochastic block model.

The results of this thesis generalize immediately to the case in which the measure P^* is arbitrary with compact support. It is also believed by the author that they generalize without much change to settings with more than two communities, although not every technical detail has been verified.

1.4 Organization of this thesis

This thesis is organized as follows.

In Chapter 2, the Hamilton-Jacobi approach is described, and it is applied to analyze the dense stochastic block model and prove Theorem 1.3. A universality property of the mutual information in the dense regime is leveraged to map the dense stochastic block model to a symmetric rank-one matrix estimation problem. The arguments in [36] are then used to analyze this matrix estimation problem via the Hamilton-Jacobi approach. This chapter draws heavily on Chapters 3 and 4 in [50]. The main novelty relative to [36, 50] is a well-posedness theory for Hamilton-Jacobi equations on positive half-space. This allows consideration of the Hamilton-Jacobi equation appearing in the symmetric rank-one matrix estimation problem directly on its natural domain as opposed to appealing to a symmetrization trick as in [50] or imposing an artificial boundary condition as in [36].

In Chapter 3, the first steps required to apply the Hamilton-Jacobi approach to the sparse stochastic block model are taken, and an appropriately translated enrichment of the mutual information is formally shown to satisfy the infinite-dimensional Hamilton-Jacobi equation (1.25) provided that all multioverlaps concentrate in the limit of large system size. This chapter is taken from Section 2 in [49], and it leads to Conjecture 1.4.

In Chapter 4, the techniques used to establish the well-posedness of Hamilton-Jacobi equations on positive half-space are refined to obtain the well-posedness of infinite-dimensional Hamilton-Jacobi equations posed on the set of non-negative measures and with a monotonic non-linearity. The strategy is to introduce an approximating family of finite-dimensional Hamilton-Jacobi equations and to use the monotonicity of the non-linearity to show that, just like in Chapter 2, no boundary condition needs to be prescribed to establish well-posedness. The solution to the infinite-dimensional Hamilton-Jacobi equation is then defined as the limit of these approximating solutions. In the special setting of a convex non-linearity, a Hopf-Lax variational representation of the solution is also established. This chapter is taken from [48], and it leads to Theorem 1.5.

In Chapter 5 a finitary version of the main result in [15] regarding the concentration of multioverlaps is established. In addition to being finitary, the most notable difference between this multioverlap result and that in [15] is that multioverlap concentration is shown for any perturbation parameter satisfying a condition that may be verified in practice, as opposed to on average over the set of perturbation parameters. This additional control is essential in the proof of Theorem 1.6. This chapter is taken from Appendix C in [49].

Finally, in Chapter 6, Theorems 1.6 and 1.7 are established using the Hamilton-Jacobi approach. Combining ideas from the theory of viscosity solutions with the finitary multioverlap concentration result in Chapter 5, one inequality between the limit mutual information and the translated solution to the infinite-dimensional Hamilton-Jacobi equation (1.25) is proved both in the assortative and disassortative regimes. An interpolation argument taken from [14] and the Hopf-Lax variational formula (1.29) are then used to establish the converse inequality as well as the variational representation (1.32) in the disassortative regime. Although the converse bound is also expected to be valid in the assortative regime, significant technical challenges stand in the way of proving it at the moment. This chapter is taken from [49].

To not disrupt the flow but provide as self-contained a presentation as possible, several basic results in analysis and probability used throughout the thesis are given in Appendix A.

Chapter 2

The dense stochastic block model

In this chapter, the Hamilton-Jacobi approach is introduced and used to compute the limit of the mutual information in the dense stochastic block model. The Hamilton-Jacobi approach was developed by Jean-Christophe Mourrat in the context of statistical mechanics to study the behaviour of mean-field systems with disordered interactions through the lens of the free energy [50, 83, 84, 85, 86, 87, 88]. The key insight underlying the Hamilton-Jacobi approach is that the mutual information in a statistical inference problem such as the dense stochastic block model can be identified with the free energy in a statistical mechanics problem up to an explicit additive constant and that this finite-volume free energy can be shown to satisfy a Hamilton-Jacobi equation up to an error term that vanishes in the limit of large system size. In Section 2.1, the problem of computing the limit of the mutual information in the dense stochastic block model is reformulated using the language of statistical mechanics by introducing a relevant Gibbs measure and free energy. In Section 2.2, the symmetric rank-one matrix estimation problem [12, 13, 16, 36, 46, 52, 68, 69] is described, and a universality property of the free energy is used to relate it to the dense stochastic block model. In Section 2.3, the Gaussian integration by parts formula is leveraged to derive a Hamilton-Jacobi equation posed on positive half-space for an "enrichment" of the free energy in the symmetric rank-one matrix estimation problem. The well-posedness of this Hamilton-Jacobi equation is established in Section 2.4, where the arguments in [33] are used to ignore the boundary of the domain. This is possible because the non-linearity in the Hamilton-Jacobi equation "points in the right direction". With this well-posedness theory at hand, the Hamilton-Jacobi equation for the symmetric rank-one matrix estimation problem is revisited in Section 2.5, and it is shown that the limit of the enriched free energy is its unique solution. The Hopf-Lax variational formula then yields a variational representation for the limit free energy in the symmetric rank-one matrix estimation problem which can be combined with the results in Sections 2.2 and 2.3 to establish Theorem 1.3. Unfortunately, for more general models such as those studied in [36], the supersolution criterion for the Hamilton-Jacobi equation cannot be verified directly. To ensure the generalizability of the Hamilton-Jacobi approach described in this chapter, in Section 2.6, a selection principle for Hamilton-Jacobi equations is developed. This selection principle ensures that a convex function that satisfies a Hamilton-Jacobi equation on a dense set must satisfy the equation everywhere, and it is applied in Section 2.7 to provide an alternative proof of Theorem 1.3. The contents of this chapter rely heavily on Chapters 3 and 4 in [50] which are in turn based on [36].

2.1 From statistical inference to statistical mechanics

The community detection problem associated with the dense stochastic block model consists in recovering the assignment vector

$$\sigma^* := (\sigma_1, \ldots, \sigma_N) \in \Sigma_N := \{-1,+1\}^N \tag{2.1}$$

given the random undirected graph $\mathbf{G}_N := (G_{ij})_{i,j\leqslant N}$ with vertex set $\{1,\ldots,N\}$ constructed by stipulating that an edge between node i and node j is present with conditional probability

$$\mathbb{P}\{G_{ij} = 1 \mid \sigma^*\} = c_N + \sigma_i^* \sigma_j^* \Delta_N \tag{2.2}$$

for some c_N and Δ_N satisfying assumptions (**A1**) and (**A2**). Recall that the labels $\sigma_i^* \sim P^*$ are taken to be i.i.d. Bernoulli random variables with probability of success $p \in (0,1)$ and expectation $\overline{m}$,

$$p := P^*(1) = \mathbb{P}\{\sigma_i^* = 1\} \quad \text{and} \quad \overline{m} := \mathbb{E}\sigma_1^* = 2p-1. \tag{2.3}$$

This means that the assignment vector σ^* follows a product distribution,

$$\sigma^* \sim P_N^* := (P^*)^{\otimes N}. \tag{2.4}$$

To understand the mutual information

$$I(\mathbf{G}_N;\sigma^*) := \mathbb{E}\log\frac{\mathbb{P}(\mathbf{G}_N,\sigma^*)}{\mathbb{P}(\mathbf{G}_N)\mathbb{P}(\sigma^*)} = \mathbb{E}\int_{\mathbb{R}^N}\log\left(\frac{\mathrm{d}P_{\sigma^*|\mathbf{G}_N}}{\mathrm{d}P_N^*}(\sigma)\right)\mathrm{d}P_{\sigma^*|\mathbf{G}_N}(\sigma) \tag{2.5}$$

between the assignment vector σ^* and the graph $\mathbf{G}_N$, it will be useful to get a better grasp on the conditional law $P_{\sigma^*|\mathbf{G}_N}$ of the assignment vector σ^* given the graph $\mathbf{G}_N$. Observing that

$$\mathbb{P}\big\{\mathbf{G}_N = (G_{ij}) \mid \sigma^* = \sigma\big\} = \prod_{i<j}(c_N + \Delta_N\sigma_i^*\sigma_j^*)^{G_{ij}}(1-c_N-\Delta_N\sigma_i^*\sigma_j^*)^{1-G_{ij}}, \tag{2.6}$$

Bayes' formula can be used to obtain the law of the assignment vector σ^* conditionally on the observation of $\mathbf{G}_N$. It can be written in the form of a Gibbs measure,

$$\mathbb{P}\big\{\sigma^* = \sigma \mid \mathbf{G}_N = (G_{ij})\big\} = \frac{\exp H_N^{\mathrm{SBM}}(\sigma)P_N^*(\sigma)}{\int_{\Sigma_N}\exp H_N^{\mathrm{SBM}}(\tau)\,\mathrm{d}P_N^*(\tau)}, \tag{2.7}$$

for the Hamiltonian

$$H_N^{\mathrm{SBM}}(\sigma) := \sum_{i<j}\log\left[\left(1+\frac{\Delta_N}{c_N}\sigma_i\sigma_j\right)^{G_{ij}}\left(1-\frac{\Delta_N}{1-c_N}\sigma_i\sigma_j\right)^{1-G_{ij}}\right]. \tag{2.8}$$

Denoting its associated average free energy by

$$\overline{F}_N^{\mathrm{SBM}} := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp H_N^{\mathrm{SBM}}(\sigma)\,\mathrm{d}P_N^*(\sigma), \tag{2.9}$$

in the limit of large N, this average free energy coincides with the mutual information (2.5) up to an explicit additive constant.

Proposition 2.1. *Under assumptions (***A1***) and (***A2***), the limits of the average free energy (2.9) and of the mutual information (2.5) differ by an additive constant,*

$$\frac{1}{N}I(\mathbf{G}_N;\sigma^*) = \frac{\lambda}{4} - \overline{F}_N^{\mathrm{SBM}} + o(1). \tag{2.10}$$

Proof. The explicit form of the likelihood in (2.6) and the definition of the Hamiltonian in (2.8) imply that

$$\frac{1}{N}I(\mathbf{G}_N;\sigma^*) = \frac{1}{N}\mathbb{E}H_N^{\mathrm{SBM}}(\sigma^*) - \overline{F}_N^{\mathrm{SBM}}. \tag{2.11}$$

Since the coordinates of the assignment vector σ^* are i.i.d., the first term simplifies to

$$\frac{1}{N}\mathbb{E}H_N^{\mathrm{SBM}}(\sigma^*) = \frac{1}{N}\cdot\binom{N}{2}\cdot\left[\mathbb{E}G_{12}\log\left(1+\frac{\Delta_N}{c_N}\sigma_1^*\sigma_2^*\right)+\mathbb{E}(1-G_{12})\log\left(1-\frac{\Delta_N}{1-c_N}\sigma_1^*\sigma_2^*\right)\right].$$

Averaging with respect to the randomness of G_{12} conditionally on the randomness of the assignment vector σ^* reveals that this is equal to

$$\frac{N-1}{2}\left[c_N\mathbb{E}\left(1+\frac{\Delta_N}{c_N}\sigma_1^*\sigma_2^*\right)\log\left(1+\frac{\Delta_N}{c_N}\sigma_1^*\sigma_2^*\right)+(1-c_N)\mathbb{E}\left(1-\frac{\Delta_N}{1-c_N}\sigma_1^*\sigma_2^*\right)\log\left(1-\frac{\Delta_N}{1-c_N}\sigma_1^*\sigma_2^*\right)\right].$$

Taylor expanding the logarithm shows that

$$\frac{1}{N}\mathbb{E}H_N^{\mathrm{SBM}}(\sigma^*) = \frac{N-1}{4}\cdot\left[\frac{\Delta_N^2}{c_N(1-c_N)}+\mathcal{O}\left(\frac{\Delta_N^3}{c_N^2}+\frac{\Delta_N^3}{(1-c_N)^2}\right)\right].$$

Recalling the definition of the constant λ_N in (1.9), observing that $c_N\in(0,1)$, and remembering (2.11) gives

$$\frac{1}{N}I(\mathbf{G}_N;\sigma^*) = \frac{N-1}{N}\cdot\frac{\lambda_N}{4} - \overline{F}_N^{\mathrm{SBM}} + \mathcal{O}\left(\frac{N-1}{N}\cdot\frac{\lambda_N^{3/2}}{\sqrt{Nc_N(1-c_N)}}\right).$$

Invoking assumptions (**A1**) and (**A2**) completes the proof. ■

This result reduces the task of understanding the limit of the mutual information (2.5), an information-theoretic quantity, to computing the limit of the free energy (2.9), a statistical mechanics quantity. The limit of this free energy will be determined indirectly by leveraging a universality property that allows one to map the free energy in the dense stochastic block model to that in the symmetric rank-one matrix estimation problem. The limit of the free energy in the symmetric rank-one matrix estimation problem can be computed using the Hamilton-Jacobi approach.

2.2 A symmetric rank-one matrix estimation problem

The problem of symmetric rank-one matrix estimation consists of recovering the symmetric rank-one matrix generated by a vector

$$\overline{x} := (\overline{x}_1,\dots,\overline{x}_N)\in\mathbb{R}^N \tag{2.12}$$

of independent entries sampled from a bounded probability measure P_1 on the real line, given its noisy observation

$$Y := \sqrt{\frac{2t}{N}}\bar{x}\bar{x}^\top + W \in \mathbb{R}^{N\times N}. \tag{2.13}$$

The noise matrix $W := (W_{ij})_{1\leqslant i,j\leqslant N} \in \mathbb{R}^{N\times N}$ is made of independent standard Gaussian random variables independent of the vector $\bar{x}$, and the parameter $t \geq 0$ is called the *signal-to-noise ratio*. The vector $\bar{x}$ follows a product distribution,

$$P_N := (P_1)^{\otimes N}. \tag{2.14}$$

As in the dense stochastic block model, it will be useful to determine the conditional law of the signal $\bar{x}$ given the observation Y. For every $x \in \mathbb{R}^N$ and $y \in \mathbb{R}^{N\times N}$, a formal computation reveals that

$$\mathbb{P}\{\bar{x} = x \mid Y = y\} = \frac{\mathbb{P}\{\bar{x} = x \text{ and } Y = y\}}{\mathbb{P}\{Y = y\}} = \frac{\exp\Big(-\frac{1}{2}\Big|y - \sqrt{\frac{2t}{N}}xx^\top\Big|^2\Big)\,\mathrm{d}P_N(x)}{\int_{\mathbb{R}^N}\exp\Big(-\frac{1}{2}\Big|y - \sqrt{\frac{2t}{N}}x'x'^\top\Big|^2\Big)\,\mathrm{d}P_N(x')}, \tag{2.15}$$

where, for any two matrices $a, b \in \mathbb{R}^{d\times d}$,

$$a\cdot b := \mathrm{tr}(ab^\top) = \sum_{i=1}^d a_i b_i \quad \text{and} \quad |a| := (a\cdot a)^{\frac{1}{2}} = \left(\sum_{i=1}^d |a_i|^2\right)^{\frac{1}{2}} \tag{2.16}$$

denote the entry-wise scalar product and the Euclidean norm, respectively. More precisely, introducing the Hamiltonian

$$H_N^\circ(t,x) := \sqrt{\frac{2t}{N}}x\cdot Yx - \frac{t}{N}|x|^4 = \sqrt{\frac{2t}{N}}x\cdot Wx + \frac{2t}{N}(x\cdot\bar{x})^2 - \frac{t}{N}|x|^4, \tag{2.17}$$

for every bounded measurable function $f:\mathbb{R}^N \to \mathbb{R}$,

$$\mathbb{E}\big[f(\bar{x}) \mid Y\big] = \frac{\int_{\mathbb{R}^N} f(x)\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x)}{\int_{\mathbb{R}^N}\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x)}. \tag{2.18}$$

This is verified rigorously in Exercise 4.2 of [50]. In other words, the conditional law of $\bar{x}$ given Y is the Gibbs measure associated with the Hamiltonian $H_N^\circ(t,\cdot)$. The associated free energy is denoted by

$$F_N^\circ(t) := \frac{1}{N}\log\int_{\mathbb{R}^N}\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x). \tag{2.19}$$

Notice that the Hamiltonian $H_N^\circ(t,\sigma)$ is a random quantity, as it depends on $\bar{x}$ and W through Y. In particular, the free energy (2.19) is also random, and its average is denoted by

$$\bar{F}_N^\circ(t) := \mathbb{E}F_N^\circ(t) = \frac{1}{N}\mathbb{E}\log\int_{\mathbb{R}^N}\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x). \tag{2.20}$$

Slightly abusing terminology, the average free energy (2.20) will often be referred to as simply the free energy. It will be convenient to introduce notation for a random variable whose law is the random Gibbs measure (2.18). For every bounded measurable function $f:\mathbb{R}^N\to\mathbb{R}$, write

$$\langle f(x)\rangle := \int_{\mathbb{R}^N} f(x)\,\mathrm{d}P_{\bar{x}|Y}(x) = \frac{\int_{\mathbb{R}^N} f(x)\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x)}{\int_{\mathbb{R}^N}\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x)}. \tag{2.21}$$

Although this is kept implicit in the notation, the bracket $\langle\cdot\rangle$ is a random quantity that depends on t and on Y. Whenever expressions such as $\langle g(x,\bar{x})\rangle$ are written, it is understood that the variable x is integrated against the conditional probability measure $P_{\bar{x}|Y}$, while keeping $\bar{x}$ fixed. In more explicit notation,

$$\langle g(x,\bar{x})\rangle = \int_{\mathbb{R}^N} g(x,\bar{x})\,\mathrm{d}P_{\bar{x}|Y}(x) = \frac{\int_{\mathbb{R}^N} g(x,\bar{x})\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x)}{\int_{\mathbb{R}^N}\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x)}, \tag{2.22}$$

not to be confused with $\mathbb{E}[g(\bar{x},\bar{x})\mid Y]$ for instance. If no information about the signal $\bar{x}$ is given by the observation of Y, when $t=0$, then x is simply an independent copy of $\bar{x}$. On the other hand, if perfect information about $\bar{x}$ is possessed, for instance, if $\bar{x}$ is observed with no noise, then x is equal to $\bar{x}$. It will also be convenient to introduce independent copies of x under the Gibbs average $\langle\cdot\rangle$, often called *replicas*, which are denoted by x', x'', or also x^1, x^2, x^3, and so on if an arbitrary number of replicas needs to be considered. Explicitly, for every bounded measurable function $f:\mathbb{R}^N\times\mathbb{R}^N\to\mathbb{R}$,

$$\langle f(x,x')\rangle = \frac{\int_{\mathbb{R}^N}\int_{\mathbb{R}^N} f(x,x')\exp\big(H_N^\circ(t,x)+H_N^\circ(t,x')\big)\,\mathrm{d}P_N(x)\,\mathrm{d}P_N(x')}{\big(\int_{\mathbb{R}^N}\exp H_N^\circ(t,x)\,\mathrm{d}P_N(x)\big)^2}, \tag{2.23}$$

with the natural generalization of this expression in the case of more replicas. The fact that the Gibbs measure (2.18) is a conditional expectation will fundamentally simplify the analysis — in the language of statistical mechanics, the symmetric rank-one matrix estimation problem is always *replica symmetric*. The replica symmetry derives from the Nishimori identity which allows the replacement of one replica x by the ground-truth signal $\bar{x}$, provided that all sources of randomness are averaged.

Proposition 2.2 (Nishimori identity)**.** *For all bounded measurable functions $f:\mathbb{R}^N\times\mathbb{R}^{N\times N}\to\mathbb{R}$ and $g:\mathbb{R}^N\times\mathbb{R}^N\times\mathbb{R}^{N\times N}\to\mathbb{R}$,*

$$\mathbb{E}\big\langle f\big(x,Y\big)\big\rangle = \mathbb{E}\big\langle f\big(\bar{x},Y\big)\big\rangle \quad \textit{and} \quad \mathbb{E}\big\langle g\big(x,x',Y\big)\big\rangle = \mathbb{E}\big\langle g\big(x,\bar{x},Y\big)\big\rangle, \tag{2.24}$$

and so on with more replicas, that is, for every integer $\ell\geq 1$ and bounded measurable function $h:(\mathbb{R}^N)^\ell\times\mathbb{R}^{N\times N}\to\mathbb{R}$,

$$\mathbb{E}\big\langle h\big(x^1,x^2,\ldots,x^\ell,Y\big)\big\rangle = \mathbb{E}\big\langle h\big(x^1,x^2,\ldots,x^{\ell-1},\bar{x},Y\big)\big\rangle. \tag{2.25}$$

Proof. It suffices to prove (2.25). By Dynkin's π-λ theorem (Theorem A.5 in [50]), it suffices to verify this identity for functions that factorize over the variables. Precisely, assume that the function h can be written in the form

$$h(x^1,\ldots,x^\ell,Y) = h_1(x^1,\ldots,x^{\ell-1})h_2(x^\ell)h_3(Y), \tag{2.26}$$

for some bounded measurable functions h_1,h_2,h_3. It then follows that

$$\begin{aligned}\mathbb{E}\langle h(x^1,\ldots,x^\ell,Y)\rangle &= \mathbb{E}\big(\langle h_1(x^1,\ldots,x^{\ell-1},Y)\rangle\langle h_2(x^\ell)\rangle h_3(Y)\big)\\ &= \mathbb{E}\big(\langle h_1(x^1,\ldots,x^{\ell-1},Y)\rangle\mathbb{E}[h_2(\bar{x})\mid Y]h_3(Y)\big),\end{aligned}$$

where the last identity uses that the Gibbs measure (2.21) is the conditional law of $\bar{x}$ given the observation Y. Recalling that the measure $\langle\cdot\rangle$ depends on the randomness only through Y, and thus that $\langle h_1(x^1,\ldots,x^{\ell-1},Y)\rangle$

is Y-measurable, reveals that

$$\begin{aligned}\mathbb{E}\langle h(x^1,\ldots,x^\ell,Y)\rangle &= \mathbb{E}\big(\mathbb{E}[\langle h_1(x^1,\ldots,x^{\ell-1},Y)\rangle h_2(\overline{x})h_3(Y)\,|\,Y]\big)\\ &= \mathbb{E}\big(\langle h_1(x^1,\ldots,x^{\ell-1},Y)\rangle h_2(\overline{x})h_3(Y)\big)\\ &= \mathbb{E}\big(\langle h_1(x^1,\ldots,x^{\ell-1},Y)\rangle h_2(\overline{x})h_3(Y)\big).\end{aligned}$$

Combining this with the definition (2.26) of the function h completes the proof. ■

A universality property of the free energy (2.9) in the dense stochastic block model ensures that it is asymptotically equivalent to the free energy (2.20) in the symmetric rank-one matrix estimation problem for the appropriate choice of prior and signal-to-noise ratio.

Proposition 2.3. *Under assumptions (**A1**) and (**A2**), the free energy (2.9) in the stochastic block model is asymptotically equivalent to the free energy (2.20) in the symmetric rank-one matrix estimation problem with Bernoulli prior,*

$$\overline{F}_N^{\mathrm{SBM}} = \overline{F}_N^{\circ}\left(\frac{\lambda}{4}\right) + o(1). \tag{2.27}$$

Before delving into the proof of this result, it will be convenient to simplify the expression of the Hamiltonian H_N^{SBM} in (2.8), and discard some lower-order terms to gain intuition. Taylor expanding the logarithm in the definition of this Hamiltonian reveals that

$$\begin{aligned}H_N^{\mathrm{SBM}}(\sigma) = \widetilde{H}_N^{\mathrm{SBM}}(\sigma) &+ \sum_{i<j}\left[G_{ij}\left(\frac{\Delta_N^2}{2(1-c_N)^2} - \frac{\Delta_N^2}{2c_N^2}\right) - \frac{\Delta_N^2}{2(1-c_N)^2}\right]\\ &+ \sum_{i<j}\mathcal{O}\left(G_{ij}\frac{\Delta_N^3}{c_N^3} + (1-G_{ij})\frac{\Delta_N^3}{(1-c_N)^3}\right)\end{aligned} \tag{2.28}$$

for the Hamiltonian

$$\widetilde{H}_N^{\mathrm{SBM}}(\sigma) := \sum_{i<j}\frac{\Delta_N}{c_N(1-c_N)}(G_{ij}-c_N)\sigma_i\sigma_j. \tag{2.29}$$

Introducing the free energy

$$\widetilde{F}_N^{\mathrm{SBM}} := \frac{1}{N}\mathbb{E}\int_{\Sigma_N}\exp\widetilde{H}_N^{\mathrm{SBM}}(\sigma)\,\mathrm{d}P_N^*(\sigma) \tag{2.30}$$

associated with the Hamiltonian (2.29), the equality (2.28) implies that

$$\begin{aligned}\overline{F}_N^{\mathrm{SBM}} = \widetilde{F}_N^{\mathrm{SBM}} &+ \frac{1}{N}\sum_{i<j}\mathbb{E}\left[G_{ij}\left(\frac{\Delta_N^2}{2(1-c_N)^2} - \frac{\Delta_N^2}{2c_N^2}\right) - \frac{\Delta_N^2}{2(1-c_N)^2}\right]\\ &+ \sum_{i<j}\mathcal{O}\left(\mathbb{E}G_{ij}\frac{\Delta_N^3}{c_N^3} + (1-\mathbb{E}G_{ij})\frac{\Delta_N^3}{(1-c_N)^3}\right).\end{aligned} \tag{2.31}$$

Noticing that $\mathbb{E}G_{ij} = c_N + \Delta_N\overline{m}^2$, and remembering the definition of λ_N in (1.9), this becomes

$$\overline{F}_N^{\mathrm{SBM}} = \widetilde{F}_N^{\mathrm{SBM}} - \frac{N-1}{N}\cdot\frac{\lambda_N}{4} + \mathcal{O}\left(\frac{N-1}{N}\cdot\frac{\lambda_N^{3/2}}{\sqrt{Nc_N(1-c_N)}}\right). \tag{2.32}$$

Leveraging the assumptions (**A1**) and (**A2**) shows that

$$\overline{F}_N^{\mathrm{SBM}} = \widetilde{F}_N^{\mathrm{SBM}} - \frac{\lambda}{4} + o(1) \tag{2.33}$$

so the free energies (2.9) and (2.30) are equal up to an additive constant. The free energy (2.30) starts to look more like the free energy (2.20) in the symmetric rank-one matrix estimation problem. Indeed, introducing the centred random variables

$$\widetilde{G}_{ij} := \frac{\Delta_N}{c_N(1-c_N)}\big(G_{ij} - c_N - \Delta_N \sigma_i^* \sigma_j^*\big), \tag{2.34}$$

the Hamiltonian (2.29) may be written as

$$\widetilde{H}_N^{\mathrm{SBM}}(\sigma) = \sum_{i<j}\bigg(\widetilde{G}_{ij}\sigma_i\sigma_j + \frac{\lambda_N}{N}\sigma_i^*\sigma_j^*\sigma_i\sigma_j\bigg). \tag{2.35}$$

A direct computation shows that the random variables $\widetilde{G}_{ij}$ have variance

$$\begin{aligned}\mathbb{E}\widetilde{G}_{ij}^2 &= \frac{\lambda_N}{Nc_N(1-c_N)}\mathbb{E}\big(1-c_N-\Delta_N\sigma_i^*\sigma_j^*\big)\big(c_N+\Delta_N\sigma_i^*\sigma_j^*\big)\\ &= \frac{\lambda_N}{N} + \mathcal{O}\bigg(\frac{\lambda_N}{N}\cdot\bigg(\sqrt{\frac{\lambda_N}{Nc_N(1-c_N)}} + \frac{\lambda_N}{N}\bigg)\bigg).\end{aligned} \tag{2.36}$$

Recall that under assumption (**A2**), the average degree of a node diverges as N tends to infinity. This gives some credence to the idea that in the limit of large system size, a sort of central limit theorem takes place, and the random variables $(\widetilde{G}_{ij})_{i,j\leqslant N}$ may as well be substituted by centred Gaussian random variables with the same variance. In other words, the free energy (2.30) may be expected to be asymptotically equivalent to the Gaussian free energy

$$\overline{F}_N^{\mathrm{gauss}}(\lambda) := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp H_N^{\mathrm{gauss}}(\lambda,\sigma)\,\mathrm{d}P_N^*(\sigma) \tag{2.37}$$

associated with the Hamiltonian

$$H_N^{\mathrm{gauss}}(\lambda,\sigma) := \sum_{i<j}\bigg(\sqrt{\frac{\lambda}{N}}W_{ij}\sigma_j\sigma_j + \frac{\lambda}{N}\sigma_i^*\sigma_j^*\sigma_i\sigma_j\bigg). \tag{2.38}$$

Noticing that $\frac{W_{ij}+W_{ji}}{\sqrt{2}}$ is again a standard Gaussian, the equality in distribution

$$H_N^{\mathrm{gauss}}(\lambda,\sigma) \stackrel{d}{=} H_N^{\circ}\bigg(\frac{\lambda}{4}\bigg) + \frac{N\lambda}{4} - \sum_{i=1}^N \frac{\lambda}{2N}W_{ii} - \frac{\lambda}{2} \tag{2.39}$$

holds jointly over σ. For the choice of Bernoulli prior $P_1 = P^*$, this implies that

$$\overline{F}_N^{\mathrm{gauss}}(\lambda) = \overline{F}_N^{\circ}\bigg(\frac{\lambda}{4}\bigg) + \frac{\lambda}{4} - \frac{\lambda}{2N}, \tag{2.40}$$

where it has been used that $\mathbb{E}W_{ii} = 0$. To establish Proposition 2.3 it remains to prove rigorously that the free energies (2.30) and (2.37) are asymptotically equivalent up to an additive constant. This will be done through an interpolation argument taken from Theorem 3.9 in [95].

Proof of Proposition 2.3. By the asymptotic equivalences (2.33) and (2.40), it suffices to show that

$$\widetilde{F}_N^{\text{SBM}} = \overline{F}_N^{\text{gauss}}(\lambda) + o(1). \tag{2.41}$$

To alleviate notation, the dependence on λ will be kept implicit. The proof proceeds by interpolation. For each $t \in [0,1]$, define the interpolating Hamiltonian

$$H_{N,t}(\sigma) := \sum_{i<j} \left[\left(\sqrt{t}\widetilde{G}_{ij} + \sqrt{1-t}\sqrt{\frac{\lambda}{N}} W_{ij} \right) \sigma_i \sigma_j + \left(t \cdot \frac{\lambda_N}{N} + (1-t) \cdot \frac{\lambda}{N} \right) \sigma_i^* \sigma_j^* \sigma_i \sigma_j \right],$$

and the interpolating free energy

$$\overline{F}_N(t) := \frac{1}{N} \mathbb{E} \log \int_{\Sigma_N} \exp H_{N,t}(\sigma) \,\mathrm{d}P_N(\sigma).$$

By the fundamental theorem of calculus,

$$\left| \widetilde{F}_N^{\text{SBM}} - \overline{F}_N^{\text{gauss}} \right| = \left| \overline{F}_N(1) - \overline{F}_N(0) \right| \leqslant \sup_{t \in [0,1]} \left| \overline{F}_N'(t) \right|. \tag{2.42}$$

To compute this derivative, write $\langle \cdot \rangle_t$ for the average with respect to the Gibbs measure associated with the interpolating Hamiltonian $H_{N,t}$, and observe that

$$\begin{aligned} \overline{F}_N'(t) = \frac{1}{2N\sqrt{t}} \sum_{i<j} \mathbb{E} \widetilde{G}_{ij} \langle \sigma_i \sigma_j \rangle_t - \frac{\sqrt{\lambda}}{2N^{3/2}\sqrt{1-t}} \sum_{i<j} \mathbb{E} W_{ij} \langle \sigma_i \sigma_j \rangle_t \\ + \left(\lambda_N - \lambda \right) \cdot \frac{1}{N^2} \sum_{i<j} \mathbb{E} \left\langle \sigma_i^* \sigma_j^* \sigma_i \sigma_j \right\rangle_t . \end{aligned} \tag{2.43}$$

At this point, fix indices $i<j$ and introduce the function $F(\widetilde{G}_{ij}) = \langle \sigma_i \sigma_j \rangle_t$. A direct computation reveals that

$$\begin{aligned} \partial_{\widetilde{G}_{ij}} F &= \langle \sigma_i \sigma_j \partial_{\widetilde{G}_{ij}} H_{N,t}(\sigma) \rangle_t - \langle \sigma_i \sigma_j \rangle_t \langle \partial_{\widetilde{G}_{ij}} H_{N,t}(\sigma) \rangle_t = \sqrt{t}\left(1 - \langle \sigma_i \sigma_j \rangle_t^2 \right) \\ \partial_{\widetilde{G}_{ij}}^2 F &= -2\sqrt{t} \langle \sigma_i \sigma_j \rangle_t \cdot \partial_{\widetilde{G}_{ij}} F = -2t\left(\langle \sigma_i \sigma_j \rangle_t - \langle \sigma_i \sigma_j \rangle_t^3 \right) \end{aligned}$$

so the approximate Gaussian integration by parts formula (Exercise 4.6 in [50]) gives

$$\mathbb{E} \widetilde{G}_{ij} \langle \sigma_i \sigma_j \rangle_t = \sqrt{t} \mathbb{E} \widetilde{G}_{ij}^2 \left(1 - \mathbb{E} \langle \sigma_i \sigma_j \rangle_t^2 \right) + \mathcal{O}\left(\mathbb{E} |\widetilde{G}_{ij}|^3 \right).$$

On the other hand, the Gaussian integration by parts formula (Theorem 4.5 in [50]) shows that

$$\mathbb{E} W_{ij} \langle \sigma_i \sigma_j \rangle_t = \sqrt{1-t} \cdot \sqrt{\frac{\lambda}{N}} \cdot \left(1 - \mathbb{E} \langle \sigma_i \sigma_j \rangle_t^2 \right).$$

It follows by (2.43) that

$$\overline{F}_N'(t) = \frac{1}{2N} \sum_{i<j} \mathbb{E} \widetilde{G}_{ij}^2 \left(1 - \mathbb{E} \langle \sigma_i \sigma_j \rangle_t^2 \right) - \frac{\lambda}{2N^2} \sum_{i<j} \left(1 - \mathbb{E} \langle \sigma_i \sigma_j \rangle_t^2 \right) + \mathcal{O}\left(N \mathbb{E} |\widetilde{G}_{12}|^3 + |\lambda_N - \lambda| \right).$$

Together with (2.36) and (2.42), this implies that

$$\widetilde{F}_N^{\text{SBM}} = \overline{F}_N^{\text{gauss}} + \mathcal{O}\left(\lambda_N \cdot \left(\sqrt{\frac{\lambda_N}{Nc_N(1-c_N)}} + \frac{\lambda_N}{N}\right) + N\mathbb{E}|\widetilde{G}_{12}|^3 + |\lambda_N - \lambda|\right).$$

Observing that

$$\mathbb{E}|\widetilde{G}_{12}|^3 \leqslant \left|\frac{\Delta_N^3}{c_N^3(1-c_N)^3}\right| \mathbb{E}\left|(1-c_N-\Delta_N\overline{\sigma}_i\overline{\sigma}_j)(c_N+\Delta_N\overline{\sigma}_i\overline{\sigma}_j)\right| \leqslant \frac{4\lambda_N^{3/2}}{N\sqrt{Nc_N(1-c_N)}},$$

and remembering the assumptions (**A1**) and (**A2**) establishes (2.41) and completes the proof. ■

This result reduces the task of understanding the limit of the free energy (2.9), and hence of the mutual information (2.5), to computing the limit of the free energy (2.20) in the rank-one matrix estimation problem. To do so, the Hamilton-Jacobi approach will be used. The structure of the argument closely resembles that in [36] — the only difference will be in the well-posedness theory for the Hamilton-Jacobi equation.

2.3 A matrix estimation Hamilton-Jacobi equation

To determine the limit free energy in the symmetric rank-one matrix estimation problem using the Hamilton-Jacobi approach, a partial differential equation satisfied by the free energy (2.20) up to an error that vanishes with N needs to be derived. Notice that the limits of the random free energy (2.19) and its average (2.20), provided that they exist, are the same.

Lemma 2.4. *The free energy (2.19) concentrates about its average (2.20); that is, for every $T < +\infty$, there is a constant $C < +\infty$ such that for every $t \in [0,T]$ and $\lambda > 0$,*

$$\mathbb{P}\left\{|F_N^\circ(t) - \overline{F}_N^\circ(t)| \geqslant \lambda\right\} \leqslant 2\exp\left(-\frac{N\lambda^2}{C}\right). \tag{2.44}$$

In particular, for every $t \geqslant 0$, the free energy (2.19) and its average (2.20) are asymptotically equivalent,

$$\limsup_{N\to+\infty}\left|F_N^\circ(t) - \overline{F}_N^\circ(t)\right| = 0. \tag{2.45}$$

Proof. The free energy (2.19) in the symmetric rank-one matrix estimation problem is a function of the Gaussian noise W and the signal $\overline{x}$. To make this dependence explicit, temporarily change the notation and write the free energy as

$$F_N^\circ(t, W, \overline{x}) := F_N^\circ(t).$$

Letting $a > 0$ be such that the support of the measure P_1 is contained in the interval $[-\sqrt{a}, \sqrt{a}]$, the measure P_N is supported in the closed Euclidean ball $\overline{B}_{\sqrt{aN}}(0)$ of radius $\sqrt{aN}$ centred at the origin. This means that for $t \geqslant 0$, $\overline{x} \in \mathbb{R}^N$ and $W^1, W^2 \in \mathbb{R}^{N\times N}$,

$$F_N^\circ(t, W^1, \overline{x}) \leqslant \sqrt{\frac{2t}{N^3}} \sup_{x\in\overline{B}_{\sqrt{aN}}(0)} \left(x\cdot W^1 x - x\cdot W^2 x\right) + F_N^\circ(t, W^2, \overline{x}). \tag{2.46}$$

Using that $x\cdot(W^1-W^2)x=(W^1-W^2)\cdot(x^*x)$ and invoking the Cauchy-Schwarz inequality reveals that

$$F_N^\circ(t,W^1,\overline{x})-F_N^\circ(t,W^2,\overline{x})\leqslant\sqrt{\frac{2t}{N^3}}\sup_{x\in\overline{B}_{\sqrt{aN}}(0)}|x|^2|W^1-W^2|\leqslant a\sqrt{\frac{2t}{N}}|W^1-W^2|. \tag{2.47}$$

Since the right side of (2.47) is symmetric in the pair (W^1,W^2), this estimate gives an upper bound on the Lipschitz semi-norm of the map $W\mapsto F_N^\circ(t,W,\overline{x})$. It thus follows by the Gaussian concentration inequality (Theorem 4.7 in [50]) applied conditionally on the randomness of the signal $\overline{x}$ that, for any $\lambda>0$,

$$\mathbb{P}\big\{|F_N^\circ(t,W,\overline{x})-\mathbb{E}_W F_N^\circ(t,W,\overline{x})|\geqslant\lambda\big\}\leqslant 2\exp\bigg(-\frac{N\lambda^2}{4a^2t}\bigg), \tag{2.48}$$

where $\mathbb{E}_W$ denotes the average only with respect to the randomness of the Gaussian noise W. To obtain the concentration of the free energy, it remains to establish the concentration of $\mathbb{E}_W F_N^\circ(t,W,\overline{x})$ about its average with respect to the signal $\overline{x}$. Fix $t\geqslant 0, W\in\mathbb{R}^{N\times N}$ and $\overline{x}\in\mathbb{R}^N$. A direct computation reveals that for any $1\leqslant i\leqslant N$,

$$\partial_{\overline{x}_i}F_N^\circ(t,W,\overline{x})=\frac{4t}{N^2}\langle(x\cdot\overline{x}).x_i\rangle. \tag{2.49}$$

Letting $a>0$ be such that the support of the measure P_1 is contained in the interval $[-\sqrt{a},\sqrt{a}]$, this implies that

$$\big|\partial_{\overline{x}_i}F_N^\circ(t,W,\overline{x})\big|\leqslant\frac{4ta^{3/2}}{N}. \tag{2.50}$$

Averaging over the randomness of the Gaussian noise W and using Jensen's inequality shows that this upper bound also holds for the averaged free energy $\mathbb{E}_W F_N^\circ(t,W,\overline{x})$. Combining this with the mean value theorem shows that for any $i\in\{1,\ldots,N\}$ and $\overline{x}_1,\ldots,\overline{x}_N,\overline{x}_i'\in[-\sqrt{a},\sqrt{a}]$,

$$\begin{aligned}\Big|\mathbb{E}_W F_N^\circ(t,W,\overline{x}_1,\ldots,\overline{x}_{i-1},\overline{x}_i',\overline{x}_{i+1},\ldots,\overline{x}_n)&\\ -\mathbb{E}_W F_N^\circ(\overline{x}_1,\ldots,\overline{x}_{i-1},\overline{x}_i,\overline{x}_{i+1},\overline{x}_n)\Big|&\leqslant\frac{4ta^{3/2}}{N}|\overline{x}_i-\overline{x}_i'|=\frac{8ta^2}{N}.\end{aligned} \tag{2.51}$$

It follows by the McDiarmid inequality (Theorem 4.8 in [50]) that

$$\mathbb{P}\big\{|\mathbb{E}_W F_N^\circ(t,W,\overline{x})-\mathbb{E}F_N^\circ(t,W,\overline{x})|\geqslant\lambda\big\}\leqslant 2\exp\bigg(-\frac{N\lambda^2}{32t^2a^4}\bigg). \tag{2.52}$$

Together with the triangle inequality and the upper bound (2.48), this establishes the exponential concentration (2.44) of the free energy about its average. Since the right side of this expression is summable, invoking the Borel-Cantelli lemma completes the proof. ∎

The Gaussian integration by parts formula (Theorem 4.5 in [50]) reveals that

$$\mathbb{E}\langle x\cdot Wx\rangle=\sqrt{\frac{2t}{N}}\big(\mathbb{E}\big\langle|x|^4\big\rangle-\mathbb{E}\big\langle(x\cdot x')^2\big\rangle\big). \tag{2.53}$$

Together with a direct derivative computation, this implies that

$$\partial_t\overline{F}_N^\circ(t)=\frac{1}{N^2}\mathbb{E}\big\langle(x\cdot\overline{x})^2\big\rangle. \tag{2.54}$$

Unfortunately, there is no way of closing this equation if only derivatives in t of the free energy can be computed. Indeed, the situation is analogous to that encountered if one were to study a Curie-Weiss model without any magnetization part, where there would be no parameter h with respect to which to differentiate and close the equation — see Chapter 3 of [50] for a detailed discussion of the Curie-Weiss model and its analysis via the Hamilton-Jacobi approach. To overcome this issue, an "enriched" free energy that also depends on an additional parameter h will be defined. The enriched free energy $\overline{F}_N(t,h)$ should extend the free energy $\overline{F}_N^\circ(t)$, in the sense that there is some h_0 with $\overline{F}_N(\cdot,h_0) = \overline{F}_N^\circ(\cdot)$; it should be simple enough that the limit of its initial condition $\overline{F}_N(0,\cdot)$ can be computed explicitly; and it should be rich enough that it allows to close the equation, up to a small error term. In the context of statistical inference models, one also needs to ensure that the enrichment does not destroy the fact that the Gibbs measure is a conditional expectation. Indeed, this property gives access to the Nishimori identity (Proposition 2.2), which plays a fundamental role in simplifying statistical inference models and distinguishing them from the more complicated spin-glass models.

In the context of the symmetric rank-one matrix estimation problem, the appropriate enrichment of the free energy is obtained by assuming that, in addition to observing the noisy rank-one matrix Y in (2.13), a noisy version $\widetilde{Y}$ of the signal vector $\overline{x}$ is also observed,

$$\widetilde{Y} := \sqrt{2h}\,\overline{x} + z. \tag{2.55}$$

The noise vector $z := (z_i)_{i\leqslant N} \in \mathbb{R}^N$ is made of independent standard Gaussian random variables independent of the vector $\overline{x}$ and the noise matrix W, and the parameter $h \geq 0$ is a signal-to-noise ratio. The enriched symmetric rank-one matrix estimation problem is to infer the signal $\overline{x}$ from the observation $\mathcal{Y} := (Y,\widetilde{Y})$. Applying Bayes' formula shows that the law of the signal $\overline{x}$ given the observation of $\mathcal{Y}$ is the Gibbs measure whose Hamiltonian on $\mathbb{R}^N$ is

$$H_N(t,h,x) := H_N^\circ(t,x) + \sqrt{2h}\,\widetilde{Y}\cdot x - h|x|^2. \tag{2.56}$$

In other words, for any bounded measurable function $f:\mathbb{R}^N \to \mathbb{R}$,

$$\mathbb{E}\big[f(\overline{x}) \mid \mathcal{Y}\big] = \frac{\int_{\mathbb{R}^N} f(x)\exp H_N(t,h,x)\,\mathrm{d}P_N(x)}{\int_{\mathbb{R}^N} \exp H_N(t,h,x)\,\mathrm{d}P_N(x)}. \tag{2.57}$$

The free energy

$$F_N(t,h) := \frac{1}{N}\log\int_{\mathbb{R}^N} \exp H_N(t,h,x)\,\mathrm{d}P_N(x) \tag{2.58}$$

of this model is again random, as it depends on $\overline{x}$, W, and z, and its average is again denoted by

$$\overline{F}_N(t,h) := \mathbb{E}F_N(t,h) = \frac{1}{N}\mathbb{E}\log\int_{\mathbb{R}^N} \exp H_N(t,h,x)\,\mathrm{d}P_N(x). \tag{2.59}$$

The average free energy (2.59) will often be referred to as simply the free energy. Through a slight abuse of notation, as before, $\langle\cdot\rangle$ will denote the average with respect to the Gibbs measure (2.57), and x, x', x'', and so on, will be independent random variables sampled according to this probability measure. That is, for every bounded measurable function $f:\mathbb{R}^N \to \mathbb{R}$,

$$\langle f(x)\rangle := \frac{\int_{\mathbb{R}^N} f(x)\exp H_N(t,h,x)\,\mathrm{d}P_N(x)}{\int_{\mathbb{R}^N} \exp H_N(t,h,x)\,\mathrm{d}P_N(x)}, \tag{2.60}$$

and so on as in (2.23) with more replicas. The derivation of the identity (2.54) is unchanged for this more general Gibbs measure: for all $t,h \geq 0$,

$$\partial_t \overline{F}_N(t,h) = \frac{1}{N^2}\mathbb{E}\Big\langle (x\cdot\overline{x})^2\Big\rangle. \tag{2.61}$$

Recalling from (2.55) and (2.56) that

$$H_N(t,h,x) = H_N^\circ(t,x) + 2hx\cdot\overline{x} + \sqrt{2h}z\cdot x - h|x|^2, \tag{2.62}$$

it is also possible to compute the spatial derivative of the free energy: for all $t,h \geq 0$,

$$\partial_h \overline{F}_N(t,h) = \frac{1}{N\sqrt{2h}}\mathbb{E}\langle z\cdot x\rangle + \frac{2}{N}\mathbb{E}\langle x\cdot\overline{x}\rangle - \frac{1}{N}\mathbb{E}\Big\langle |x|^2\Big\rangle. \tag{2.63}$$

The Gaussian integration by parts formula (Theorem 4.5 in [50]) reveals that

$$\mathbb{E}\langle z\cdot x\rangle = \sqrt{2h}\Big(\mathbb{E}\Big\langle |x|^2\Big\rangle - \mathbb{E}\langle x\cdot x'\rangle\Big) \tag{2.64}$$

which together with the Nishimori identity implies that

$$\partial_h \overline{F}_N(t,h) = \frac{1}{N}\mathbb{E}\langle x\cdot\overline{x}\rangle = \frac{1}{N}\mathbb{E}|\langle x\rangle|^2. \tag{2.65}$$

It is reasonable to expect the variance of the inner product, or overlap, $N^{-1}x\cdot\overline{x}$ between a sample x from the Gibbs measure (2.57) and the ground-truth signal $\overline{x}$ to be small, simply because it is the average of a large number of variables. If this is so, then the difference between the time derivative (2.61) and the square of the spatial derivative (2.65) would also be small since

$$\partial_t \overline{F}_N(t,h) - \big(\partial_h \overline{F}_N(t,h)\big)^2 = \mathrm{Var}\left(\frac{x\cdot\overline{x}}{N}\right). \tag{2.66}$$

This suggests that the enriched free energy (2.59) should converge to the function f solving the Hamilton-Jacobi equation

$$\partial_t f(t,h) - \big(\partial_h f(t,h)\big)^2 = 0 \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}_{>0} \tag{2.67}$$

subject to the initial condition

$$\psi(h) := \lim_{N\to+\infty} \overline{F}_N(0,h) = \overline{F}_1(0,h). \tag{2.68}$$

The assumption that P_N is a product measure has been used to assert that the initial condition is independent of N. To make this argument rigorous, a well-posedness theory for the Hamilton-Jacobi equation (2.67) needs to be established. The general well-posedness theory for Hamilton-Jacobi equations on positive half-space developed in the next section relies on the assumptions that the non-linearity is locally Lipschitz continuous and non-decreasing. Given a set $\mathcal{D}\subseteq\mathbb{R}^d$, a function $h:\mathcal{D}\to\mathbb{R}$ is said to be *non-decreasing* if, for all $y,y'\in\mathcal{D}$,

$$y \leqslant y' \quad\Longrightarrow\quad h(y)\leqslant h(y'), \tag{2.69}$$

where $\leqslant$ denotes the partial order,

$$y\leqslant y' \iff y'-y\in\mathbb{R}^d_{\geqslant 0}. \tag{2.70}$$

Although the non-linearity $\mathsf{H}(p) := p^2$ in the Hamilton-Jacobi equation (2.73) is locally Lipschitz continuous, it fails to be non-decreasing on $\mathbb{R}$; however its modification

$$\overline{\mathsf{H}}(p) := p^2\mathbf{1}\{p \geqslant 0\} \tag{2.71}$$

is both locally Lipschitz continuous and non-decreasing on the real line. Moreover, the derivative computation (2.65) suggests that for all $t,h \geqslant 0$,

$$\big(\partial_h f(t,h)\big)^2 = \mathsf{H}\big(\partial_h f(t,h)\big) = \overline{\mathsf{H}}\big(\partial_h f(t,h)\big) \tag{2.72}$$

which means that the enriched free energy (2.59) should converge to the function f solving the Hamilton-Jacobi equation

$$\partial_t f(t,h) - \overline{\mathsf{H}}\big(\partial_h f(t,h)\big) = 0 \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}_{>0} \tag{2.73}$$

subject to the initial condition $f(0,\cdot) = \psi(\cdot)$. The well-posedness of this Hamilton-Jacobi equation will be established in the next section. The focus will then be on controlling the error term on the right side of (2.66), and proving that the limit of the free energy indeed solves the Hamilton-Jacobi equation (2.73). Together with the Hopf-Lax variational formula, this will lead to the following result for the limit free energy in the symmetric rank-one matrix estimation problem.

Theorem 2.5. *For every $N \geqslant 1$, denote by $\overline{F}_N : \mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0} \to \mathbb{R}$ the enriched free energy (2.59) in the symmetric rank-one matrix estimation problem. For every $t,h \geqslant 0$, the sequence $(\overline{F}_N(t,h))_{N\geqslant 1}$ converges to $f(t,h)$ as N tends to infinity, where $f : \mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0} \to \mathbb{R}$ is the unique viscosity solution to the Hamilton-Jacobi equation (2.73) subject to the initial condition $\psi : \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ defined in (2.68). Moreover, the limit free energy f admits the Hopf-Lax representation, for every $t,h \geq 0$,*

$$f(t,h) = \sup_{h'\in\mathbb{R}_{\geqslant 0}}\left(\psi(h') - \frac{(h'-h)^2}{4t}\right). \tag{2.74}$$

In particular, the limit of the free energy $\overline{F}^{\circ}_N$ defined in (2.20) is given by

$$\lim_{N\to+\infty}\overline{F}^{\circ}_N(t) = \sup_{h\in\mathbb{R}_{\geqslant 0}}\left(\psi(h) - \frac{h^2}{4t}\right). \tag{2.75}$$

Together with Propositions 2.3 and 2.1 this result implies Theorem 1.3. The remainder of this chapter will therefore be devoted to proving Theorem 2.5.

2.4 Hamilton-Jacobi equations on positive half-space

The simplest setting in which to establish a well-posedness theory for Hamilton-Jacobi equations is on Euclidean space $\mathbb{R}_{>0}\times\mathbb{R}^d$ — see Chapter 3 of [50] for a detailed discussion of this well-posedness theory. Unfortunately, the Hamilton-Jacobi equation (2.73) that appears in the context of the symmetric rank-one matrix estimation problem is defined on the positive half-space $\mathbb{R}_{>0}\times\mathbb{R}_{>0}$ as opposed to the Euclidean space $\mathbb{R}_{>0}\times\mathbb{R}$. In this section, a well-posedness theory for Hamilton-Jacobi equations on positive half-space is developed following [33, 48]. This requires additional technical effort relative to the Euclidean setting; the upshot of the analysis will be that, for all equations of relevance, no boundary condition for the solution needs to be prescribed, and in effect, the boundary can simply be ignored. Intuitively, this is possible because the characteristic lines always go towards the boundary as t increases rather than away from it.

To motivate the appropriate notion of solution to the Hamilton-Jacobi equation (2.73), temporarily assume that the initial condition ψ is convex and non-decreasing, and take Theorem 2.5 for granted. The Fenchel-Moreau theorem on positive half-space (Proposition A.6) then implies that the unique solution $f : \mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0} \to \mathbb{R}$ to the Hamilton-Jacobi equation (2.73) is given by

$$f(t,h) = \sup_{h' \in \mathbb{R}_{\geqslant 0}} \left(\psi(h') - \frac{(h'-h)^2}{4t} \right) = \sup_{m,h' \in \mathbb{R}_{\geqslant 0}} \left(mh' - \psi^*(m) - \frac{(h'-h)^2)}{4t} \right), \tag{2.76}$$

where $\psi^*(m) = \sup_{h' \in \mathbb{R}_{\geqslant 0}} (h'm - \psi(h'))$ denotes the convex dual of ψ defined in (A.7). Since the function $x \mapsto \frac{1}{2}|x|^2$ is its own convex dual (Exercise 2.8 in [50]), the solution f may be written as

$$f(t,h) = \sup_{m \in \mathbb{R}_{\geqslant 0}} \left(mh + tm^2 - \psi^*(m) \right). \tag{2.77}$$

Leveraging the envelope theorem (Theorem 2.22 in [50]), one can show that at every point $(t,h) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}$ of differentiability of f,

$$\partial_t f(t,h) = m_0^2(t,h) \qquad \text{and} \qquad \partial_h f(t,h) = m_0(t,h) \tag{2.78}$$

for a maximizer $m_0(t,h)$ of the right side of (2.77). In particular, the limit free energy f does indeed satisfy the Hamilton-Jacobi equation (2.73) at all its points of differentiability. Together with Rademacher's theorem (Theorem 2.10 in [50]), this implies that f satisfies the equation (2.73) almost everywhere. Notice that f is Lipschitz continuous because each $\overline{F}_N$ is by (2.61) and (2.65). A natural question arises at this point. Could the limit free energy f be identified as the *unique* Lipschitz continuous function with $f(0,\cdot) = \psi(\cdot)$ which satisfies the equation (2.73) almost everywhere? Unfortunately, the answer is *no*. In fact, the following construction shows that this Hamilton-Jacobi equation admits infinitely many almost-everywhere solutions that are Lipschitz continuous.

Example 2.6. For simplicity, infinitely many Lipschitz functions that satisfy the Hamilton-Jacobi equation (2.73) almost everywhere and are constant equal to zero at the initial time will be constructed. Similar but more complicated constructions can be performed for more general initial conditions.

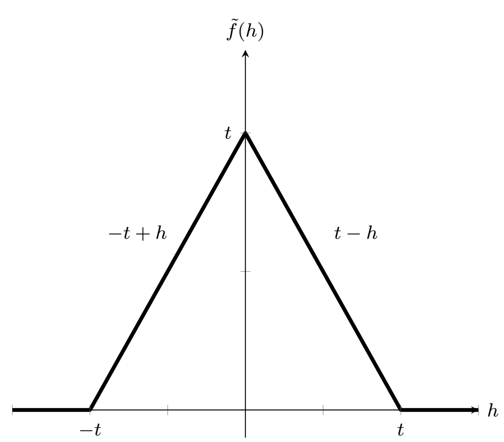

Figure 2.1: Graph of the function $h \mapsto \widetilde{f}(t,h)$, for a fixed value of $t > 0$.

Temporarily disregarding the question of the initial condition, notice that the functions $(t,h) \mapsto 0$, $(t,h) \mapsto t+h$ and $(t,h) \mapsto t-h$ are all solutions to (2.73). It thus follows that the Lipschitz function

$$\widetilde{f}(t,h) := \begin{cases} t+h & \text{if } h \in [-t,0] \\ t-h & \text{if } h \in [0,t] \\ 0 & \text{otherwise} \end{cases}$$

displayed in Figure 2.1 is an almost everywhere solution to (2.73), as it is obtained by "gluing together" these different solutions, and the measure-zero set of points where they are joined together can be disregarded. By construction, the function $\widetilde{f}$ is such that $\widetilde{f}(0,\cdot) = 0$. The null solution also satisfies these properties, so this construction yields two almost-everywhere solutions with the same initial condition. Moreover, any translation in space of $\widetilde{f}$ also satisfies this property; and the "emergence of the corner" can also be delayed to some arbitrary time. So in fact, there are uncountably many Lipschitz functions that solve the equation (2.73) almost everywhere and vanish at the initial time.

One could try to impose the uniqueness of solutions by strengthening the regularity assumptions; for instance, one could impose that a solution f to (2.73) be $C^1(\mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0}; \mathbb{R})$ and solve the equation everywhere. The problem with this idea is that in this case, the set of solutions can be empty. A notion of solution that is more stringent than the "almost-everywhere solutions" explored above, but less stringent than asking the solution to be continuously differentiable on $\mathbb{R}_{\geqslant 0} \times \mathbb{R}_{\geqslant 0}$ therefore needs to be identified. In a nutshell, the function f will be required to satisfy a certain form of the maximum principle. Observe that whenever two functions f and g satisfy the Hamilton-Jacobi equation (2.73) with "viscosity" parameter $\varepsilon > 0$,

$$\partial_t f(t,h) - \overline{\mathsf{H}}\big(\partial_h f(t,h)\big) = \varepsilon \Delta f(t,h) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}_{>0}, \tag{2.79}$$

if their initial conditions are ordered, say $f(0,\cdot) \leqslant g(0,\cdot)$, then this ordering is preserved at all later times $t \geqslant 0$ as well, $f(t,\cdot) \leqslant g(t,\cdot)$. While this monotonicity property will be obtained as a consequence of the definition of solution explained below, one can essentially also go the other way around [8].

To strive for generality, as opposed to focusing exclusively on the Hamilton-Jacobi equation (2.73), for the remainder of this section, fix a Lipschitz initial condition $\psi : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$, and a locally Lipschitz and non-decreasing non-linearity $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$, and consider the Hamilton-Jacobi equation

$$\partial_t f(t,x) - \mathsf{H}\big(\nabla f(t,x)\big) = 0 \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}, \tag{2.80}$$

subject to the initial condition $f(0,\cdot) = \psi(\cdot)$. In this equation, the gradient is taken with respect to the x variable, leaving the t variable aside: $\nabla f(x) = (\partial_{x_1} f, \ldots, \partial_{x_d} f)$. The Hamilton-Jacobi equation (2.73) arising in the context of the symmetric rank-one matrix estimation problem corresponds to the one-dimensional case, $d = 1$, the convex non-linearity $\mathsf{H} = \overline{\mathsf{H}}$ defined in (2.71), and the initial condition ψ defined in (2.68).

As previously alluded to, a natural way to define a solution to the Hamilton-Jacobi equation (2.80) is to add a small "viscosity" parameter $\varepsilon > 0$, and to consider the second-order parabolic equation

$$\partial_t f_\varepsilon(t,x) - \mathsf{H}\big(\nabla f_\varepsilon(t,x)\big) = \varepsilon \Delta f_\varepsilon(t,x) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d_{>0} \tag{2.81}$$

subject to the initial condition $f_\varepsilon(0,\cdot) = \psi(\cdot)$. That smooth solutions exist for (2.81) subject to appropriate boundary conditions can be shown by classical techniques, because the Laplacian term is dominant on very small scales, being of higher order than the other terms of the equation. The solution to the Hamilton-Jacobi equation (2.80) can then be defined as the limit of the solutions to (2.81) as the viscosity parameter ε tends to zero. Although this route will not be pursued rigorously here, suppose for a moment that for each $\varepsilon > 0$, a smooth solution $f_\varepsilon : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ to (2.81) has been defined, and that, as ε tends to zero, the sequence $(f_\varepsilon)_{\varepsilon>0}$ converges to some function $f : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ in the topology of local uniform convergence. In some sense to be discovered, one would like to say that the limit thus obtained is a solution to (2.80). The main difficulty is that a solution to (2.80) should not be imposed to be differentiable everywhere; but uniqueness is not obtained if points of non-differentiability are simply ignored. In analogy with the notion of weak solutions, it would be desirable to introduce smooth test functions and somehow move the derivatives of f onto the test functions. This way, constraints can be established regarding what a solution is allowed to do at points of non-differentiability. This transfer of the derivatives onto the test functions will not be obtained by some integration by parts here. Rather, a strategy that accords well with the fact that the approximations $(f_\varepsilon)_{\varepsilon>0}$ satisfy the maximum principle will be sought. Pick a test function $\phi \in C^\infty(\mathbb{R}_{>0} \times \mathbb{R}^d_{>0}; \mathbb{R})$, and assume that $f - \phi$ achieves a strict local maximum at the point $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$. If f is smooth at (t^*, x^*), then

$$\big(\partial_t \phi(t^*, x^*), \nabla \phi(t^*, x^*)\big) = \big(\partial_t f(t^*, x^*), \nabla f(t^*, x^*)\big), \tag{2.82}$$

so it is expected that

$$\partial_t \phi(t^*, x^*) - \mathsf{H}\big(\nabla \phi(t^*, x^*)\big) = 0. \tag{2.83}$$

This identity may no longer be valid when f is not differentiable at (t^*, x^*), but it turns out that the quantity on the left side of (2.83) must always be non-positive. Showing this requires a simple technical result that will be used consistently throughout this thesis. Write $B_r(x)$ for the Euclidean ball of radius $r > 0$ centred at $x \in \mathbb{R}^d$, and $\overline{B}_r(x)$ for its closure.

Lemma 2.7. *Let $(f_N)_{N\geqslant 1}$ be a sequence of continuous functions on some open set $U \subseteq \mathbb{R}^d$ converging locally uniformly to a function $f : U \to \mathbb{R}$. If f has a strict local maximum at $x \in U$, then there exists $r > 0$ and a sequence of points $(x_N)_{N\geqslant 1} \subseteq U$ converging to x such that, for every N sufficiently large, one has $f_N(x_N) = \sup_{B_r(x)} f_N$.*

Proof. Fix $r > 0$ sufficiently small that x is a strict maximum of f in $\overline{B}_r(x) \subseteq U$. By continuity of f_N and compactness of $\overline{B}_r(x)$, let $x_N \in \overline{B}_r(x)$ be such that $f_N(x_N) \geq f_N(y)$ for every $y \in \overline{B}_r(x)$. Since $(x_N)_{N\geqslant 1}$ stays in a compact set, it admits a subsequence converging to some $x^* \in \overline{B}_r(x)$. Together with the local uniform convergence of f_N to f, this implies that $f(x^*) \geqslant f(y)$ for every $y \in \overline{B}_r(x)$. Since x is a strict maximum, it must be that $x^* = x$, so the only limit point of $(x_N)_{N\geqslant 1}$ is x. This implies that $(x_N)_{N\geqslant 1}$ converges to x as required. In particular, for N sufficiently large x_N is in the open ball $B_r(x)$, and is thus a local maximum of f_N. ■

Using Lemma 2.7, find a sequence $(t^*_\varepsilon, x^*_\varepsilon)_{\varepsilon>0}$ converging to (t^*, x^*) with the property that, for every $\varepsilon > 0$ sufficiently small, the function $f_\varepsilon - \phi$ achieves a local maximum at $(t^*_\varepsilon, x^*_\varepsilon) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, and therefore

$$\partial_t (f_\varepsilon - \phi)(t^*_\varepsilon, x^*_\varepsilon) = 0, \qquad \nabla (f_\varepsilon - \phi)(t^*_\varepsilon, x^*_\varepsilon) = 0, \quad \text{and} \quad \Delta (f_\varepsilon - \phi)(t^*_\varepsilon, x^*_\varepsilon) \leqslant 0. \tag{2.84}$$

It follows by (2.81) that

$$\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t^*_\varepsilon, x^*_\varepsilon) = \big(\partial_t f - \mathsf{H}(\nabla f)\big)(t^*_\varepsilon, x^*_\varepsilon) = \varepsilon \Delta f_\varepsilon(t^*_\varepsilon, x^*_\varepsilon) \leqslant \varepsilon \Delta \phi(t^*_\varepsilon, x^*_\varepsilon).$$

Letting ε tend to zero and leveraging the smoothness of ϕ shows that

$$\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t^*, x^*) \leqslant 0. \tag{2.85}$$

This argument shows that whenever a smooth function ϕ is such that $f - \phi$ has a strict local maximum at $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, the inequality (2.85) holds at (t^*, x^*). An analogous argument shows that whenever a smooth function ϕ is such that $f - \phi$ has a strict local minimum at $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$,

$$\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t^*, x^*) \geqslant 0. \tag{2.86}$$

As will be seen, these properties of f just derived are sufficient to determine it uniquely once the initial condition $f(0,\cdot)$ is fixed. They will therefore be taken as the definition of being a solution to the Hamilton-Jacobi equation (2.80). In fact, given any domain $\mathcal{D} \subseteq \mathbb{R}^d$, they will also be taken as the definition of being a solution to the Hamilton-Jacobi equation

$$\partial_t f(t,x) = \mathsf{H}\big(\nabla f(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathcal{D}. \tag{2.87}$$

It may seem surprising that no boundary condition has to be imposed for this equation; this is because the non-linearity "points in the right direction" in the sense that it is non-decreasing.

Definition 2.8. An upper semi-continuous function $u : \mathbb{R}_{\geqslant 0} \times \mathcal{D} \to \mathbb{R}$ is a *viscosity subsolution* to the Hamilton-Jacobi equation (2.87) if, for every $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathcal{D}$ and $\phi \in C^\infty\big(\mathbb{R}_{>0} \times \mathcal{D}; \mathbb{R}\big)$ with the property that $u - \phi$ has a local maximum at the point $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathcal{D}$,

$$\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t^*, x^*) \leqslant 0. \tag{2.88}$$

A lower semi-continuous function $v : \mathbb{R}_{\geqslant 0} \times \mathcal{D} \to \mathbb{R}$ is a *viscosity supersolution* to the Hamilton-Jacobi equation (2.87) if, for every $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathcal{D}$ and $\phi \in C^\infty\big(\mathbb{R}_{>0} \times \mathcal{D}; \mathbb{R}\big)$ with the property that $v - \phi$ has a local minimum at the point $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathcal{D}$,

$$\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t^*, x^*) \geqslant 0. \tag{2.89}$$

A continuous function $f \in C\big(\mathbb{R}_{\geqslant 0} \times \overline{\mathcal{D}}; \mathbb{R}\big)$ is a *viscosity solution* to to the Hamilton-Jacobi equation (2.87) if it is both a viscosity subsolution and a viscosity supersolution to (2.87).

When $f - \phi$ has a local maximum at (t^*, x^*), one says that "ϕ touches f from above at (t^*, x^*)". The reason is that, when discussing well-posedness, adding a constant to the test function ϕ is irrelevant to the discussion, so it may as well be assumed that indeed $(f - \phi)(t^*, x^*) = 0$. The point (t^*, x^*) is often termed the "contact point". This wording has some intuitive appeal, as one can imagine taking some arbitrary smooth function ϕ that is way above f, and then progressively "sliding it down" until the graphs of ϕ and f touch — at the contact point. This is illustrated in Figure 2.2. Notice that not every point can be "touched" in this way. For instance, there is no smooth function that touches the absolute value function from above at the origin.

As shown in Exercises 3.2 – 3.4 of [50], in the definition of a viscosity subsolution or viscosity supersolution, replacing "local maximum" by "strict local maximum" or by "global maximum", or replacing the requirement that $\phi \in C^\infty(\mathbb{R}_{>0} \times \mathcal{D}; \mathbb{R})$ by the requirement that $\phi \in C^1(\mathbb{R}_{>0} \times \mathcal{D}; \mathbb{R})$ leads to equivalent definitions. In Exercise 3.5 of [50], it is also shown that a continuously differentiable function that satisfies the equation (2.87) everywhere is indeed a viscosity solution, so the notion of viscosity solution is more "permissive" than prescribing the function to be $C^1(\mathbb{R}_{\geqslant 0} \times \overline{\mathcal{D}}; \mathbb{R})$ and to solve the equation everywhere. One can also show that a viscosity solution must satisfy the equation (2.87) at every point of differentiability (Theorem 10.1.2.1 in [53]). By the Rademacher theorem (Theorem 2.10 in [50]), a Lipschitz viscosity solution must therefore satisfy the equation (2.87) almost everywhere. In other words, the notion of a viscosity solution is indeed more stringent than that of an "almost-everywhere solution" explored at the start of this section.

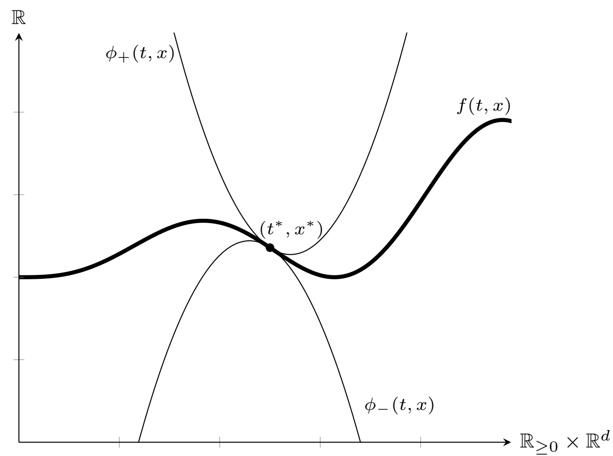

Figure 2.2: The function ϕ_+ touches the function f from above at the point (t^*, x^*) while the function ϕ_- touches the function f from below at the point (t^*, x^*).

The well-posedness of the Hamilton-Jacobi equation (2.80) will be established over the space of functions with Lipschitz initial condition that grow at most linearly in time. Given functions $h : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ and $u : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$, it will be convenient to introduce the semi-norms

$$\|h\|_{\mathrm{Lip}} := \sup_{x \neq x' \in \mathbb{R}^d_{\geqslant 0}} \frac{|h(x) - h(x')|}{|x - x'|} \quad \text{and} \quad [u]_0 := \sup_{\substack{t>0 \\ x \in \mathbb{R}^d_{\geqslant 0}}} \frac{|u(t,x) - u(0,x)|}{t}, \tag{2.90}$$

as well as the space of functions with Lipschitz initial condition that grow at most linearly in time,

$$\mathfrak{L} := \left\{ u : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R} \mid u(0, \cdot) \text{ is Lipschitz continuous and } [u]_0 < +\infty \right\}, \tag{2.91}$$

and its subset of uniformly Lipschitz continuous functions,

$$\mathfrak{L}_{\mathrm{unif}} := \left\{ u \in \mathfrak{L} \mid \sup_{t \geqslant 0} \|u(t, \cdot)\|_{\mathrm{Lip}} < +\infty \right\}. \tag{2.92}$$

The main well-posedness result of this section reads as follows.

Proposition 2.9. *If $\psi : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ is a Lipschitz initial condition, and $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$ is a locally Lipschitz and non-decreasing non-linearity, then the Hamilton-Jacobi equation (2.80) admits a unique viscosity solution $f \in \mathfrak{L}_{\mathrm{unif}}$ with*

$$\sup_{t>0} \|f(t, \cdot)\|_{\mathrm{Lip}} = \|\psi\|_{\mathrm{Lip}}. \tag{2.93}$$

Moreover, if $u,v \in \mathfrak{L}_{\mathrm{unif}}$ *are a continuous subsolution and a continuous supersolution to* (2.80)*, then*

$$\sup_{\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}} \big(u(t,x)-v(t,x)\big) = \sup_{\mathbb{R}^d_{\geqslant 0}} \big(u(0,x)-v(0,x)\big). \tag{2.94}$$

To be more specific, given $\delta_0>0$*, introduce the Lipschitz constants*

$$L := \max\Big(\sup_{t>0}\|u(t,\cdot)\|_{\mathrm{Lip}}, \sup_{t>0}\|v(t,\cdot)\|_{\mathrm{Lip}}\Big) \quad \text{and} \quad V := \sup\left\{\frac{|\mathsf{H}(p')-\mathsf{H}(p)|}{|p'-p|} \;\Big|\; |p|,|p'|\leqslant L+\delta_0\right\}, \tag{2.95}$$

then for every $R\in\mathbb{R}$ *and* $M>2L$*, the map*

$$(t,x)\mapsto u(t,x)-v(t,x)-M\big(|x|+Vt-R\big)_+ \tag{2.96}$$

achieves its supremum on $\{0\}\times\mathbb{R}^d_{\geqslant 0}$*.*

In the statement above, the notation $r_+ := \max(0,r)$ is used to denote the positive part of a real number $r\in\mathbb{R}$. This result will be combined with the Hopf-Lax variational formula to establish Theorem 2.5. Recall the definition of the convex dual in (A.7).

Proposition 2.10 (Hopf-Lax formula)**.** *If* $\psi:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ *is a Lipschitz initial condition, and* $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ *is a locally Lipschitz, non-decreasing, and convex non-linearity, then the Hopf-Lax function*

$$f(t,x) := \sup_{y\in\mathbb{R}^d_{\geqslant 0}}\left(\psi(x+y)-t\mathsf{H}^*\Big(\frac{y}{t}\Big)\right) \tag{2.97}$$

is the unique viscosity solution in $\mathfrak{L}_{\mathrm{unif}}$ *to the Hamilton-Jacobi equation* (2.80)*.*

Since the proofs in this section are quite long and technical, the reader may consider skipping these proofs on the first reading and simply taking Propositions 2.9 and 2.10 for granted. This will allow the reader to get a clearer picture of how the Hamilton-Jacobi approach can be used to study mean-field disordered systems before investing the time and energy required to understand Hamilton-Jacobi equations on positive half-space.

To prove Proposition 2.9, its analog for the Hamilton-Jacobi equation

$$\partial_t f(t,x)-\mathsf{H}\big(\nabla f(t,x)\big)=0 \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}, \tag{2.98}$$

will first be established, and it will then be shown that solutions to the Hamilton-Jacobi equations (2.98) and (2.80) coincide. To be more precise, in Section 2.4.1, a comparison principle for the Hamilton-Jacobi equation (2.98) which ensures the uniqueness of solutions over the space $\mathfrak{L}$ is established. It is also shown that any solution in $\mathfrak{L}$ with a Lipschitz initial condition must in fact belong to the solution space $\mathfrak{L}_{\mathrm{unif}}$. In Section 2.4.2, this comparison principle is leveraged to apply the classical Perron argument to the Hamilton-Jacobi equation (2.98), and show that it admits a solution in the solution space $\mathfrak{L}_{\mathrm{unif}}$. Full details have been provided when applying Perron's method; although the arguments will certainly be seen as classical by experts, the hope is that the reader will find these details helpful. In Section 2.4.3, it is shown that solutions to (2.80) and (2.98) coincide. This is then used to prove Proposition 2.9 by translating the well-posedness theory for the Hamilton-Jacobi equation (2.98) into a well-posedness theory for the Hamilton-Jacobi equation (2.80). Finally, in Section 2.4.4, under appropriate convexity assumptions, the Hopf-Lax formula stated in Proposition 2.10 for the unique solution to the Hamilton-Jacobi equations (2.80) and (2.98) is established.

2.4.1 Comparison principle and Lipschitz continuity of solutions on $\mathbb{R}^d_{\geqslant 0}$

The comparison principle formalizes the idea that the Hamilton-Jacobi equation (2.98) should preserve the ordering of initial conditions. In fact, since the equation is invariant under the addition of a constant to the solution, it will be shown that if u is a subsolution and v is a supersolution to (2.98), then the function $u-v$ achieves its supremum at time zero. In particular, if $u(0,\cdot) \leqslant v(0,\cdot)$, then this ordering is preserved by the evolution.

If one assumes for a moment that u and v are smooth functions, then one can argue heuristically to get a sense of why this result might be true. For smooth functions u and v, saying that u and v are a subsolution and a supersolution to the Hamilton-Jacobi equation (2.98) amounts to saying that

$$\partial_t u - \mathsf{H}(\nabla u) \leqslant 0 \quad \text{and} \quad \partial_t v - \mathsf{H}(\nabla v) \geqslant 0. \tag{2.99}$$

Arguing by contradiction, suppose that

$$\sup_{\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}} (u-v) > \sup_{\{0\} \times \mathbb{R}^d_{\geqslant 0}} (u-v). \tag{2.100}$$

Up to subtracting a small increasing function of t to u, such as εt for some sufficiently small $\varepsilon > 0$, it is possible to ensure that (2.100) is still valid, and also that (2.99) has been improved into

$$\partial_t u - \mathsf{H}(\nabla u) < 0 \quad \text{and} \quad \partial_t v - \mathsf{H}(\nabla v) \geqslant 0. \tag{2.101}$$

Assuming that the supremum on the left side of (2.100) is achieved at some point $(t^*, x^*) \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{>0}$, then it must be that $t^* > 0$, so the first order derivatives of u and v must coincide at this point. This contradicts (2.101), so (2.100) cannot be true. Of course, there is much left to be desired with this argument, since it was assumed that u and v are smooth, and also that the supremum on the left side of (2.100) is achieved at some point outside the boundary of the positive half-space $\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$.

The simplest issue to address is how one can go around the possibility of the supremum not being achieved. To simplify the discussion, assume temporarily that the space $\mathbb{R}^d_{\geqslant 0}$ is replaced by the unit torus $\mathbb{T}^d = \mathbb{R}^d / \mathbb{Z}^d$ so that the variable x lives in a compact space without boundary. Take $u, v : \mathbb{R}_{\geqslant 0} \times \mathbb{T}^d \to \mathbb{R}$ to be a viscosity subsolution and supersolution to (2.98) respectively, and assume that they are smooth. The starting point is again to argue by contradiction, assuming that there exists some time $T > 0$ with

$$\sup_{[0,T] \times \mathbb{T}^d} (u-v) > \sup_{\{0\} \times \mathbb{T}^d} (u-v). \tag{2.102}$$

Denoting $\chi(t) := \frac{\varepsilon}{T-t}$, select $\varepsilon > 0$ sufficiently small that

$$\sup_{[0,T] \times \mathbb{T}^d} (u-v-\chi) > \sup_{\{0\} \times \mathbb{T}^d} (u-v-\chi). \tag{2.103}$$

Since $u-v$ is uniformly bounded over $[0,T] \times \mathbb{T}^d$, it is clear that approximate optimizers of the left side of (2.103) will remain away from the final time T. Using also that $[0,T] \times \mathbb{T}^d$ is compact, construct an optimizer $(t^*, x^*) \in [0,T] \times \mathbb{T}^d$ for this supremum. By construction, it is clear that $t^* \in (0,T)$. Once this is verified, the differential condition at the maximum can be used to ascertain that

$$\partial_t(u-v)(t^*,x^*) - \frac{\varepsilon}{(T-t^*)^2} = \partial_t(u-v-\chi)(t^*,x^*) = 0 \quad \text{and} \quad \nabla(u-v)(t^*,x^*) = 0. \tag{2.104}$$

Since v is a smooth supersolution to the equation, this can be combined with the second inequality in (2.99) to obtain that

$$\frac{\varepsilon}{(T-t^*)^2} \leqslant \big(\partial_t u - \mathsf{H}(\nabla u)\big)(t^*,x^*). \tag{2.105}$$

This contradicts the assumption (2.99) that u is a smooth subsolution to the Hamilton-Jacobi equation (2.98). The role of the perturbation function χ is therefore two-fold. First, it ensures that the optimum t^* is detached from the right endpoint of the interval $[0,T]$, that is, $t^* < T$. Second, it allows one to strengthen the inequalities (2.99) into those in (2.101), if it is understood that the function u is redefined to be $u-\chi$.

To preclude the supremum on the left side of (2.100) from being achieved at a point whose spatial component belongs to the boundary of the positive half-space $\mathbb{R}^d_{\geqslant 0}$, the function χ will be further perturbed using the reciprocal of the function $\mathrm{d}:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ defined by

$$\mathrm{d}(x) := \inf_{\substack{|y|=1\\ y\in\mathbb{R}^d_{\geqslant 0}}} y\cdot x. \tag{2.106}$$

This function essentially measures the distance to the boundary of the domain $\mathbb{R}^d_{\geqslant 0}$ in the sense that it vanishes exactly on the boundary $\partial\mathbb{R}^d_{\geqslant 0}$. This, along with other basic properties of the function d, is the content of the following technical result. Recall that the *superdifferential* of a function $h:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ at a point $x\in\mathbb{R}^d_{>0}$ is the set

$$\partial h(x) := \big\{p\in\mathbb{R}^d \mid h(x') \leqslant h(x) + p\cdot(x'-x) + o(x'-x) \text{ as } x'\to x \text{ in } \mathbb{R}^d_{\geqslant 0}\big\}. \tag{2.107}$$

Lemma 2.11. *The function* $\mathrm{d}:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}_{\geqslant 0}$ *defined in* (2.106) *satisfies the following basic properties.*

(i) The infimum defining $\mathrm{d}(x)$ *is achieved for every* $x\in\mathbb{R}^d_{\geqslant 0}$.

(ii) $\mathrm{d}(x)=0$ *if and only if* $x\in\partial\mathbb{R}^d_{\geqslant 0}$.

(iii) d *is Lipschitz continuous with Lipschitz constant at most one.*

(iv) d *is concave and non-decreasing.*

(v) If $x\in\mathbb{R}^d_{>0}$, *then* $\partial\,\mathrm{d}(x)\subseteq\mathbb{R}^d_{\geqslant 0}$. *Moreover, any* $p\in\partial\,\mathrm{d}(x)$ *is such that* $|p|\leqslant 1$.

(vi) If $h:\mathbb{R}^d\to\mathbb{R}$ *is a differentiable function and* $x\mapsto h(x)-\frac{1}{\mathrm{d}(x)}$ *achieves a local maximum at a point* $x_0\in\mathbb{R}^d_{>0}$, *then* $-\mathrm{d}(x_0)^2\nabla h(x_0)\in\partial\,\mathrm{d}(x_0)$.

Proof. Each property is treated separately.

(i) Consider a sequence $(y_n)_{n\geqslant 1}\subseteq\mathbb{R}^d_{\geqslant 0}$ with $|y_n|=1$ and $y_n\cdot x\to\mathrm{d}(x)$. Since $(y_n)_{n\geqslant 1}$ is uniformly bounded, it admits a subsequential limit $y\in\mathbb{R}^d_{\geqslant 0}$ with $|y|=1$ and $y\cdot x=\mathrm{d}(x)$. The fact that $\mathbb{R}^d_{\geqslant 0}$ is closed has played its part. This shows that the infimum in the definition of $\mathrm{d}(x)$ is attained.

(ii) If $\mathrm{d}(x)=0$, then there exists $y\in\mathbb{R}^d_{\geqslant 0}$ with $|y|=1$ and $y\cdot x=0$. This shows that $x\in\partial\mathbb{R}^d_{\geqslant 0}$. On the other hand, if $x\in\partial\mathbb{R}^d_{\geqslant 0}$, then there exists a non-zero $z\in\mathbb{R}^d_{\geqslant 0}$ with $z\cdot x=0$. Taking $y:=z/|z|$ gives $y\in\mathbb{R}^d_{\geqslant 0}$ with $|y|=1$ and $\mathrm{d}(x)\leqslant y\cdot x=0$. This shows that $\mathrm{d}(x)=0$.

(iii) Fix $x, y \in \mathbb{R}^d_{\geqslant 0}$, and let $z \in \mathbb{R}^d_{\geqslant 0}$ with $|z| = 1$ be such that $\mathrm{d}(y) = z \cdot y$. By the Cauchy-Schwarz inequality,

$$\mathrm{d}(x) - \mathrm{d}(y) \leqslant z \cdot x - z \cdot y = z \cdot (y - x) \leqslant |z||y - x| = |y - x|.$$

Reversing the roles of x and y shows that d is Lipschitz continuous with Lipschitz constant at most one.

(iv) Fix $x, y \in \mathbb{R}^d_{\geqslant 0}$ as well as $t \in [0,1]$, and let $z \in \mathbb{R}^d_{\geqslant 0}$ achieve the infimum for $\mathrm{d}(tx + (1-t)y)$. It is clear that

$$\mathrm{d}(tx + (1-t)y) = z \cdot \big(tx + (1-t)y\big) = t(z \cdot x) + (1-t)(z \cdot y) \geqslant t\,\mathrm{d}(x) + (1-t)\,\mathrm{d}(y).$$

This shows that d is concave. To see that d is non-decreasing, fix $x, x' \in \mathbb{R}^d_{\geqslant 0}$ with $x' - x \in \mathbb{R}^d_{\geqslant 0}$, and let $y \in \mathbb{R}^d_{\geqslant 0}$ attain the infimum defining $\mathrm{d}(x')$. Since $x' - x \in \mathbb{R}^d_{\geqslant 0}$,

$$\mathrm{d}(x') - \mathrm{d}(x) \geqslant y \cdot x' - y \cdot x = (x' - x) \cdot y \geqslant 0$$

as required.

(v) Fix $x \in \mathbb{R}^d_{>0}$, $z \in \mathbb{R}^d_{\geqslant 0}$, and $p \in \partial\,\mathrm{d}(x)$. Notice that $\partial\,\mathrm{d}(x) \neq \varnothing$ by Proposition A.8 as d is concave, and thus $-\mathrm{d}$ is convex. Since $\varepsilon z \in \mathbb{R}^d_{\geqslant 0}$ for every $\varepsilon > 0$, the non-decreasingness of d and the definition of the superdifferential imply that

$$0 \leqslant \mathrm{d}(x + \varepsilon z) - \mathrm{d}(x) \leqslant p \cdot \varepsilon z + o(\varepsilon z).$$

Dividing by ε and letting ε tend to zero shows that $p \cdot z \geqslant 0$ for all $z \in \mathbb{R}^d_{\geqslant 0}$. It follows that $p \in \mathbb{R}^d_{\geqslant 0}$. Now, fix $p \in \partial\,\mathrm{d}(x)$ as well as $y \in \mathbb{R}^d_{\geqslant 0}$ with $|y| = 1$, and find $\varepsilon > 0$ small enough so that $x - \varepsilon y \in \mathbb{R}^d_{\geqslant 0}$. The definition of the superdifferential implies that

$$\mathrm{d}(x - \varepsilon y) \leqslant \mathrm{d}(x) - \varepsilon p \cdot y + o(\varepsilon y).$$

Rearranging and using the 1-Lipschitz continuity of d reveals that

$$\varepsilon p \cdot y \leqslant \varepsilon |y| + o(\varepsilon y) = \varepsilon + o(\varepsilon y).$$

Dividing by ε and letting ε tend to zero shows that $p \cdot y \leqslant 1$ for every $y \in \mathbb{R}^d$ with $|y| = 1$. Choosing $y := p/|p|$ gives $|p| \leqslant 1$.

(vi) Fix $z \in \mathbb{R}^d_{\geqslant 0}$. Since $x_0 \in \mathbb{R}^d_{>0}$ is a local maximum of the map $x \mapsto h(x) - \frac{1}{\mathrm{d}(x)}$, for every $\varepsilon > 0$ small enough,

$$h(x_0) - \frac{1}{\mathrm{d}(x_0)} \geqslant h(x_0 + \varepsilon z) - \frac{1}{\mathrm{d}(x_0 + \varepsilon z)}.$$

Rearranging and using the 1-Lipschitz continuity of d as well as the differentiability of h reveals that

$$\begin{aligned}\mathrm{d}(x_0 + \varepsilon z) &\leqslant \mathrm{d}(x_0) - \mathrm{d}(x_0)\,\mathrm{d}(x_0 + \varepsilon z)\big(h(x_0 + \varepsilon z) - h(x_0)\big)\\ &= \mathrm{d}(x_0) - \mathrm{d}(x_0)^2 \nabla h(x_0) \cdot \varepsilon z + o(\varepsilon z).\end{aligned}$$

This shows that $-\mathrm{d}(x_0)^2 \nabla h(x_0) \in \partial\,\mathrm{d}(x_0)$ and completes the proof. ■

In summary, to establish the comparison principle rigorously, three main issues need to be resolved. The first, and most fundamental, is of course that u and v cannot be assumed to be differentiable. The second is that the spatial variable takes values in $\mathbb{R}^d_{\geqslant 0}$ rather than the torus. The third is that the spatial variable of the perturbed difference $u-v$ should not be maximized on the boundary of $\mathbb{R}^d_{\geqslant 0}$. To tackle the first of these matters, the variables will be doubled, and rather than optimizing $u(t,x)-v(t,x)$, a function that involves $u(t,x)-v(t',x')$ plus a smooth penalty term that strongly encourages (t,x) and (t',x') to stay close together will be optimized instead. This will naturally provide smooth test functions that touch u and v from above and below respectively. The second problem, that the variable x lives in an unbounded space, will be tackled by introducing another "cutoff" function, similar to the function χ used above, but in the space variable. The third problem, that the difference $u-v$ perturbed by the function χ should be maximized away from the boundary $\partial\mathbb{R}^d_{\geqslant 0}$, will be handled by introducing yet another cutoff function that is proportional to the reciprocal of the distance-like function (2.106).

Proposition 2.12 (Comparison principle)**.** *Suppose $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ is a locally Lipschitz and non-decreasing non-linearity, and let $u,v\in\mathfrak{L}_{\mathrm{unif}}$ be a viscosity subsolution and a viscosity supersolution to the Hamilton-Jacobi equation (2.98), respectively. Given $\delta_0>0$, introduce the Lipschitz constants*

$$L:=\max\Big(\sup_{t\geqslant 0}\|u(t,\cdot)\|_{\mathrm{Lip}},\sup_{t\geqslant 0}\|v(t,\cdot)\|_{\mathrm{Lip}}\Big)\quad\text{and}\quad V:=\sup\left\{\frac{|\mathsf{H}(p')-\mathsf{H}(p)|}{|p'-p|}\mid |p|,|p'|\leqslant L+\delta_0\right\}. \tag{2.108}$$

For every $R\in\mathbb{R}$ and $M>2L$, the mapping

$$(t,x)\mapsto u(t,x)-v(t,x)-M\big(|x|+Vt-R\big)_+ \tag{2.109}$$

achieves its supremum at a point in $\{0\}\times\mathbb{R}^d$.

Since the proof of Proposition 2.12 is a bit long, the reader is encouraged to ignore any term related to the cutoff in space $M(|x|+Vt-R)_+$ and its smoothed variants on the first reading, in effect showing the comparison principle with the unbounded space domain $\mathbb{R}^d_{\geqslant 0}$ replaced by the compact space $\mathbb{T}^d$.

Proof of Proposition 2.12. Suppose for the sake of contradiction that there exists $T>0$ with

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}}(u-v-\varphi)>\sup_{\{0\}\times\mathbb{R}^d_{\geqslant 0}}(u-v-\varphi), \tag{2.110}$$

where $\varphi(t,x):=M\big(|x|+Vt-R\big)_+$. The proof proceeds in three steps. First (2.110) is smoothened and perturbed, then a variable doubling argument is used to obtain a system of inequalities, and finally, this system of inequalities is contradicted.

Step 1: smoothing and perturbing. Let $\varepsilon_0\in(0,1)$ be a parameter to be determined, and let $\theta\in C^\infty(\mathbb{R};\mathbb{R})$ be an increasing function such that, for every $r\in\mathbb{R}$,

$$(r-\varepsilon_0)_+\leqslant\theta(r)\leqslant r_+$$

Introduce the function

$$\Phi(t,x):=M\theta\Big(\big(\varepsilon_0+|x|^2\big)^{1/2}+Vt-R\Big)$$

defined on $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$. The choice of θ and the bound $(a+b)_+\leqslant a_++b_+$ imply that

$$\varphi(t,x)\leqslant\Phi(t,x)+M\varepsilon_0\leqslant\varphi(t,x)+M\varepsilon_0^{1/2}+M\varepsilon_0,$$

where the second inequality uses that $(a+b)^{\frac{1}{2}} \leqslant a^{\frac{1}{2}} + b^{\frac{1}{2}}$ for $a, b > 0$. It follows by (2.110) that

$$\sup_{\{0\}\times\mathbb{R}^d_{\geqslant 0}} (u - v - \Phi) < \sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}} (u - v - \Phi) + M\varepsilon_0 + M\varepsilon_0^{1/2},$$

so choosing $\varepsilon_0 > 0$ small enough guarantees that

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}} (u - v - \Phi) > \sup_{\{0\}\times\mathbb{R}^d_{\geqslant 0}} (u - v - \Phi). \tag{2.111}$$

This is a smoothed version of the hypothesis (2.110). A cutoff function in time and a cutoff function in space are now also added to ensure that the supremum in (2.111) is achieved away from the boundary of the domain $\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$. Recall the definition of the distance-like function $\mathrm{d} : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ in (2.106). For small parameters $\varepsilon, \varepsilon', \delta, \delta' > 0$ to be determined, introduce the functions

$$\chi_1(t,x) := \Phi(t,x) + \frac{\varepsilon}{T-t} + \varepsilon' t \quad \text{and} \quad \chi_2(t,x) := \frac{\delta}{\mathrm{d}(x)} + \delta'|x|,$$

defined on $\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$. Choosing $\varepsilon, \varepsilon', \delta, \delta' > 0$ small enough ensures that

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}} (u - v - \chi_1 - \chi_2) > \sup_{\{0\}\times\mathbb{R}^d_{\geqslant 0}} (u - v - \chi_1 - \chi_2). \tag{2.112}$$

This is the smoothed and perturbed version of the hypothesis (2.110) that will be used to reach a contradiction.

Step 2: system of inequalities. For each $\alpha \geqslant 1$, define the function $\Psi_\alpha : [0,T] \times \mathbb{R}^d_{\geqslant 0} \times [0,T] \times \mathbb{R}^d_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R} \cup \{-\infty\}$ by

$$\Psi_\alpha(t,x,t',x',y) := u(t,x) - v(t',x') - \frac{\alpha}{2}\left(|t-t'|^2 + |x-x'|^2 + |x-y|^2\right) - \chi_1(t,x) - \chi_2(t,y). \tag{2.113}$$

It is now argued that the function Ψ_α achieves its supremum at a point $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha, y_\alpha)$ which remains bounded as α tends to infinity with the coordinate y_α also staying bounded away from the boundary $\partial\mathbb{R}^d_{\geqslant 0}$. To do so, write $C < +\infty$ for a constant whose value might change throughout the argument, and which may depend on L, M, R, T, V, $u(0,0)$, $[u]_0$, $v(0,0)$ and $[v]_0$. Fix an arbitrary $y' \in \mathbb{R}^d_{>0}$ as well as some $x \in \mathbb{R}^d_{\geqslant 0}$ with $|x| > R+1$ and $\alpha \geqslant 1$. The bound $\Phi(t,x) \geqslant M\big(|x| + Vt - R - 1\big)_+$ reveals that

$$\begin{aligned}
\Psi_\alpha(t,x,t',x',y) &\leqslant u(0,x) - v(0,x') + t[u]_0 + t'[v]_0 - \frac{\alpha}{2}|x-x'|^2 - \Phi(t,x) - \delta'|y| \\
&\leqslant L\big(|x| + |x'|\big) - \frac{\alpha}{2}|x-x'|^2 - M|x| - \delta'|y| + C \\
&\leqslant (2L - M)|x| + \Big(L - \frac{\alpha}{2}\Big)|x-x'|^2 - \delta'|y| + C.
\end{aligned}$$

Observe also that the supremum of (2.113) is bounded from below by $\Psi_\alpha(0,y',0,y',y')$, which does not depend on α, and that $M > 2L$. This implies that x_α, x'_α and y_α remain bounded as α tends to infinity, and that

$$\alpha\Big(|t_\alpha - t'_\alpha|^2 + |x_\alpha - x'_\alpha|^2 + |x_\alpha - y_\alpha|^2\Big) + \frac{\varepsilon}{T - t_\alpha} + \frac{\delta}{\mathrm{d}(y_\alpha)} \leq C. \tag{2.114}$$

It follows that, up to the extraction of a subsequence, there exist $t_0 \in [0,T]$ and $x_0 \in \mathbb{R}^d_{\geqslant 0}$ such that $t_\alpha \to t_0$,

$t'_\alpha \to t_0$, $x_\alpha \to x_0$, $x'_\alpha \to x_0$ and $y_\alpha \to x_0$ as $\alpha \to +\infty$. By (2.114) and property (ii) in Lemma 2.11, it must be that $t_0 \in [0,T)$ and $x_0 \in \mathbb{R}^d_{>0}$. On the other hand, the semi-continuity of u, v, χ_1 and χ_2 together with the bounds

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}} (u - v - \chi_1 - \chi_2) \leqslant \Psi_\alpha(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha, y_\alpha) \leqslant u(t_\alpha, x_\alpha) - v(t'_\alpha, x'_\alpha) - \chi_1(t_\alpha, x_\alpha) - \chi_2(t_\alpha, y_\alpha)$$

imply that

$$(u - v - \chi_1 - \chi_2)(t_0, x_0) = \sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}} (u - v - \chi_1 - \chi_2).$$

By (2.112), it must therefore be that $t_0 \in (0,T)$ and that $t_\alpha, t'_\alpha \in (0,T)$ for all α large enough. This means that $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha, y_\alpha)_{\alpha \geqslant 1}$ is a sequence of quintuples such that Ψ_α achieves its supremum at $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha, y_\alpha)$, and with $t_\alpha, t'_\alpha \in (0,T)$ and $x_\alpha, x'_\alpha, y_\alpha \in \mathbb{R}^d_{>0}$ for α large enough. With this in mind, fix $\alpha \geqslant 1$ large enough, and introduce the functions $\phi, \phi' \in C^\infty((0,T)\times\mathbb{R}^d_{\geqslant 0};\mathbb{R})$ defined by

$$\phi(t,x) := v(t'_\alpha, x'_\alpha) + \frac{\alpha}{2}\big(|t - t'_\alpha|^2 + |x - x'_\alpha|^2 + |x - y_\alpha|^2\big) + \chi_1(t,x) + \chi_2(t, y_\alpha),$$
$$\phi'(t', x') := u(t_\alpha, x_\alpha) - \frac{\alpha}{2}\big(|t' - t_\alpha|^2 + |x' - x_\alpha|^2 + |x_\alpha - y_\alpha|^2\big) - \chi_1(t_\alpha, x_\alpha) - \chi_2(t_\alpha, y_\alpha).$$

Since $(t_\alpha, x_\alpha, t'_\alpha, x'_\alpha, y_\alpha)$ maximizes Ψ_α, the function $u - \phi$ achieves a local maximum at the point $(t_\alpha, x_\alpha) \in \mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$, while the function $v - \phi'$ achieves a local minimum at $(t'_\alpha, x'_\alpha) \in \mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$. It follows by the definition of a viscosity subsolution and supersolution that

$$\big(\partial_t \phi - \mathsf{H}(\nabla\phi)\big)(t_\alpha, x_\alpha) \leqslant 0 \quad \text{and} \quad \big(\partial_t \phi' - \mathsf{H}(\nabla\phi')\big)(t'_\alpha, x'_\alpha) \geqslant 0. \tag{2.115}$$

This is the system of inequalities that will be contradicted.

Step 3: reaching a contradiction. A direct computation shows that

$$\big(\partial_t \phi - \mathsf{H}(\nabla\phi)\big)(t_\alpha, x_\alpha) = \varepsilon' + \alpha(t_\alpha - t'_\alpha) + \partial_t \Phi(t_\alpha, x_\alpha) + \frac{\varepsilon}{(T - t_\alpha)^2} - \mathsf{H}\big(\nabla\phi(t_\alpha, x_\alpha)\big) \tag{2.116}$$

and

$$\big(\partial_t \phi' - \mathsf{H}(\nabla\phi')\big)(t'_\alpha, x'_\alpha) = \alpha(t_\alpha - t'_\alpha) - \mathsf{H}\big(\nabla\phi'(t'_\alpha, x'_\alpha)\big). \tag{2.117}$$

To compare these two quantities, the non-decreasingness and local Lipschitz continuity of the non-linearity H will be used to replace the gradient

$$\nabla\phi(t_\alpha, x_\alpha) = \alpha(x_\alpha - x'_\alpha) + \alpha(x_\alpha - y_\alpha) + \nabla\Phi(t_\alpha, x_\alpha). \tag{2.118}$$

by the gradient

$$\nabla\phi'(t'_\alpha, x'_\alpha) = \alpha(x_\alpha - x'_\alpha)$$

in (2.116). With the definition of V in mind, it is first shown that

$$|\nabla\phi(t_\alpha, x_\alpha)| \leqslant L \qquad \text{and} \qquad |\nabla\phi'(t'_\alpha, x'_\alpha)| \leqslant L. \tag{2.119}$$

Fix $z \in \mathbb{R}^d$ and $\eta > 0$. Since $u - \phi$ achieves a local maximum at $(t_\alpha, x_\alpha) \in \mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$, and u is uniformly

Lipschitz continuous with Lipschitz constant L,

$$\phi(t_\alpha,x_\alpha+\eta z)-\phi(t_\alpha,x_\alpha)\geqslant u(t_\alpha,x_\alpha+\eta z)-u(t_\alpha,x_\alpha)\geqslant -\eta L|z|.$$

Dividing by η and letting η tend to zero reveals that

$$\nabla\phi(t_\alpha,x_\alpha)\cdot z\geqslant -L|z|.$$

Choosing $z:=-\nabla\phi(t_\alpha,x_\alpha)$ gives the first inequality in (2.119); the second inequality is obtained in an identical manner. These bounds would be sufficient if the term $\alpha|x_\alpha-y_\alpha|$ could be made arbitrarily small. To overcome this issue, the non-decreasingness of H will be leveraged. Since the function $y\mapsto\Psi_\alpha(t_\alpha,x_\alpha,t'_\alpha,x'_\alpha,y)$ achieves its maximum at $y_\alpha\in\mathbb{R}^d_{>0}$, properties (v) and (vi) in Lemma 2.11 imply that

$$\frac{1}{\delta}\,\mathrm{d}(y_\alpha)^2\left(\alpha(y_\alpha-x_\alpha)+\delta'\frac{y_\alpha}{|y_\alpha|}\right)\in\partial\,\mathrm{d}(y_\alpha)\subseteq\mathbb{R}^d_{\geqslant 0}\quad\text{and}\quad \mathrm{d}(y_\alpha)^2\left|\alpha(y_\alpha-x_\alpha)+\delta'\frac{y_\alpha}{|y_\alpha|}\right|\leqslant\delta.$$

Decreasing δ if necessary, and setting $p_\alpha:=\delta'\frac{y_\alpha}{|y_\alpha|}$ gives a vector $p_\alpha\in\mathbb{R}^d_{\geqslant 0}$ with $|p_\alpha|=\delta'$ as well as

$$p_\alpha-\alpha(x_\alpha-y_\alpha)\in\mathbb{R}^d_{\geqslant 0}\quad\text{and}\quad|p_\alpha-\alpha(x_\alpha-y_\alpha)|\leqslant\delta_0. \tag{2.120}$$

To obtain the second bound in this display, after possibly decreasing δ, the fact that $(y_\alpha)_{\alpha\geqslant 1}$ converges to $y_0\in\mathbb{R}^d_{>0}$ has been combined with property (ii) in Lemma 2.11. Remembering (2.118), and combining the non-decreasingness of the non-linearity H with (2.120) yields

$$\mathsf{H}\big(\nabla\phi(t_\alpha,x_\alpha)\big)\leqslant\mathsf{H}\big(\alpha(x_\alpha-x'_\alpha)+p_\alpha+\nabla\Phi(t_\alpha,x_\alpha)\big)\leqslant\mathsf{H}\big(\nabla\phi'(t'_\alpha,x'_\alpha)\big)+V\delta'+V|\nabla\Phi(t_\alpha,x_\alpha)|.$$

The second inequality implicitly uses that by (2.118) and (2.119),

$$|\alpha(x_\alpha-x'_\alpha)+p_\alpha+\nabla\Phi(t_\alpha,x_\alpha)|\leqslant|\nabla\phi(t_\alpha,x_\alpha)|+|p_\alpha-\alpha(x_\alpha-y_\alpha)|\leqslant L+\delta_0.$$

It follows by (2.116) that

$$\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha)>\varepsilon'+\alpha(t_\alpha-t'_\alpha)+\partial_t\Phi(t_\alpha,x_\alpha)-\mathsf{H}\big(\nabla\phi'(t'_\alpha,x'_\alpha)\big)-V\delta'-V|\nabla\Phi(t_\alpha,x_\alpha)|.$$

A direct computation shows that $V|\nabla\Phi(t_\alpha,x_\alpha)|\leqslant\partial_t\Phi(t_\alpha,x_\alpha)$, so in fact

$$\begin{aligned}\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha)&>\varepsilon'+\alpha(t_\alpha-t'_\alpha)-\mathsf{H}\big(\nabla\phi'(t'_\alpha,x'_\alpha)\big)-V\delta'=\big(\partial'_t\phi-\mathsf{H}(\nabla\phi')\big)(t'_\alpha,x'_\alpha)+\varepsilon'-V\delta'\\&\geqslant\varepsilon'-V\delta',\end{aligned}$$

where (2.117) and the second inequality in (2.115) have been used. Choosing $\delta'<\varepsilon'/V$ contradicts the first inequality in (2.115) and completes the proof. ■

The comparison principle is now extended to viscosity solutions in $\mathfrak{L}$ provided that they are initially Lipschitz continuous. This is done by proving that the Lipschitz semi-norm of the solution to the Hamilton-Jacobi equation (2.98) is maximized at the initial time. This will give a uniqueness theory for the Hamilton-Jacobi equation (2.98) on the solution space $\mathfrak{L}$.

Proposition 2.13. *If* $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ *is a locally Lipschitz non-linearity, and* $f\in\mathfrak{L}$ *is a viscosity solution to the Hamilton-Jacobi equation* (2.98)*, then*

$$\sup_{t\geqslant 0}\|f(t,\cdot)\|_{\mathrm{Lip}}=\|f(0,\cdot)\|_{\mathrm{Lip}}. \tag{2.121}$$

Proof. Let $L:=\|f(0,\cdot)\|_{\mathrm{Lip}}$ denote the Lipschitz semi-norm of the initial condition, and suppose for the sake of contradiction that there exists $T>0$ with

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}}\big(f(t,x)-f(t,x')-L|x-x'|\big)>0\geqslant\sup_{\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}}\big(f(0,x)-f(0,x')-L|x-x'|\big). \tag{2.122}$$

The proof proceeds in three steps. First (2.122) is perturbed, then a variable doubling argument is used to obtain a system of inequalities, and finally, this system of inequalities is contradicted.

Step 1: perturbing. Let $\delta_0>0$ be fixed, and let $\theta\in C^\infty(\mathbb{R};\mathbb{R})$ be an increasing function such that, for every $r\in\mathbb{R}$,

$$(r-1)_+\leqslant\theta(r)\leqslant r_+\quad\text{and}\quad|\theta'(r)|\leqslant 1.$$

For a constant $R>0$ to be chosen and the local Lipschitz constant

$$V:=\sup\left\{\frac{|\mathsf{H}(p')-\mathsf{H}(p)|}{|p'-p|}\mid |p|,|p'|\leqslant L+\delta_0\right\},$$

introduce the function

$$\Phi(t,x):=\delta_0\theta\Big(\big(1+|x|^2\big)^{1/2}+Vt-R\Big)$$

defined on $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$. For small parameters $\varepsilon,\varepsilon'>0$ to be determined, introduce the functions

$$\chi_1(t,x):=\Phi(t,x)+\frac{\varepsilon}{T-t}+\varepsilon' t\quad\text{and}\quad\chi_2(t',x'):=\Phi(t',x')+\frac{\varepsilon}{T-t'}$$

defined on $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$. Choosing $R>0$ large enough and $\varepsilon,\varepsilon'>0$ small enough ensures that

$$\begin{aligned}\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}}&\big(f(t,x)-f(t,x')-L|x-x'|-\chi_1(t,x)-\chi_2(t,x')\big)\\&>0\geqslant\sup_{\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}}\big(f(0,x)-f(0,x')-L|x-x'|-\chi_1(0,x)-\chi_2(0,x')\big).\end{aligned} \tag{2.123}$$

This is the perturbed version of the hypothesis (2.122) that will be used to reach a contradiction.

Step 2: system of inequalities. For each $\alpha\geqslant 1$, define the function $\Psi_\alpha:\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}\cup\{-\infty\}$ by

$$\Psi_\alpha(t,x,t',x'):=f(t,x)-f(t',x')-L|x-x'|-\frac{\alpha}{2}|t-t'|^2-\chi_1(t,x)-\chi_2(t',x'). \tag{2.124}$$

It is now argued that the function Ψ_α achieves its supremum at a point $(t_\alpha,x_\alpha,t'_\alpha,x'_\alpha)$ which remains bounded as α tends to infinity. To do so, write $C<+\infty$ for a constant whose value might change throughout the argument, and which may depend on δ_0, R, T, V, $f(0,0)$ and $[f]_0$. For every $x\in\mathbb{R}^d_{\geqslant 0}$ with $|x|>R+1$ and $\alpha\geqslant 1$, the bound $\Phi(t,x) \geqslant \delta_0(|x|+Vt-R-1)$ reveals that

$$\Psi_\alpha(t,x,t',x') \leqslant f(0,x)-f(0,x')+(t+t')[f]_0 - L|x-x'| - \Phi(t,x)-\Phi(t',x') \leqslant -\delta_0(|x|+|x'|)+C.$$

Observe also that the supremum of (2.124) is bounded from below by $\Psi_\alpha(0,0,0,0)$, which does not depend on α. This implies that x_α and x'_α remain bounded as α tends to infinity, and that

$$\alpha|t_\alpha - t'_\alpha|^2 + \frac{\varepsilon}{T-t_\alpha} \leqslant C. \tag{2.125}$$

It follows that, up to the extraction of a subsequence, there exist $t_0 \in [0,T]$ and $x_0, x'_0 \in \mathbb{R}^d_{\geqslant 0}$ such that $t_\alpha \to t_0$, $t'_\alpha \to t_0$, $x_\alpha \to x_0$ and $x'_\alpha \to x'_0$ as $\alpha \to +\infty$. By (2.125) it must be that $t_0 \in [0,T)$. On the other hand, the continuity of f together with the fact that for all $(t,x,x') \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$,

$$f(t_\alpha,x_\alpha)-f(t_\alpha,x'_\alpha)-L|x_\alpha - x'_\alpha| - \chi_1(t_\alpha,x_\alpha)-\chi_2(t_\alpha,x'_\alpha) \geqslant \Psi_\alpha(t_\alpha,x_\alpha,t'_\alpha,x'_\alpha) \geqslant \Psi_\alpha(t,x,t,x'),$$

implies that

$$\begin{aligned} f(t_0,x_0)-f(t_0,x'_0)-L|x_0-x'_0|-\chi_1(t_0,x_0)-\chi_2(t_0,x'_0) \\ \geqslant \sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}} \big(f(t,x)-f(t,x')-L|x-x'|-\chi_1(t,x)-\chi_2(t,x')\big) \end{aligned} \tag{2.126}$$

By (2.123) it must therefore be that $t_0 > 0$. Moreover, if it were the case that $x_0 = x'_0$, then the left side of the inequality (2.123) would be bounded from above by $-\chi_1(t_0,x_0)-\chi_2(t_0,x_0) \leqslant 0$; however, this quantity is strictly positive. This means that $t_\alpha, t'_\alpha \in (0,T)$ and $x_\alpha \neq x'_\alpha$ for α large enough, so $(t_\alpha,x_\alpha,t'_\alpha,x'_\alpha)_{\alpha\geqslant 1}$ is a sequence of quadruples such that Ψ_α achieves its supremum at $(t_\alpha,x_\alpha,t'_\alpha,x'_\alpha)$, and with $t_\alpha,t'_\alpha \in (0,T)$ and $x_\alpha \neq x'_\alpha$ for α large enough. With this in mind, fix $\alpha \geqslant 1$ large enough, and introduce the smooth functions $\phi,\phi' \in C^\infty\big((0,T)\times\mathbb{R}^d_{\geqslant 0};\mathbb{R}\big)$ defined by

$$\begin{aligned} \phi(t,x) &:= f(t'_\alpha,x'_\alpha)+L|x-x'_\alpha|+\frac{\alpha}{2}|t-t'_\alpha|^2+\chi_1(t,x)+\chi_2(t'_\alpha,x'_\alpha), \\ \phi'(t',x') &:= f(t_\alpha,x_\alpha)-L|x'-x_\alpha|-\frac{\alpha}{2}|t'-t_\alpha|^2-\chi_1(t_\alpha,x_\alpha)-\chi_2(t',x'). \end{aligned}$$

Since $(t_\alpha,x_\alpha,t'_\alpha,x'_\alpha)$ maximizes Ψ_α, the function $f-\phi$ achieves a local maximum at $(t_\alpha,x_\alpha) \in \mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}$, while $f-\phi'$ achieves a local minimum at $(t'_\alpha,x'_\alpha) \in \mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}$. It follows by the definition of a viscosity solution that

$$\big(\partial_t\phi - \mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha) \leqslant 0 \quad \text{and} \quad \big(\partial_t\phi' - \mathsf{H}(\nabla\phi')\big)(t'_\alpha,x'_\alpha) \geqslant 0. \tag{2.127}$$

This is the system of inequalities that will be contradicted.

Step 3: reaching a contradiction. A direct computation shows that

$$\big(\partial_t\phi - \mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha) = \varepsilon' + \alpha(t_\alpha - t'_\alpha) + \partial_t\Phi(t_\alpha,x_\alpha) + \frac{\varepsilon}{(T-t_\alpha)^2} - \mathsf{H}\big(\nabla\phi(t_\alpha,x_\alpha)\big), \tag{2.128}$$

and

$$\big(\partial_t\phi' - \mathsf{H}(\nabla\phi')\big)(t'_\alpha,x'_\alpha) = \alpha(t_\alpha - t'_\alpha) - \partial_t\Phi(t'_\alpha,x'_\alpha) - \frac{\varepsilon}{(T-t'_\alpha)^2} - \mathsf{H}\big(\nabla\phi'(t'_\alpha,x'_\alpha)\big). \tag{2.129}$$

To compare these two quantities, the non-decreasingness and local Lipschitz continuity of the non-linearity H

will be used to replace the gradient

$$\nabla \phi(t_\alpha, x_\alpha) = L\frac{x_\alpha - x'_\alpha}{|x_\alpha - x'_\alpha|} + \nabla \Phi(t_\alpha, x_\alpha)$$

by the gradient

$$\nabla \phi'(t'_\alpha, x'_\alpha) = L\frac{x_\alpha - x'_\alpha}{|x_\alpha - x'_\alpha|} - \nabla \Phi(t'_\alpha, x'_\alpha)$$

in (2.128). Since $|\theta'| \leqslant 1$, a direct computation shows that $|\nabla \Phi| \leqslant \delta_0$. Together with the definition of V and the fact that $x_\alpha \neq x'_\alpha$, this implies that

$$\mathsf{H}\big(\nabla \phi(t_\alpha, x_\alpha)\big) \leqslant \mathsf{H}\big(\nabla \phi'(t'_\alpha, x'_\alpha)\big) + V|\nabla \Phi(t_\alpha, x_\alpha)| + V|\nabla \Phi(t'_\alpha, x'_\alpha)|.$$

It follows by (2.128) and (2.129) that

$$\begin{aligned}&\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t_\alpha, x_\alpha)\\ &\quad > \varepsilon' + \partial_t \Phi(t_\alpha, x_\alpha) - V|\nabla \Phi(t_\alpha, x_\alpha)| + \partial_t \Phi(t'_\alpha, x'_\alpha) - V|\nabla \Phi(t'_\alpha, x'_\alpha)| + \big(\partial_t \phi' - \mathsf{H}(\nabla \phi')\big)(t'_\alpha, x'_\alpha)\end{aligned}$$

A direct computation shows that $V|\nabla \Phi(t_\alpha, x_\alpha)| \leqslant \partial_t \Phi(t_\alpha, x_\alpha)$ and $V|\nabla \Phi(t'_\alpha, x'_\alpha)| \leqslant \partial_t \Phi(t'_\alpha, x'_\alpha)$, so in fact

$$\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t_\alpha, x_\alpha) > \varepsilon' + \big(\partial_t \phi' - \mathsf{H}(\nabla \phi')\big)(t'_\alpha, x'_\alpha) \geqslant \varepsilon',$$

where the second inequality in (2.127) has been used. This contradicts the first inequality in (2.127) and completes the proof. ∎

Corollary 2.14. *If* $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$ *is a locally Lipschitz and non-decreasing non-linearity, and* $u, v \in \mathfrak{L}_{\mathrm{unif}}$ *are a viscosity subsolution and a viscosity supersolution to the Hamilton-Jacobi equation (2.98), respectively, then*

$$\sup_{\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}} (u - v) = \sup_{\{0\} \times \mathbb{R}^d_{\geqslant 0}} (u - v). \tag{2.130}$$

Proof. Suppose for the sake of contradiction that there is a point $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$ such that

$$(u - v)(t^*, x^*) > \sup_{\{0\} \times \mathbb{R}^d_{\geqslant 0}} (u - v). \tag{2.131}$$

In the notation of Proposition 2.12, choose $M := 2L + 1$ and $R := |x^*| + Vt^*$, so that

$$u(t^*, x^*) - v(t^*, x^*) - M\big(|x^*| + Vt^* - R\big)_+ = (u - v)(t^*, x^*).$$

By the assumption (2.131), this is strictly greater than

$$\sup_{x \in \mathbb{R}^d} \big(u(0, x) - v(0, x)\big) \geqslant \sup_{x \in \mathbb{R}^d} \big(u(0, x) - v(0, x) - M(|x| - R)_+\big).$$

This contradicts Proposition 2.12 and thus completes the proof. ∎

Corollary 2.15. *If* $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$ *is a locally Lipschitz and non-decreasing non-linearity, and* $u, v \in \mathfrak{L}$ *are viscosity solutions to the Hamilton-Jacobi equation (2.98) with the same initial condition* $u(0, \cdot) = v(0, \cdot)$*, then* $u = v$.

Proof. The definition of the space $\mathfrak{L}$ ensures that u and v have Lipschitz initial conditions, so Proposition 2.13 implies that $u, v \in \mathfrak{L}_{\text{unif}}$. Applying Corollary 2.14 with u treated as a viscosity subsolution and v as a viscosity supersolution reveals that $u \leqslant v$. A symmetric argument shows that $v \leqslant u$, and therefore that $u = v$. This completes the proof. ∎

This settles the uniqueness side of the well-posedness theory for the Hamilton-Jacobi equation (2.98) on the solution space $\mathfrak{L}$. The matter of existence is now discussed.

2.4.2 Existence of solutions on $\mathbb{R}^d_{\geqslant 0}$

The existence of solutions to the Hamilton-Jacobi equation (2.98) can be established using the classical Perron method [17, 39, 48]. Full details of the Perron method will be provided; although the arguments will certainly be seen as classical by experts, the hope is that the reader will find them helpful. This section closely follows Chapter 5 in [17]. It will be convenient to fix a positive constant

$$K > \sup\left\{|\mathsf{H}(y)| \mid y \in \mathbb{R}^d \text{ with } |y| \leqslant \|\psi\|_{\text{Lip}}\right\}, \tag{2.132}$$

and to define the continuous functions $u_\pm : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ by

$$u_\pm(t,x) := \psi(x) \pm Kt. \tag{2.133}$$

The importance of these functions is that they are a viscosity subsolution and a viscosity supersolution to the Hamilton-Jacobi equation (2.98).

Lemma 2.16. *If $\psi : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ is a Lipschitz initial condition, and $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$ is a locally Lipschitz and non-decreasing non-linearity, then the functions u_- and u_+ defined in (2.133) are a subsolution and a supersolution to the Hamilton-Jacobi equation (2.98), respectively.*

Proof. Consider a smooth function $\phi \in C^\infty\left(\mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}; \mathbb{R}\right)$ with the property that $u_- - \phi$ has a local maximum at a point $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$. For any $x \in \mathbb{R}^d_{\geqslant 0}$ and $\varepsilon > 0$,

$$\phi(t^*, x^* + \varepsilon x) - \phi(t^*, x^*) \geqslant u_-(t^*, x^* + \varepsilon x) - u_-(t^*, x^*) = \psi(x^* + \varepsilon x) - \psi(x^*) \geqslant -\varepsilon \|\psi\|_{\text{Lip}} |x|$$

Dividing by ε and letting ε tend to zero shows that for all $x \in \mathbb{R}^d_{\geqslant 0}$,

$$\nabla \phi(t^*, x^*) \cdot x \geqslant -\|\psi\|_{\text{Lip}} |x| = p \cdot x$$

where the vector $p := -\frac{\|\psi\|_{\text{Lip}}}{|x|} x$ has been introduced. It follows that $\nabla \phi(t^*, x^*) - p \in \mathbb{R}^d_{\geqslant 0}$, so the non-decreasingness of H and the fact that $t^* > 0$ imply that

$$\big(\partial_t \phi - \mathsf{H}(\nabla \phi)\big)(t^*, x^*) \leqslant \partial_t u_-(t^*, x^*) - \mathsf{H}(p) = -K - \mathsf{H}(p). \tag{2.134}$$

Remembering the definition of K shows that u_- is a subsolution to the Hamilton-Jacobi equation (2.98). An identical argument shows that u_+ is a supersolution to the Hamilton-Jacobi equation (2.98). This completes the proof. ∎

The Perron method consists of proving that the function $f:\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ defined by

$$f(t,x):=\sup_{u\in\mathcal{S}}u(t,x) \tag{2.135}$$

for the set

$$\mathcal{S}:=\left\{u:\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}\mid u_-\leqslant u\leqslant u_+ \text{ and } \overline{u} \text{ is a subsolution to (2.98)}\right\} \tag{2.136}$$

is a viscosity solution to the Hamilton-Jacobi equation (2.98). See Appendix A.4 for the definitions and basic properties of lower and upper semi-continuous envelopes of a function u, which are denoted by $\underline{u}$ and $\overline{u}$ respectively. The strategy will be to show that $\overline{f}$ is a viscosity subsolution to the Hamilton-Jacobi equation (2.98) while $\underline{f}$ is a viscosity supersolution to this equation. The comparison principle in Corollary 2.14 will then imply that f is a viscosity solution to the Hamilton-Jacobi equation (2.98). Throughout this section,

$$B_r(t^*,x^*):=\left\{(t,x)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}\mid |t-t^*|^2+|x-x^*|^2\leqslant r^2\right\} \tag{2.137}$$

will denote the Euclidean ball of radius $r>0$ centred at the point $(t^*,x^*)\in\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$. It is readily verified that $\overline{f}$ is a subsolution.

Lemma 2.17. *If* $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ *is a locally Lipschitz non-linearity, then the upper semi-continuous envelope* $\overline{f}$ *of the function* (2.135) *is a viscosity subsolution to the Hamilton-Jacobi equation* (2.98)*. In particular,* $f\in\mathcal{S}$*.*

Proof. Consider a smooth function $\phi\in C^\infty\left(\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0};\mathbb{R}\right)$ with the property that $\overline{f}-\phi$ has a strict local maximum at the point $(t^*,x^*)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}$. To be more precise, suppose that

$$\left(\overline{f}-\phi\right)(t^*,x^*)>\left(\overline{f}-\phi\right)(t,x)$$

for all $(t,x)\in B_r(t^*,x^*)\setminus\{(t^*,x^*)\}$. By definition of the upper semi-continuous envelope and continuity of ϕ, it is possible to find points $(t_n,x_n)_{n\geqslant 1}\subseteq B_r(t^*,x^*)$ converging to (t^*,x^*) such that, for every $n\geqslant 1$,

$$(f-\phi)(t_n,x_n)\geqslant(\overline{f}-\phi)(t^*,x^*)-\frac{1}{n}.$$

Similarly, by definition of f, it is possible find a sequence of functions $(u_n)_{n\geqslant 1}\subseteq\mathcal{S}$ such that, for every $n\geqslant 1$,

$$f(t_n,x_n)-\frac{1}{n}\leqslant u_n(t_n,x_n).$$

If $(t_n',x_n')\in B_r(t^*,x^*)$ denotes the maximum of $\overline{u}_n-\phi$ on $B_r(t^*,x^*)$, then the fact that $\overline{u}_n$ is a subsolution to (2.98) implies that

$$\left(\partial_t\phi-\mathsf{H}(\nabla\phi)\right)(t_n',x_n')\leqslant 0. \tag{2.138}$$

Notice that $\overline{u}_n-\phi$ achieves its maximum on the compact set $B_r(t^*,x^*)$ as it is an upper semi-continuous function by Proposition A.13. Remembering that $\overline{u}_n\geqslant u_n$ reveals that

$$(\overline{f}-\phi)(t_n',x_n')\geqslant(\overline{u}_n-\phi)(t_n',x_n')\geqslant(u_n-f)(t_n,x_n)+(f-\phi)(t_n,x_n)\geqslant(\overline{f}-\phi)(t^*,x^*)-\frac{2}{n},$$

where the first inequality uses that $\overline{u}_n\leqslant\overline{f}$ as $u_n\leqslant f$. If (t_∞',x_∞') denotes any subsequential limit of (t_n',x_n'), then the upper semi-continuity of $\overline{f}$ established in Proposition A.13 gives

$$(\overline{f}-\phi)(t'_\infty,x'_\infty)\geqslant(\overline{f}-\phi)(t^*,x^*).$$

Since (t^*,x^*) is a strict local maximum of $\overline{f}-\phi$ on $B_r(t^*,x^*)$, this implies that $(t'_\infty,x'_\infty)=(t^*,x^*)$. Letting n tend to infinity in (2.138) shows that $\overline{f}$ is viscosity subsolution to the Hamilton-Jacobi equation (2.98). It is clear by the definition of f in (2.135) that $u_-\leqslant f\leqslant u_+$, so $f\in\mathcal{S}$. This completes the proof. ■

Showing that $\underline{f}$ is a viscosity supersolution requires more work, and relies upon the following modification of Lemma 2.12 in [17].

Lemma 2.18. *If $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ is a locally Lipschitz and non-decreasing non-linearity, and if $u\in\mathcal{S}$ is such that $\underline{u}$ is not a viscosity supersolution to the Hamilton-Jacobi equation (2.98), then there exist $v\in\mathcal{S}$ and $(t,x)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}$ with $v(t,x)>u(t,x)$.*

Proof. The assumption that $\underline{u}$ is not a viscosity supersolution to the Hamilton-Jacobi equation (2.98) gives $\phi\in C^\infty\big(\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0};\mathbb{R}\big)$ and $(t^*,x^*)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}$ with the property that $\underline{u}-\phi$ has a strict local minimum at (t^*,x^*) but $(\partial_t\phi-\mathsf{H}(\nabla\phi))(t^*,x^*)<0$. To be more precise, there exist $\varepsilon>0$ and $r>0$ such that for all $(t,x)\in B_r(t^*,x^*)\setminus\{(t^*,x^*)\}$,

$$\big(\underline{u}-\phi\big)(t,x)>\big(\underline{u}-\phi\big)(t^*,x^*),\tag{2.139}$$

and for all $(t,x)\in B_r(t^*,x^*)$,

$$\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t,x)<-\varepsilon\tag{2.140}$$

Notice that $\underline{u}(t^*,x^*)<u_+(t^*,x^*)$. Indeed, if this were not the case, the assumption that $u\in\mathcal{S}$ would imply that $\underline{u}(t^*,x^*)=u_+(t^*,x^*)$, and therefore, for $(t,x)\in B_r(t^*,x^*)\setminus\{(t^*,x^*)\}$,

$$\big(u_+-\phi\big)(t,x)\geqslant\big(\underline{u}-\phi\big)(t,x)>\big(\underline{u}-\phi\big)(t^*,x^*)=\big(u_+-\phi\big)(t^*,x^*).$$

In other words, the supersolution u_+ would be such that $u_+-\phi$ achieves a local maximum at (t^*,x^*). This would contradict (2.140). Decreasing $r>0$ if necessary and using the continuity of ϕ and u_+, it is therefore possible to find $\delta>0$ such that for all $(t,x)\in B_r(t^*,x^*)$,

$$\underline{u}(t^*,x^*)+\delta<u_+(t^*,x^*)-\delta\leqslant u_+(t,x)\quad\text{and}\quad\phi(t,x)\leqslant\phi(t^*,x^*)+\frac{\delta}{2}.\tag{2.141}$$

With this in mind, given $\varepsilon'<\frac{1}{4}\min(r^2,\delta)$, introduce the function

$$w(t,x):=\phi(t,x)+\varepsilon'-|x-x^*|^2-|t-t^*|^2+(\underline{u}-\phi)(t^*,x^*),$$

and define $v:\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ by

$$v(t,x):=\begin{cases}\max\big(u(t,x),w(t,x)\big) & \text{if } (t,x)\in B_r(t^*,x^*),\\ u(t,x) & \text{if } (t,x)\notin B_r(t^*,x^*).\end{cases}$$

It is clear from the assumption $u\in\mathcal{S}$ that $v\geqslant u\geqslant u_-$. Moreover, for $(t,x)\in B_r(t^*,x^*)$,

$$w(t,x)\leqslant\phi(t^*,x^*)+\frac{\delta}{2}+\frac{\delta}{2}+(\underline{u}-\phi)(t^*,x^*)=\underline{u}(t^*,x^*)+\delta\leqslant u_+(t,x),$$

where (2.141) and the fact that $\varepsilon' \leqslant \delta/2$ have been used. Together with the assumption $u \in \mathcal{S}$, this shows that $v \leqslant u_+$, and therefore that $u_- \leqslant v \leqslant u_+$. Furthermore, the definition of the lower semi-continuous envelope gives points $(t_n,x_n)_{n\geqslant 1} \subseteq B_r(t^*,x^*)$ with $(t_n,x_n) \to (t^*,x^*)$ and $u(t_n,x_n) \to \underline{u}(t^*,x^*)$. Since $v \geqslant w$ on $B_r(t^*,x^*)$, it follows that

$$\liminf_{n\to+\infty} v(t_n,x_n) \geqslant \liminf_{n\to+\infty} w(t_n,x_n) = \phi(t^*,x^*) + \varepsilon' + (\underline{u}-\phi)(t^*,x^*) = \underline{u}(t^*,x^*) + \varepsilon'.$$

This means that for any n large enough,

$$v(t_n,x_n) \geqslant u(t_n,x_n) + \frac{\varepsilon'}{2} > u(t_n,x_n),$$

so there exists a point $(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$ with $v(t,x) > u(t,x)$. All that remains is to verify that $\overline{v}$ is a subsolution to the Hamilton-Jacobi equation (2.98). Consider $\beta \in C^\infty\big(\mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0};\mathbb{R}\big)$ and $(t_0,x_0) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$ with the property that $\overline{v} - \beta$ has a strict local maximum on $B_{r'}(t_0,x_0)$ at (t_0,x_0). The definition of the upper semi-continuous envelope gives points $(t_n,x_n)_{n\geqslant 1} \subseteq B_{r'}(t_0,x_0)$ converging to (t_0,x_0) with

$$v(t_n,x_n) \geqslant \overline{v}(t_0,x_0) - \frac{1}{n}. \tag{2.142}$$

Since $v(t_n,x_n)$ is either equal to $w(t_n,x_n)$ or $u(t_n,x_n)$, passing to a subsequence, it is possible to assume that $v(t_n,x_n) = w(t_n,x_n)$ for all $n \geqslant 1$ or that $v(t_n,x_n) = u(t_n,x_n)$ for all $n \geqslant 1$. These two cases are treated separately.

Case 1: $v(t_n,x_n) = w(t_n,x_n)$ for all $n \geqslant 1$. In this case, it must be that $(t_n,x_n) \in B_r(t^*,x^*)$ for all $n \geqslant 1$. Indeed, for $(t,x) \notin B_{r/2}(t^*,x^*)$,

$$(w-u)(t,x) \leqslant (w-\underline{u})(t,x) \leqslant (\phi-\underline{u})(t,x) + \varepsilon' - \frac{r^2}{4} + (\underline{u}-\phi)(t^*,x^*) < \varepsilon' - \frac{r^2}{4} \leqslant 0,$$

where (2.139) and the fact that $\varepsilon' < r^2/4$ have been used. If (t_n',x_n') denotes the maximum of $w-\beta$ on $B_r(t^*,x^*) \cap B_{r'}(t_0,x_0)$, arguing as in the proof of Lemma 2.16 shows that

$$\begin{aligned}
\partial_t \beta(t_n',x_n') &= \partial_t w(t_n',x_n') = \partial_t \phi(t_n',x_n') - 2(t_n'-t^*),\\
\nabla \beta(t_n',x_n') &\geqslant \nabla w(t_n',x_n') = \nabla \phi(t_n',x_n') - 2(x_n'-x^*).
\end{aligned}$$

It follows by local Lipschitz continuity and non-decreasingness of the non-linearity H as well as (2.140) that for some large enough constant V,

$$\begin{aligned}
\big(\partial_t\beta - \mathsf{H}(\nabla\beta)\big)(t_n',x_n') &\leqslant \partial_t\phi(t_n',x_n') - \mathsf{H}\big(\nabla\phi(t_n',x_n')\big) + 2|t_n'-t^*| + V|x_n'-x^*|\\
&\leqslant -\varepsilon + 2|t_n'-t^*| + V|x_n'-x^*|.
\end{aligned}$$

Decreasing r if necessary, it is therefore possible to ensure that $\big(\partial_t\beta - \mathsf{H}(\nabla\beta)\big)(t_n',x_n') \leqslant 0$. To leverage this bound, observe that by (2.142), the continuity of β, and the fact that (t_n,x_n) converges to (t_0,x_0),

$$(\overline{v}-\beta)(t_n',x_n') \geqslant (w-\beta)(t_n',x_n') \geqslant (w-\beta)(t_n,x_n) = (v-\beta)(t_n,x_n) \geqslant (\overline{v}-\beta)(t_0,x_0) - \frac{2}{n}.$$

In particular, any subsequential limit (t'_∞, x'_∞) of (t'_n, x'_n) must satisfy

$$(\overline{v}-\beta)(t'_\infty, x'_\infty) \geqslant (\overline{v}-\beta)(t_0, x_0) \quad \text{and} \quad \big(\partial_t \beta - \mathsf{H}(\nabla \beta)\big)(t'_\infty, x'_\infty) \leqslant 0.$$

Since (t_0, x_0) is a strict local maximum of $\overline{v}-\beta$ on $B_{r'}(t_0,x_0)$, the first of these inequalities shows that $(t'_\infty, x'_\infty) = (t_0, x_0)$ while the second implies the required subsolution criterion.

Case 2: $v(t_n,x_n) = u(t_n,x_n)$ for all $n \geqslant 1$. In this case, let (t'_n, x'_n) denote the maximum of $\overline{u}-\beta$ on $B_{r'}(t_0,x_0)$. Since $\overline{u}$ is a viscosity subsolution to the Hamilton-Jacobi equation (2.98),

$$\big(\partial_t \beta - \mathsf{H}(\nabla \beta)\big)(t'_n, x'_n) \leqslant 0.$$

On the other hand, the inequality $v \geqslant u$ and (2.142) reveal that

$$\begin{aligned}(\overline{v}-\beta)(t'_n, x'_n) &\geqslant (\overline{u}-\beta)(t'_n, x'_n) \geqslant (\overline{u}-\beta)(t_n, x_n) \geqslant (u-\beta)(t_n,x_n) = (v-\beta)(t_n,x_n)\\ &\geqslant (\overline{v}-\beta)(t_0,x_0) - \frac{1}{n},\end{aligned}$$

so any subsequential limit (t'_∞, x'_∞) of (t'_n, x'_n) must satisfy

$$(\overline{v}-\beta)(t'_\infty, x'_\infty) \geqslant (\overline{v}-\beta)(t_0, x_0) \quad \text{and} \quad \big(\partial_t \beta - \mathsf{H}(\nabla \beta)\big)(t'_\infty, x'_\infty) \leqslant 0.$$

Since (t_0, x_0) is a strict local maximum of $\overline{v}-\beta$ on $B_{r'}(t_0,x_0)$, the first of these inequalities shows that $(t'_\infty, x'_\infty) = (t_0, x_0)$ while the second implies the required subsolution criterion. This completes the proof. ∎

Corollary 2.19. *If $\mathsf{H}:\mathbb{R}^d \to \mathbb{R}$ is a locally Lipschitz and non-decreasing non-linearity, then the lower semi-continuous envelope $\underline{f}$ of the function* (2.135) *is a viscosity supersolution to the Hamilton-Jacobi equation* (2.98)*.*

Proof. Suppose for the sake of contradiction that $\underline{f}$ is not a supersolution to the Hamilton-Jacobi equation (2.98). Combining Lemmas 2.17 and 2.18 gives a function $v \in \mathcal{S}$ and a point $(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$ with $v(t,x) > f(t,x)$. The contradiction

$$f(t,x) = \sup_{u \in \mathcal{S}} u(t,x) \geqslant v(t,x) > f(t,x)$$

completes the proof. ∎

Together with Lemma 2.17, the comparison principle in Corollary 2.14, and the Lipschitz bound in Proposition 2.13, this result allows one to establish the well-posedness of the Hamilton-Jacobi equation (2.98).

Proposition 2.20. *If $\psi:\mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ is a Lipschitz initial condition, and $\mathsf{H}:\mathbb{R}^d \to \mathbb{R}$ is a locally Lipschitz and non-decreasing non-linearity, then the Hamilton-Jacobi equation* (2.98) *admits a unique viscosity solution $f \in \mathfrak{L}$ subject to the initial condition ψ. Moreover,*

$$\sup_{t>0} \|f(t,\cdot)\|_{\mathrm{Lip}} = \|\psi\|_{\mathrm{Lip}}. \tag{2.143}$$

Proof. Denote by f the function defined in (2.135). Combining Lemma 2.17 and Corollary 2.19 shows that $\overline{f}$ is a viscosity subsolution to the Hamilton-Jacobi equation (2.98) while $\underline{f}$ is a viscosity supersolution to this equation. By Proposition A.13 and continuity of u_- and u_+,

$$u_- \leqslant \underline{f} \leqslant f \leqslant \overline{f} \leqslant u_+.$$

Moreover, any function $h : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ with $u_- \leqslant h \leqslant u_+$ satisfies the bounds

$$\begin{aligned}
\psi(x) = u_-(0,x) \leqslant h(0,x) &\leqslant u_+(0,x) = \psi(x),\\
h(t,x) - h(0,x) &\leqslant u_+(t,x) - \psi(x) = Kt,\\
h(0,x) - h(t,x) &\leqslant \psi(x) - u_-(t,x) \leqslant Kt,\\
h(t,x) - h(t,x') &\leqslant u_+(t,x) - u_-(t,x') = \psi(x) - \psi(x') \leqslant \|\psi\|_{\mathrm{Lip}}|x-x'|,\\
h(t,x) - h(t,x') &\geqslant u_-(t,x) - u_+(t,x') = \psi(x) - \psi(x') \geqslant -\|\psi\|_{\mathrm{Lip}}|x-x'|,
\end{aligned}$$

for all $t > 0$ and every $x, x' \in \mathbb{R}^d_{\geqslant 0}$. In particular, it belongs to the solution space $\mathfrak{L}_{\mathrm{unif}}$. This means that $\underline{f}, f, \overline{f} \in \mathfrak{L}_{\mathrm{unif}}$. It follows by the comparison principle in Corollary 2.14 that $\overline{f} \leqslant \underline{f}$. Since $\underline{f} \leqslant f \leqslant \overline{f}$ by Proposition A.13, it must be that $f = \underline{f} = \overline{f}$. In particular, the function $f \in \mathfrak{L}$ is a continuous viscosity solution to the Hamilton-Jacobi equation (2.98). The uniqueness of such a viscosity solution is guaranteed by Corollary 2.15. Recalling Proposition 2.13 gives the Lipschitz bound (2.143) and completes the proof. ∎

This settles the well-posedness theory for the Hamilton-Jacobi equation (2.98) on the solution space $\mathfrak{L}$. The non-decreasingness of the non-linearity H is now leveraged to show that solutions to (2.80) and (2.98) coincide. This allows translation of the well-posedness theory just established for (2.98) into a well-posedness theory for the Hamilton-Jacobi equation (2.80), thereby proving Proposition 2.9.

2.4.3 Equivalence of solutions on $\mathbb{R}^d_{\geqslant 0}$ and $\mathbb{R}^d_{>0}$

The notion of solution to the Hamilton-Jacobi equation (2.98) does not require the imposition of a boundary condition. Intuitively, this is possible because the characteristic lines always go towards the boundary as t increases rather than away from it, and it suggests that the boundary can simply be ignored. In other words, solutions to the Hamilton-Jacobi equations (2.98) and (2.80) should coincide. To show that this is indeed the case, one can argue as in Proposition 2.1 of [33] which is inspired by [43, 104]. The main difference between [33] and [43, 104] is in the definition of the distance-like function (2.106) to the boundary of the domain on which the Hamilton-Jacobi equation is defined.

Proposition 2.21. *If $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$ is a locally Lipschitz and non-decreasing non-linearity, then a continuous function $u : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ is a viscosity subsolution to the Hamilton-Jacobi equation* (2.98) *if and only if it is a viscosity subsolution to the Hamilton-Jacobi equation* (2.80). *An identical statement holds for viscosity supersolutions.*

Proof. The argument for viscosity subsolutions and viscosity supersolutions being almost identical, the focus is exclusively on the case of viscosity subsolutions. To begin with, suppose that u is a viscosity subsolution to the Hamilton-Jacobi equation (2.98), and let $\phi \in C^\infty\left(\mathbb{R}_{>0} \times \mathbb{R}^d_{>0}; \mathbb{R}\right)$ be a function with the property that $u - \phi$ has a local maximum at $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$. After modifying ϕ outside a small enough neighbourhood of (t^*, x^*) so that it becomes a smooth function defined on the larger domain $\mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$, apply the subsolution criterion for (2.98) to obtain the result.

Conversely, suppose that u is a continuous viscosity subsolution to the Hamilton-Jacobi equation (2.80), and consider a smooth function $\phi \in C^\infty\left(\mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}; \mathbb{R}\right)$ with the property that $u - \phi$ has a local maximum at $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$. If $x^* \in \mathbb{R}^d_{>0}$ there is nothing to prove, so assume that $x^* \in \partial \mathbb{R}^d_{\geqslant 0}$. Perturbing the test function ϕ by a small quadratic function if necessary, suppose further that (t^*,x^*) is a strict local maximum of $u-\phi$. To be more precise, assume that

$$u(t,x) - \phi(t,x) < u(t^*,x^*) - \phi(t^*,x^*) \tag{2.144}$$

for any (t,x) other than (t^*,x^*) in the closure of the open neighbourhood

$$\mathcal{O}_r := (t^* - r, t^* + r) \times \left(B_r(x^*) \cap \mathbb{R}^d_{>0}\right) \subseteq \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}.$$

The proof of the subsolution criterion for (2.98) now proceeds in three steps. First, it is shown that there exists an almost maximizer of $u-\phi$ in $\mathcal{O}_r$, then a variable doubling argument is used to obtain a system of inequalities, and finally, this system of inequalities is leveraged to establish the required subsolution criterion.

Step 1: almost maximizer of $u-\phi$ in $\mathcal{O}_r$. Recall the definition of the distance-like function $\mathrm{d}:\mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ in (2.106). For a small parameter $\varepsilon > 0$ to be determined, introduce the function

$$\psi_\varepsilon(s,y) := u(s,y) - \phi(s,y) - \frac{\varepsilon}{\mathrm{d}(y)}$$

defined on $\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$. Since ψ_ε is upper semi-continuous with values in $\mathbb{R} \cup \{-\infty\}$, it is maximized on the compact set $\overline{\mathcal{O}}_r$ at some point $(s_\varepsilon, y_\varepsilon) \in \overline{\mathcal{O}}_r$. The strict inequality (2.144) implies that choosing $\varepsilon > 0$ small enough, it is actually possible to ensure that $(s_\varepsilon, y_\varepsilon) \in \mathcal{O}_r$. Indeed, the continuity of $u-\phi$ gives $(t,x) \in \mathcal{O}_r$ independent of $\varepsilon > 0$ such that for $\varepsilon > 0$ small enough and $(s,y) \in \overline{\mathcal{O}}_r$ with $s \in \{t^* - r, t^* + r\}$ or $y \in \partial B_r(x^*) \cap \mathbb{R}^d_{\geqslant 0}$,

$$\psi_\varepsilon(s,y) < \psi_\varepsilon(t,x).$$

Due to the term $\varepsilon/\mathrm{d}(y)$ in the definition of ψ_ε, this inequality extends to all $(s,y) \in \partial\overline{\mathcal{O}}_r$, thus showing that $(s_\varepsilon, y_\varepsilon) \in \mathcal{O}_r$.

Step 2: system of inequalities. Fix a smooth function $\zeta_\varepsilon : \mathbb{R} \times \mathbb{R}^d \to [0,1]$ with

$$\operatorname{supp} \zeta_\varepsilon \subseteq \mathcal{O}_r \quad \text{and} \quad \zeta_\varepsilon(s_\varepsilon, y_\varepsilon) = 1, \tag{2.145}$$

and for each $\theta > 0$ define the θ-modulus of continuity of u on $\overline{\mathcal{O}}_r$,

$$\omega_u(\theta) := \sup\left\{|u(t,x) - u(s,y)| \mid (t,x),(s,y) \in \overline{\mathcal{O}}_r \text{ and } |t-s|^2 + |x-y|^2 \leqslant \theta^2\right\}.$$

Given a small parameter $\delta > 0$ that will eventually be sent to zero, use the uniform continuity of u on the compact set $\overline{\mathcal{O}}_r$ to find $\theta = \theta(\delta) > 0$ with

$$\omega_u(\theta) < \delta. \tag{2.146}$$

Define the function $\Psi_{\varepsilon,\delta,\theta} : \overline{\mathcal{O}}_r \times \overline{\mathcal{O}}_r \to \mathbb{R}$ by

$$\Psi_{\varepsilon,\delta,\theta}(t,x,s,y) := u(t,x) - \phi(s,y) - \frac{\varepsilon}{\mathrm{d}(y)} - \frac{2M_u}{\theta^2}\left(|t-s|^2 + |x-y|^2\right) + \delta\zeta_\varepsilon(s,y),$$

where $M_u := \sup_{\overline{\mathcal{O}}_r} |u|$. It is now argued that the function $\Psi_{\varepsilon,\delta,\theta}$ achieves its supremum at a point $(t_0,x_0,s_0,y_0) \in \mathcal{O}_r \times \mathcal{O}_r$. Given $(t,x,s,y) \in \overline{\mathcal{O}}_r \times \overline{\mathcal{O}}_r$ with $|t-s|^2 + |x-y|^2 \leqslant \theta^2$, the triangle inequality and the definition of the modulus of continuity reveal that

$$\Psi_{\varepsilon,\delta,\theta}(t,x,s,y) \leqslant \omega_u(\theta) + u(s,y) - \phi(s,y) - \frac{\varepsilon}{\mathrm{d}(y)} + \delta\zeta_\varepsilon(s,y) = \omega_u(\theta) + \psi_\varepsilon(s,y) + \delta\zeta_\varepsilon(s,y).$$

On the other hand, given $(t,x,s,y) \in \overline{\mathcal{O}}_r \times \overline{\mathcal{O}}_r$ with $|t-s|^2 + |x-y|^2 > \theta^2$, the triangle inequality, the definition of M_u, and the non-negativity of the modulus of continuity imply that

$$\Psi_{\varepsilon,\delta,\theta}(t,x,s,y) \leqslant u(s,y) - \phi(s,y) - \frac{\varepsilon}{\mathrm{d}(y)} + \delta\zeta_\varepsilon(s,y) \leqslant \omega_u(\theta) + \psi_\varepsilon(s,y) + \delta\zeta_\varepsilon(s,y).$$

It follows that for every $(t,x) \in \overline{\mathcal{O}}_r$ and $(s,y) \in \overline{\mathcal{O}}_r \setminus \operatorname{supp}\zeta_\varepsilon$,

$$\begin{aligned}\Psi_{\varepsilon,\delta,\theta}(t,x,s,y) &\leqslant \omega_u(\theta) + \psi_\varepsilon(s_\varepsilon,y_\varepsilon) + \delta\zeta_\varepsilon(s,y) = \omega_u(\theta) + \Psi_{\varepsilon,\delta,\theta}(s_\varepsilon,y_\varepsilon,s_\varepsilon,y_\varepsilon) - \delta \\ &< \Psi_{\varepsilon,\delta,\theta}(s_\varepsilon,y_\varepsilon,s_\varepsilon,y_\varepsilon),\end{aligned}$$

where (2.145) and (2.146) have been used. It must therefore be that $(s_0,y_0) \in \operatorname{supp}\zeta_\varepsilon \subseteq \mathcal{O}_r$. To show that (t_0,x_0) also belongs to this open set, suppose that $|t_0-s_0|^2 + |x_0-y_0|^2 > \theta^2$. The triangle inequality and the definition of M_u ensure that

$$\Psi_{\varepsilon,\delta,\theta}(t_0,x_0,s_0,y_0) \leqslant u(t_0,x_0) - \phi(s_0,y_0) - \frac{\varepsilon}{\mathrm{d}(y_0)} - 2M_u + \delta\zeta_\varepsilon(s_0,y_0) \leqslant \Psi_{\varepsilon,\delta,\theta}(s_0,y_0,s_0,y_0),$$

so, up to replacing (t_0,x_0) with (s_0,y_0), assume without loss of generality that

$$|t_0-s_0|^2 + |x_0-y_0|^2 \leqslant \theta^2. \tag{2.147}$$

Decreasing θ if necessary and recalling that $\mathcal{O}_r$ is open shows that indeed $(t_0,x_0,s_0,y_0) \in \mathcal{O}_r \times \mathcal{O}_r$. With this in mind, introduce the functions $\varphi \in C^\infty(\mathbb{R}_{>0} \times \mathbb{R}^d_{>0};\mathbb{R})$ and $h \in C^\infty(\mathbb{R}_{>0};\mathbb{R})$ defined by

$$\begin{aligned}\varphi(t,x) &:= \phi(s_0,y_0) + \frac{\varepsilon}{\mathrm{d}(y_0)} + \frac{2M_u}{\theta^2}\big(|t-s_0|^2 + |x-y_0|^2\big) + \delta\zeta_\varepsilon(s_0,y_0), \\ h(y) &:= \frac{1}{\varepsilon}\Big(u(t_0,x_0) - \phi(s_0,y) - \frac{2M_u}{\theta^2}\big(|t_0-s_0|^2 + |y-x_0|^2\big) + \delta\zeta_\varepsilon(s_0,y)\Big).\end{aligned}$$

Since (t_0,x_0,s_0,y_0) maximizes $\Psi_{\varepsilon,\delta,\theta}$, the function $u-\varphi$ achieves a local maximum at the point $(t_0,x_0) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, the function $y \mapsto h(y) - \frac{1}{\mathrm{d}(y)}$ achieves a local maximum at the point $y_0 \in \mathbb{R}_{>0}$, and the function $s \mapsto \Psi_{\varepsilon,\delta,\theta}(t_0,x_0,s,y_0)$ achieves a local maximum at the point $t_0 \in \mathbb{R}_{>0}$. It follows by the definition of a viscosity subsolution and property (vi) in Lemma 2.11 that

$$\big(\partial_t\varphi - \mathsf{H}(\nabla\varphi)\big)(t_0,x_0) \leqslant 0, \quad -\mathrm{d}(y_0)^2\nabla h(y_0) \in \partial\,\mathrm{d}(y_0) \quad \text{and} \quad \partial_s\Psi_{\varepsilon,\delta,\theta}(t_0,x_0,s_0,y_0) = 0. \tag{2.148}$$

This is the system of inequalities that will be leveraged to verify the subsolution criterion for ϕ at the contact point (t^*,x^*).

Step 3: subsolution criterion. A direct computation shows that

$$\big(\partial_t\varphi - \mathsf{H}(\nabla\varphi)\big)(t_0,x_0) = \frac{4M_u}{\theta^2}(t_0-s_0) - \mathsf{H}\bigg(\frac{4M_u}{\theta^2}(x_0-y_0)\bigg) \leqslant 0, \tag{2.149}$$

and

$$-\mathrm{d}(y_0)^2\nabla h(y_0) = \frac{\mathrm{d}(y_0)^2}{\varepsilon}\left(\nabla\phi(s_0,y_0) + \frac{4M_u}{\theta^2}(y_0 - x_0) - \delta\zeta_\varepsilon(s_0,y_0)\right) \in \partial\,\mathrm{d}(y_0).$$

Invoking property (v) in Lemma 2.11 gives $p \in \mathbb{R}^d_{\geqslant 0}$ with

$$\frac{4M_u}{\theta^2}(x_0 - y_0) = \nabla\phi(s_0,y_0) - p - \delta\zeta_\varepsilon(s_0,y_0),$$

Together with (2.149) and the non-decreasingness of the non-linearity H, this implies that

$$\frac{4M_u}{\theta^2}(t_0 - s_0) - \mathsf{H}\big(\nabla\phi(s_0,y_0) - \delta\zeta_\varepsilon(s_0,y_0)\big) \leqslant 0.$$

Another direct computation reveals that

$$-\partial_t\Psi_{\varepsilon,\delta,\theta}(t_0,x_0,s_0,y_0) = \partial_s\phi(s_0,y_0) - \frac{4M_u}{\theta^2}(t_0 - s_0) - \delta\partial_t\zeta_\varepsilon(s_0,y_0) = 0,$$

so in fact,

$$\partial_t\phi(s_0,y_0) - \mathsf{H}\big(\nabla\phi(s_0,y_0) - \delta\zeta_\varepsilon(s_0,y_0)\big) - \delta\partial_t\zeta_\varepsilon(s_0,y_0) \leqslant 0. \tag{2.150}$$

Recalling that $(s_0,y_0) \in \operatorname{supp}\zeta_\varepsilon$ depends on ε,δ and θ, and that $\theta = \theta(\delta)$ was chosen small enough in terms of δ, one would now like to let θ, then δ, and finally ε tend to zero to establish the subsolution criterion at the contact point $(t^*,x^*) \in \mathbb{R}_{>0}\times\partial\mathbb{R}^d_{\geqslant 0}$. Denote by $(t_1,x_1) \in \operatorname{supp}\zeta_\varepsilon \subseteq \mathcal{O}_r$ and $(s_1,y_1) \in \operatorname{supp}\zeta_\varepsilon \subseteq \mathcal{O}_r$ subsequential limits of the sequences (t_0,x_0) and (s_0,y_0) as θ and then δ tend to zero. The bound (2.147) ensures that $s_1 = t_1$ and $y_1 = x_1$. On the other hand, the fact that for $(t,x) \in \mathcal{O}_r$,

$$u(t_0,x_0) - \phi(s_0,y_0) - \frac{\varepsilon}{\mathrm{d}(y_0)} + \delta\zeta_\varepsilon(s_0,y_0) \geqslant \Psi_{\varepsilon,\delta,\theta}(t_0,x_0,s_0,y_0) \geqslant \psi_\varepsilon(t,x) + \delta\zeta_\varepsilon(t,x),$$

and the continuity of u,ϕ,d, and ζ_ε, imply that for all $(t,x) \in \mathcal{O}_r$,

$$u(t_1,x_1) - \phi(t_1,x_1) \geqslant u(t_1,x_1) - \phi(t_1,x_1) - \frac{\varepsilon}{\mathrm{d}(x_1)} \geqslant u(t,x) - \phi(t,x) - \frac{\varepsilon}{\mathrm{d}(x)}.$$

At this point, denote by $(t_2,x_2) \in \overline{\mathcal{O}_r}$ a subsequential limit of the sequence (t_1,x_1) as ε tends to zero. By continuity of u and ϕ, for all $(t,x) \in \mathcal{O}_r$,

$$u(t_2,x_2) - \phi(t_2,x_2) \geqslant u(t,x) - \phi(t,x).$$

Since $(t^*,x^*) \in \mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}$ is a strict local maximum of $u - \phi$ on $\overline{\mathcal{O}_r}$, and this inequality extends to the closure $\overline{\mathcal{O}_r}$ by continuity of u and ϕ, it must be the case that $t_2 = t^*$ and $x_2 = x^*$. It follows by letting δ, then θ, and finally ε tend to zero in (2.150) that

$$\big(\partial_t\phi - \mathsf{H}(\nabla\phi)\big)(t^*,x^*) \leqslant 0.$$

This completes the proof. ■

Proof of Proposition 2.9. This is now an immediate consequence of Corollary 2.14 and of Propositions 2.12, 2.20 and 2.21. ■

2.4.4 Variational representations of viscosity solutions

In the context of the Hamilton-Jacobi approach, a variational formula for the solution to the Hamilton-Jacobi equation (2.80) allows one to go beyond the identification of the limit free energy as the unique viscosity solution to a Hamilton-Jacobi equation. Indeed, it allows one to write an explicit variational formula such as (2.74) for this limit free energy. Two classical variational formulas will be stated, the Hopf-Lax formula in the setting when the non-linearity H is convex, and the Hopf formula in the setting when the initial condition ψ is convex. Only the Hopf-Lax formula will be proved as this is the variational formula that will be used to establish Theorem 2.5. These variational formulas give an alternative proof to the existence of solutions to the Hamilton-Jacobi equation (2.80) under the stated convexity assumptions.

In the Euclidean setting, under different assumptions, the Hopf-Lax and the Hopf formulas allow one to write the solution to the Hamilton-Jacobi equation

$$\partial_t f(t,x) - \mathsf{H}\big(\nabla f(t,x)\big) = 0 \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^d \tag{2.151}$$

with initial condition $\psi:\mathbb{R}^d\to\mathbb{R}$ as a saddle-point problem for the functional defined, for each $(t,x)\in\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d$ and $(y,p)\in\mathbb{R}^d\times\mathbb{R}^d$, by

$$\mathcal{J}_{t,x}(y,p) := \psi(y) + p\cdot(x-y) + t\mathsf{H}(p). \tag{2.152}$$

More precisely, the Hopf-Lax formula states that when the non-linearity H is convex, then the function

$$f(t,x) := \sup_{y\in\mathbb{R}^d}\inf_{p\in\mathbb{R}^d}\mathcal{J}_{t,x}(y,p) = \sup_{y\in\mathbb{R}^d}\bigg(\psi(y) - t\mathsf{H}_*\Big(\frac{y-x}{t}\Big)\bigg) \tag{2.153}$$

is the unique viscosity solution to the Hamilton-Jacobi equation (2.151). Here H_* denotes the convex dual of H defined in (A.4). Similarly, the Hopf formula states that when the initial condition ψ is convex, then the function

$$f(t,x) := \sup_{p\in\mathbb{R}^d}\inf_{y\in\mathbb{R}^d}\mathcal{J}_{t,x}(y,p) = \sup_{p\in\mathbb{R}^d}\inf_{y\in\mathbb{R}^d}\big(\psi(y) + p\cdot(x-y) + t\mathsf{H}(p)\big) \tag{2.154}$$

is the unique viscosity solution to the Hamilton-Jacobi equation (2.151). Notice that for each fixed $(y,p)\in\mathbb{R}^d\times\mathbb{R}^d$, the mapping $(t,x)\mapsto\mathcal{J}_{t,x}(y,p)$ is a solution to (2.151), so the Hopf-Lax and Hopf formulas can be thought of as representations of the solution with initial condition ψ as envelopes of this family of solutions. At a point of differentiability of this envelope function, the function will be tangent to one particular solution in this set indexed by $(y,p)\in\mathbb{R}^d\times\mathbb{R}^d$, and since the equation is of first order, a function that is tangent to a smooth solution at a point must solve the equation at that point. Of course, this is only an informal discussion, and the interested reader is referred to Chapter 3 in [50] for precise statements and rigorous proofs of the Hopf-Lax and Hopf formulas in the Euclidean setting.

The Hopf-Lax and Hopf formulas admit natural analogs in the setting of positive half-space. A first guess at what these may be is that the optimization domain $\mathbb{R}^d\times\mathbb{R}^d$ in (2.153) and (2.154) is simply replaced by $\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$. This guess turns out to be correct for the Hopf formula provided that the initial condition is non-decreasing.

Proposition 2.22 (Hopf formula)**.** *If $\psi:\mathbb{R}^d\to\mathbb{R}$ is a Lipschitz, non-decreasing and convex initial condition,* *and* $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ *is a locally Lipschitz and non-decreasing non-linearity, then the Hopf function*

$$f(t,x):=\sup_{p\in\mathbb{R}^d_{\geqslant 0}}\inf_{y\in\mathbb{R}^d_{\geqslant 0}}\mathcal{J}_{t,x}(y,p)=\sup_{p\in\mathbb{R}^d_{\geqslant 0}}\inf_{y\in\mathbb{R}^d_{\geqslant 0}}\big(\psi(y)+p\cdot(x-y)+t\mathsf{H}(p)\big) \tag{2.155}$$

is the unique viscosity solution in $\mathfrak{L}$ *to the Hamilton-Jacobi equation* (2.80).

Proof. See Proposition 6.3 in [33]. ∎

To the best of the author's knowledge, for the Hopf-Lax formula, this guess turns out to be too naive; it is more convenient to replace the optimization domain $\mathbb{R}^d$ and the convex dual H_* by the optimization domain $\mathbb{R}^d_{\geqslant 0}$ and the convex dual H^* in the representation

$$f(t,x)=\sup_{y\in\mathbb{R}^d}\left(\psi(x+y)-t\mathsf{H}_*\Big(\frac{y}{t}\Big)\right) \tag{2.156}$$

of the Euclidean Hopf-Lax function (2.153). A precise statement of the Hopf-Lax formula for positive half-space is given in Proposition 2.10. To prove this result, it is first verified that the Hopf-Lax function (2.97) satisfies the right initial condition, and that the supremum in its definition is attained. It is then shown that the Hopf-Lax function satisfies a semigroup property from which it is deduced that it belongs to the solution space $\mathfrak{L}$. Finally, it is shown that the Hopf-Lax function is the unique viscosity solution in $\mathfrak{L}$ to the Hamilton-Jacobi equation (2.80).

Lemma 2.23. *Under the assumptions of Proposition 2.10, the Hopf-Lax function* (2.97) *satisfies the right initial condition,*

$$f(0,\cdot)=\psi(\cdot). \tag{2.157}$$

Proof. At $t=0$, the formula in (2.97) is interpreted as

$$f(0,x)=\sup_{y\in\mathbb{R}^d_{\geqslant 0}}\inf_{p\in\mathbb{R}^d_{\geqslant 0}}\big(\psi(x+y)-p\cdot y\big).$$

Taking $y=0$ on the right side of this expression gives the lower bound

$$f(0,x)\geqslant\psi(x).$$

On the other hand, given $y\in\mathbb{R}^d_{\geqslant 0}$, choosing $p:=\|\psi\|_{\mathrm{Lip}}\frac{y}{|y|}$ gives the upper bound

$$f(0,x)=\sup_{y\in\mathbb{R}^d}\big(\psi(x+y)-\|\psi\|_{\mathrm{Lip}}|y|\big)\leqslant\psi(x),$$

where the Lipschitz continuity of the initial condition ψ has been used. Combining these lower and upper bounds completes the proof. ∎

Lemma 2.24. *Under the assumptions of Proposition 2.10, for any* $(t,x)\in\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$, *there exists* $y\in\mathbb{R}^d_{\geqslant 0}$ *with*

$$f(t,x)=\psi(x+y)-t\mathsf{H}^*\Big(\frac{y}{t}\Big). \tag{2.158}$$

Proof. Fix $\lambda > 0$ and $y \in \mathbb{R}^d_{\geqslant 0}$, and observe that

$$\mathsf{H}^*(y) = \sup_{p \in \mathbb{R}^d_{\geqslant 0}} \big(p \cdot y - \mathsf{H}(p)\big) \geqslant \lambda |y| - \sup_{|z| \leqslant \lambda} |\mathsf{H}(z)|,$$

where the inequality is obtained by taking $p := \lambda \frac{y}{|y|}$. Since H is locally Lipschitz continuous, and therefore locally bounded, the supremum on the right side of this expression is finite. Dividing by $|y|$, letting λ tend to infinity, and then letting $|y|$ tend to infinity reveals that

$$\liminf_{\substack{|y| \to +\infty \\ y \in \mathbb{R}^d_{\geqslant 0}}} \frac{\mathsf{H}^*(y)}{|y|} = +\infty. \tag{2.159}$$

This confirms that $t\mathsf{H}^*(\frac{y}{t})$ should be interpreted as $+\infty$ when $t = 0$ and $y \neq 0$. It also implies that, given $(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$, there exists $R > 0$ large enough such that for all $y \in \mathbb{R}^d_{\geqslant 0}$ with $|y| > tR$,

$$t\mathsf{H}^*\Big(\frac{y}{t}\Big) \geqslant (\|\psi\|_{\mathrm{Lip}} + 1)|y|.$$

It follows by the Lipschitz continuity of ψ that for all $y \in \mathbb{R}^d_{\geqslant 0}$ with $|y| > tR$,

$$\psi(x+y) - t\mathsf{H}^*\Big(\frac{y}{t}\Big) \leqslant \psi(x) - |y|.$$

This means that the supremum defining the Hopf-Lax function in (2.97) may be restricted to a bounded set. Together with the fact that the function $y \mapsto \psi(x+y) - t\mathsf{H}^*(\frac{y}{t})$ is upper semi-continuous and locally bounded from above, as $\mathsf{H}^*(z) \geqslant -\mathsf{H}(0)$, this implies that the supremum on the right side of (2.97) is achieved at some point $y \in \mathbb{R}^d_{\geqslant 0}$. This completes the proof. ∎

Lemma 2.25 (Semigroup property)**.** *Under the assumptions of Proposition 2.10, for every pair $t > s \geqslant 0$ and $x \in \mathbb{R}^d_{\geqslant 0}$,*

$$f(t,x) = \sup_{y \in \mathbb{R}^d_{\geqslant 0}} \bigg(f(s,x+y) - (t-s)\mathsf{H}^*\Big(\frac{y}{t-s}\Big)\bigg). \tag{2.160}$$

Proof. Fix $y,z \in \mathbb{R}^d_{\geqslant 0}$. Since H^* is convex,

$$\mathsf{H}^*\Big(\frac{y+z}{t}\Big) \leqslant \frac{s}{t}\mathsf{H}^*\Big(\frac{y}{s}\Big) + \frac{t-s}{t}\mathsf{H}^*\Big(\frac{z}{t-s}\Big).$$

Taking $y+z \in \mathbb{R}^d_{\geqslant 0}$ in (2.97), and leveraging this bound yields

$$f(t,x) \geqslant \psi(x+y+z) - s\mathsf{H}^*\Big(\frac{y}{s}\Big) - (t-s)\mathsf{H}^*\Big(\frac{z}{t-s}\Big).$$

Taking the supremum over all $y \in \mathbb{R}^d_{\geqslant 0}$ gives

$$f(t,x) \geqslant f(s,x+z) - (t-s)\mathsf{H}^*\Big(\frac{z}{t-s}\Big),$$

and taking the supremum over all $z \in \mathbb{R}^d_{\geqslant 0}$ establishes the lower bound

$$f(t,x) \geqslant \sup_{y \in \mathbb{R}^d_{\geqslant 0}} \left(f(s,x+y) - (t-s)\mathsf{H}^*\left(\frac{y}{t-s}\right)\right).$$

To obtain the matching upper bound, invoke Lemma 2.24 to find $y \in \mathbb{R}^d_{\geqslant 0}$ with

$$f(t,x) = \psi(x+y) - t\mathsf{H}^*\left(\frac{y}{t}\right).$$

Defining $z := \frac{t-s}{t}y$, observe that

$$\frac{z}{t-s} = \frac{y-z}{s} = \frac{y}{t}.$$

In particular, taking $y - z \in \mathbb{R}^d_{\geqslant 0}$ in (2.97) gives

$$\begin{aligned} f(s,x+z) - (t-s)\mathsf{H}^*\left(\frac{z}{t-s}\right) &\geqslant \psi(x+y) - s\mathsf{H}^*\left(\frac{y-z}{s}\right) - (t-s)\mathsf{H}^*\left(\frac{z}{t-s}\right) \\ &= \psi(x+y) - t\mathsf{H}^*\left(\frac{y}{t}\right) \\ &= f(t,x). \end{aligned}$$

Taking the supremum over $z \in \mathbb{R}^d_{\geqslant 0}$ establishes the matching upper bound and completes the proof. ■

Lemma 2.26. *Under the assumptions of Proposition 2.10, the Hopf-Lax function f belongs to the solution space $\mathfrak{L}_{\mathrm{unif}}$ with*

$$\sup_{t \geqslant 0} \|f(t,\cdot)\|_{\mathrm{Lip}} = \|\psi\|_{\mathrm{Lip}}. \tag{2.161}$$

Proof. Fix $(t,x,x') \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$, and invoke Lemma 2.24 to find $y \in \mathbb{R}^d_{\geqslant 0}$ with

$$f(t,x) = \psi(x+y) - t\mathsf{H}^*\left(\frac{y}{t}\right).$$

Taking $y \in \mathbb{R}^d_{\geqslant 0}$ in (2.97) gives the lower bound

$$f(t,x') \geqslant \psi(x'+y) - t\mathsf{H}^*\left(\frac{y}{t}\right).$$

It follows that

$$f(t,x) - f(t,x') \leqslant \psi(x+y) - \psi(x'+y) \leqslant \|\psi\|_{\mathrm{Lip}}|x-x'|.$$

Reversing the roles of x and x' gives $y' \in \mathbb{R}^d$ with

$$f(t,x') - f(t,x) \leqslant \psi(x'+y') - \psi(x+y') \leqslant \|\psi\|_{\mathrm{Lip}}|x-x'|.$$

Combining these two bounds establishes the spatial Lipschitz continuity (2.161) of the Hopf-Lax function. To conclude that $f \in \mathfrak{L}_{\mathrm{unif}}$, there remains to show that f is also continuous in time; it is now shown that it is, in fact, Lipschitz continuous in time as well. Fix $x \in \mathbb{R}^d_{\geqslant 0}$ and $t > s \geqslant 0$. The semigroup property in Lemma 2.25 with $y = 0$ and the non-decreasingness of the non-linearity H imply that

$$f(t,x) \geqslant f(s,x) - (t-s)\mathsf{H}^*(0) \geqslant f(s,x) - (t-s)|\mathsf{H}(0)|. \tag{2.162}$$

Combining the semigroup property in Lemma 2.25 with the spatial Lipschitz continuity (2.161) reveals that

$$\begin{aligned} f(t,x) &\leqslant f(s,x) + \sup_{y\in\mathbb{R}^d_{\geqslant 0}} \left(\|\psi\|_{\mathrm{Lip}}|y| - (t-s)\mathsf{H}^*\left(\frac{y}{t-s}\right)\right) \\ &= f(s,x) + (t-s) \sup_{z\in\mathbb{R}^d_{\geqslant 0}} \left(\|\psi\|_{\mathrm{Lip}}|z| - \mathsf{H}^*(z)\right) \\ &\leqslant f(s,x) + (t-s) \sup_{|p|\leqslant\|\psi\|_{\mathrm{Lip}}} \sup_{z\in\mathbb{R}^d_{\geqslant 0}} \left(z\cdot p - \mathsf{H}^*(z)\right), \end{aligned}$$

where the final inequality uses that $\|\psi\|_{\mathrm{Lip}}|z| = z\cdot\frac{\|\psi\|_{\mathrm{Lip}}z}{|z|}$. Invoking the Fenchel-Moreau theorem on positive half-space (Proposition A.6) and remembering (2.162) establishes the temporal Lipschitz continuity of the Hopf-Lax function,

$$|f(t,x) - f(s,x)| \leqslant |t-s| \sup_{|p|\leqslant\|\psi\|_{\mathrm{Lip}}} |\mathsf{H}(p)|. \tag{2.163}$$

The convexity of the non-linearity H has played its part. This completes the proof. ■

Proof of Proposition 2.10. The proof proceeds in two steps. First, it is shown that the Hopf-Lax function (2.97) is a viscosity supersolution to the Hamilton-Jacobi equation (2.80), and then that it is also a viscosity subsolution to this equation. Together with Lemma 2.26 and the uniqueness result in Proposition 2.9, this proves that the Hopf-Lax function (2.97) is the unique viscosity solution in $\mathfrak{L}_{\mathrm{unif}}$ to the Hamilton-Jacobi equation (2.80).

Step 1: viscosity supersolution. Consider a smooth function $\phi\in C^\infty(\mathbb{R}_{>0}\times\mathbb{R}^d_{>0};\mathbb{R})$ with the property that $f-\phi$ has a local minimum at the point $(t^*,x^*)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$. By definition of a local minimum, for every $s\in(0,t^*)$ sufficiently small and $y\in\mathbb{R}^d_{\geqslant 0}$,

$$\phi(t^*,x^*) - \phi(t^*-s,x^*+sy) \geqslant f(t^*,x^*) - f(t^*-s,x^*+sy).$$

It follows by the semigroup property in Lemma 2.25 that

$$\phi(t^*,x^*) - \phi(t^*-s,x^*+sy) \geqslant s\mathsf{H}^*\left(\frac{sy}{s}\right) = s\mathsf{H}^*(y).$$

Dividing by s and letting s tend to zero reveals that

$$\partial_t\phi(t^*,x^*) - y\cdot\nabla\phi(t^*,x^*) + \mathsf{H}^*(y) \geqslant 0.$$

Taking the supremum over $y\in\mathbb{R}^d_{\geqslant 0}$ and invoking the Fenchel-Moreau theorem on positive half-space (Proposition A.6) shows that

$$\left(\partial_t\phi - \mathsf{H}(\nabla\phi)\right)(t^*,x^*) = \partial_t\phi(t^*,x^*) - \mathsf{H}^{**}\left(\nabla\phi(t^*,x^*)\right) \geqslant 0.$$

The convexity of the non-linearity H has played its part. This verifies the supersolution criterion.

Step 2: viscosity subsolution. Consider a smooth function $\phi\in C^\infty(\mathbb{R}_{>0}\times\mathbb{R}^d_{>0};\mathbb{R})$ with the property that $f-\phi$ has a local maximum at the point $(t^*,x^*)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$, and suppose for the sake of contradiction that there is

$\delta > 0$ such that for all (t',x') sufficiently close to (t^*,x^*),

$$\big(\partial_t \phi - \mathsf{H}(\nabla\phi)\big)(t',x') \geqslant \delta > 0.$$

Using the convexity of H and the Fenchel-Moreau theorem on positive half-space, this may be recast as the assumption that for all (t',x') sufficiently close to (t^*,x^*) and all $y \in \mathbb{R}^d_{\geqslant 0}$,

$$\partial_t \phi(t',x') - y\cdot\nabla\phi(t',x') + \mathsf{H}^*(y) \geqslant \delta. \tag{2.164}$$

Leveraging the semigroup property in Lemma 2.25 and arguing exactly as in the proof of Lemma 2.26, it is possible to find $R>0$ with the property that for all $s>0$ sufficiently small there is $y_s \in \mathbb{R}^d_{\geqslant 0}$ with $|y_s| \leqslant Rs$ and

$$f(t^*,x^*) = f(t^*-s,x^*+y_s) - s\mathsf{H}^*\Big(\frac{y_s}{s}\Big).$$

If $u(r) := (t^* + (r-1)s, x^* + (1-r)y_s)$, then it follows by the fundamental theorem of calculus and the absurd assumption (2.164) with $y := \frac{y_s}{s}$ that

$$\begin{aligned}
\phi(t^*,x^*) - \phi(t^*-s,x^*+y_s) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}r}\phi(u(r))\,\mathrm{d}r \\
&= \int_0^1 \big(s\partial_t\phi - y_s\cdot\nabla\phi\big)(u(r))\,\mathrm{d}r \\
&\geqslant s\delta - s\mathsf{H}^*\Big(\frac{y_s}{s}\Big) \\
&= s\delta + f(t^*,x^*) - f(t^*-s,x^*+y_s).
\end{aligned}$$

Rearranging shows that for s sufficiently small,

$$f(t^*-s,x^*+y_s) - \phi(t^*-s,x^*+y_s) \geqslant s\delta + f(t^*,x^*) - \phi(t^*,x^*).$$

This contradicts the local maximality of $f-\phi$ at (t^*,x^*) and completes the proof. ∎

As will be seen below, in the symmetric rank-one matrix estimation problem, as well as the broader class of models from statistical inference considered in [30], the free energy is always convex. If one can show that it converges to the solution to a Hamilton-Jacobi equation, then one can represent its limit variationally using the Hopf variational formula. This convexity property of the free energy is however lost in the realm of the sparse stochastic block model studied later in this thesis or of spin glasses [86]. In some cases, the non-linearity appearing in the relevant Hamilton-Jacobi equation is convex. This allows one to appeal to the Hopf-Lax formula instead and to still represent the solution variationally. However, in the assortative sparse stochastic block model, and many spin-glass models of interest such as the bipartite model, the non-linearity in the equation is neither convex nor concave. In this case, it is not yet known whether the free energy converges to the viscosity solution of the relevant Hamilton-Jacobi equation. Partial results for the sparse stochastic block model first obtained in [49] are the content of Theorems 1.6 and 1.7. Partial results for spin glasses have also been obtained in [83, 86]. In this context, it is interesting to wonder whether, for general solutions to Hamilton-Jacobi equations, any aspect of the variational structure displayed in the Hopf and Hopf-Lax formulas is preserved in the absence of any convexity or concavity assumption on the initial condition or the non-linearity. More specifically, if the solution $f(t,x)$ to the Hamilton-Jacobi equation (2.80) can always be represented as a critical value of the function $\mathcal{J}_{t,x}$; in other words, whether it is always possible to find a point (y^*, p^*) such that $\nabla \mathcal{J}_{t,x}(y^*, p^*) = 0$ and $f(t,x) = \mathcal{J}_{t,x}(y^*, p^*)$. For a detailed discussion on this topic, the interested reader is referred to Chapter 3 in [50].

2.5 Revisiting the matrix estimation Hamilton-Jacobi equation

Equipped with a well-posedness theory for general Hamilton-Jacobi equations on positive half-space, the Hamilton-Jacobi equation (2.73) derived in the context of the symmetric rank-one matrix estimation problem is now revisited. This Hamilton-Jacobi equation admits a unique viscosity solution by Proposition 2.9. To prove Theorem 2.5, the goal will be to show that the limit of the free energy (2.59) is this unique solution. Appealing to the Hopf-Lax formula in Proposition 2.10 will then give the variational representation (2.74).

To show that the limit of the free energy (2.59) is a viscosity solution to the Hamilton-Jacobi equation (2.73), the idea will be to start from (2.66) and obtain more tractable bounds on the left side of this equality.

Proposition 2.27. *The enriched free energy* (2.59) *in the symmetric rank-one matrix estimation problem satisfies the approximate Hamilton-Jacobi equation*

$$0 \leqslant \partial_t \overline{F}_N(t,h) - \big(\partial_h \overline{F}_N(t,h)\big)^2 \leqslant \frac{1}{N}\partial_h^2 \overline{F}_N(t,h) + \mathbb{E}\big(\partial_h F_N - \partial_h \overline{F}_N\big)^2(t,h). \tag{2.165}$$

Proof. The lower bound in (2.165) is immediate from (2.66) and the non-negativity of the variance. To establish the upper bound, by (2.66), it suffices to show that

$$\mathrm{Var}\bigg(\frac{x\cdot \overline{x}}{N}\bigg) \leqslant \frac{1}{N}\partial_h^2 \overline{F}_N(t,h) + \mathbb{E}\big(\partial_h F_N - \partial_h \overline{F}_N\big)^2(t,h). \tag{2.166}$$

It is reasonable to expect this variance term to be related to $\partial_h^2 \overline{F}_N(t,h)$. To compute this derivative, it will be convenient to introduce the Hamiltonian

$$H_N'(h,x) := \frac{1}{\sqrt{2h}} z\cdot x + 2x\cdot \overline{x} - |x|^2 \tag{2.167}$$

in such a way that the derivative of the free energy (2.58) prior to averaging may be expressed concisely as

$$\partial_h F_N(t,h) = \frac{1}{N}\langle H_N' \rangle.$$

Differentiating this expression reveals that

$$\partial_h^2 F_N(t,h) = \frac{1}{N}\langle \partial_h H_N'\rangle + \frac{1}{N}\big\langle (H_N')^2\big\rangle - \frac{1}{N}\langle H_N'\rangle^2 = \frac{1}{N}\big\langle (H_N')^2\big\rangle - \frac{1}{N}\langle H_N'\rangle^2 - \frac{1}{N(2h)^{3/2}}\langle z\cdot x\rangle.$$

Together with the Gaussian integration by parts formula (Theorem 4.5 in [50]) and the Nishimori identity, this implies that

$$\begin{aligned}\partial_h^2 \overline{F}_N(t,h) &= \frac{1}{N}\mathbb{E}\big\langle (H_N')^2\big\rangle - \frac{1}{N}\mathbb{E}\langle H_N'\rangle^2 - \frac{1}{2hN}\big(\mathbb{E}\big\langle |x|^2\big\rangle - \mathbb{E}\langle x\cdot \overline{x}\rangle\big)\\ &= \frac{1}{N}\mathbb{E}\big\langle (H_N')^2\big\rangle - \frac{1}{N}\mathbb{E}\langle H_N'\rangle^2 - \frac{1}{2hN}\big(\mathbb{E}\big\langle |x|^2\big\rangle - \mathbb{E}|\langle x\rangle|^2\big).\end{aligned}$$

Notice that the variance of H_N' may be written as

$$\begin{aligned}\mathrm{Var}\big(H_N'\big) &= \mathbb{E}\Big\langle\big(H_N' - \langle H_N'\rangle\big)^2\Big\rangle + \mathbb{E}\big(\langle H_N'\rangle - \mathbb{E}\langle H_N'\rangle\big)^2 \\ &= \mathbb{E}\big\langle (H_N')^2\big\rangle - \mathbb{E}\langle H_N'\rangle^2 + N^2\mathbb{E}\big(\partial_h F_N - \partial_h \overline{F}_N\big)^2(t,h).\end{aligned}$$

Up to lower-order terms, the proof consists in showing that the variance of $x\cdot\bar{x}$ is bounded from above by the variance of H_N', which in turn is essentially the right side of (2.165) up to scaling. To justify this precisely, write

$$\frac{1}{N}\partial_h^2\overline{F}_N(t,h) = \frac{1}{N^2}\mathrm{Var}\big(H_N'\big) - \mathbb{E}\big(\partial_h F_N - \partial_h\overline{F}_N\big)^2(t,h) - \frac{1}{2hN^2}\big(\mathbb{E}\big\langle |x|^2\big\rangle - \mathbb{E}|\langle x\rangle|^2\big).$$

To relate this back to the variance term in (2.166), observe that by the derivative computation (2.65),

$$\frac{1}{N^2}\mathrm{Var}\big(H_N'\big) - \frac{1}{N^2}\mathrm{Var}(x\cdot\bar{x}) = \frac{1}{N^2}\mathbb{E}\big\langle (H_N')^2\big\rangle - \frac{1}{N^2}\mathbb{E}\big\langle (x\cdot\bar{x})^2\big\rangle.$$

It follows that

$$\mathrm{Var}\bigg(\frac{x\cdot\bar{x}}{N}\bigg) \leqslant \frac{1}{N}\partial_h^2\overline{F}_N(t,h) + \mathbb{E}\big(\partial_h F_N - \partial_h\overline{F}_N\big)^2(t,h) + \frac{1}{2hN^2}\mathbb{E}\big\langle |x|^2\big\rangle - \frac{1}{N^2}\mathbb{E}\big\langle (H_N')^2\big\rangle + \frac{1}{N^2}\mathbb{E}\big\langle (x\cdot\bar{x})^2\big\rangle. \quad (2.168)$$

What is important for the sequel is to verify that the last two terms in this expression are of lower order in N due to a cancellation between $\mathbb{E}\big\langle (H_N')^2\big\rangle$ and $\mathbb{E}\big\langle (x\cdot\bar{x})^2\big\rangle$. In fact, it will be shown that these terms are non-positive. Observe that

$$\mathbb{E}\big\langle (H_N')^2\big\rangle = \frac{1}{2h}\mathbb{E}\big\langle (z\cdot x)^2\big\rangle + \frac{2}{\sqrt{2h}}\mathbb{E}\big\langle z\cdot x(2x\cdot\bar{x} - |x|^2)\big\rangle + 4\mathbb{E}\big\langle x\cdot\bar{x}(x\cdot\bar{x} - |x|^2)\big\rangle + \mathbb{E}\big\langle |x|^4\big\rangle,$$

and fix indices $1\leqslant i,j\leqslant N$. Two applications of the Gaussian integration by parts formula reveal that for $i\neq j$,

$$\mathbb{E}\langle z_i z_j x_i x_j\rangle = \sqrt{2h}\,\mathbb{E}\big\langle z_j x_j(x_i^2 - x_i x_i')\big\rangle = 2h\mathbb{E}\big\langle (x_i^2 - x_i x_i')(x_j^2 + x_j x_j' - 2x_j x_j'')\big\rangle,$$

while for $i=j$,

$$\mathbb{E}\big\langle z_i^2 x_i^2\big\rangle = \sqrt{2h}\,\mathbb{E}\big\langle z_i x_i(x_i^2 - x_i x_i')\big\rangle + \mathbb{E}\big\langle x_i^2\big\rangle = 2h\mathbb{E}\big\langle (x_i^2 - x_i x_i')(x_i^2 + x_i x_i' - 2x_i x_i'')\big\rangle + \mathbb{E}\big\langle x_i^2\big\rangle.$$

Together with the Nishimori identity, this shows that

$$\frac{1}{2h}\mathbb{E}\big\langle (z\cdot x)^2\big\rangle = \mathbb{E}\big\langle |x|^4\big\rangle - 2\mathbb{E}\big\langle |x|^2(x\cdot\bar{x})\big\rangle - \mathbb{E}\big\langle (x\cdot\bar{x})^2\big\rangle + 2\mathbb{E}\big\langle (x\cdot\bar{x})(x\cdot x')\big\rangle + \frac{1}{2h}\mathbb{E}\big\langle |x|^2\big\rangle.$$

A similar computation using the Gaussian integration by parts formula gives

$$\begin{aligned}\frac{2}{\sqrt{2h}}\mathbb{E}\big\langle z\cdot x(2x\cdot\bar{x} - |x|^2)\big\rangle &= 2\mathbb{E}\big\langle (2x\cdot\bar{x} - |x|^2)(|x|^2 - (x\cdot x'))\big\rangle \\ &= 6\mathbb{E}\big\langle |x|^2(x\cdot\bar{x})\big\rangle - 4\mathbb{E}\big\langle (x\cdot\bar{x})(x\cdot x')\big\rangle - 2\mathbb{E}\big\langle |x|^4\big\rangle.\end{aligned}$$

It follows by the Cauchy-Schwarz inequality that

$$\mathbb{E}\big\langle (H_N')^2\big\rangle = 3\mathbb{E}\big\langle (x\cdot\bar{x})^2\big\rangle - 2\mathbb{E}\big\langle (x\cdot\bar{x})(x\cdot x')\big\rangle + \frac{1}{2h}\mathbb{E}\big\langle |x|^2\big\rangle \geqslant \mathbb{E}\big\langle (x\cdot\bar{x})^2\big\rangle + \frac{1}{2h}\mathbb{E}\big\langle |x|^2\big\rangle.$$

This shows that the last two terms in (2.168) are non-positive and establishes the upper bound (2.166), thereby completing the proof. ■

To control the error term on the right side of the approximate Hamilton-Jacobi equation (2.165), at first glance, it seems like one needs to establish the concentration of the derivative of the enriched free energy (2.58) about its average. However, even in the simpler setting of the Curie-Weiss model, it is possible to construct examples in which the derivative of the free energy is not small at some special points — those at which the limit free energy is not differentiable. It should therefore not be expected that the derivative of the free energy concentrates at all points, and indeed, the concentration of this derivative will only be needed at contact points (t_N, h_N). It turns out that at these points, the variance of the derivative of the free energy can be controlled in terms of

$$\mathbb{E} \sup_{(t,h)\in[0,M]^2} |F_N(t,h) - \overline{F}_N(t,h)|^2 \tag{2.169}$$

for an adequate choice of $M > 0$. Leveraging the Gaussian concentration inequality and a covering argument, it can be shown that (2.169) is essentially of order N^{-1}, and therefore vanishes when taking the limit along the sequence $(t_N, h_N)_{N\geqslant 1}$ of contact points. The starting point is a Gaussian-type estimate on the tail of the random variable in (2.169).

Lemma 2.28. *For each $M \in \mathbb{R}_{\geqslant 0}$, there exists $C < +\infty$ such that for all $\lambda > CN^{-1/2}\sqrt{\log(N)}$,*

$$\mathbb{P}\left\{ \sup_{(t,h)\in[0,M]^2} |F_N(t,h) - \overline{F}_N(t,h)| \geq \lambda \right\} \leq \exp\left(-\frac{N\lambda^2}{C} \right). \tag{2.170}$$

Proof. Write $C < +\infty$ for a constant that does not depend on N or λ whose value may change throughout the proof, and introduce the random norm

$$\|W\|_* := \sup_{|x|\leqslant 1} |Wx|.$$

The proof proceeds in two steps. First a concentration inequality for the random norm $\|W\|_*$ is established by showing that there exists a constant $C < +\infty$ such that for every $a \geqslant C$,

$$\mathbb{P}\left\{ \|W\|_*^2 \geqslant aN \right\} \leqslant \exp\left(-\frac{aN}{C} \right), \tag{2.171}$$

and then a covering argument is leveraged to deduce (2.170).

Step 1: concentration inequality. To begin with, a pointwise upper bound on $|Wx|$ for each unit vector $x \in \mathbb{R}^N$ is established, and then a covering argument is used to extend this pointwise bound into the uniform bound required to control the random norm $\|W\|_*$. Fix $\mu > 0$ and $x \in \mathbb{R}^N$ of unit norm. Chebyshev's inequality reveals that

$$\mathbb{P}\left\{ |Wx|^2 \geqslant aN \right\} \leqslant \exp(-\mu a N)\mathbb{E}\exp\left(\mu |Wx|^2 \right). \tag{2.172}$$

The random variables $((Wx)_i)_{i\leqslant N}$ are independent and standard Gaussian. It follows that

$$\mathbb{E}\exp\left(\mu |Wx|^2 \right) = \mathbb{E}\prod_{i=1}^N \exp\left(\mu (Wx)_i^2 \right) = \left(\mathbb{E}\exp\left(\mu Z^2 \right) \right)^N$$

for a standard Gaussian random variable Z. A change of variables shows that for $\mu < 1/2$,

$$\mathbb{E}\exp(\mu Z^2) = \frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-x^2(\frac{1}{2}-\mu)}\,\mathrm{d}x = \frac{1}{\sqrt{1-2\mu}}\frac{1}{\sqrt{2\pi}}\int_{\mathbb{R}} e^{-\frac{y^2}{2}}\,\mathrm{d}y = \frac{1}{\sqrt{1-2\mu}},$$

so in fact $\mathbb{E}\exp(\mu|Wx|^2) \leqslant \exp(CN)$. Substituting this into (2.172) gives the pointwise upper bound

$$\mathbb{P}\big\{|Wx|^2 \geqslant aN\big\} \leq \exp\Big(\Big(C - \frac{a}{C}\Big)N\Big). \tag{2.173}$$

To extend this into a uniform bound, a covering argument is used. Recall that the smallest number of balls of radius $\varepsilon > 0$ required to cover a compact set K is known as the *covering number* of the set K, and it is denoted by $\mathcal{N}(K,\varepsilon)$. A related quantity is the packing number of a compact set K. A subset $\mathcal{N}$ of K is ε*-separated* if $|x-y| > \varepsilon$ for all distinct points $x,y \in \mathcal{N}$. The *packing number* of K is the size of the largest possible ε-separated subset of K, and it is denoted by $\mathcal{P}(K,\varepsilon)$. By maximality, it is clear that the balls of radius $\varepsilon > 0$ centred at the points in an ε-separated subset of K cover the set K. In particular, the covering number is bounded by the packing number,

$$\mathcal{N}(K,\varepsilon) \leqslant \mathcal{P}(K,\varepsilon).$$

This observation is now used to find an upper bound on the covering number of the unit ball $B_1(0) \subseteq \mathbb{R}^N$. Denote by m Lebesgue measure on $\mathbb{R}^N$, fix $\varepsilon > 0$ and let P be a maximal ε-separated subset of $B_1(0)$. Since $|x-y| > \varepsilon$ for all distinct points $x,y \in P$, the balls $(B_{\varepsilon/2}(x))_{x\in P}$ are disjoint and contained in the expanded unit ball $B_{1+\varepsilon/2}(0)$. It follows by the additivity and scaling properties of the Lebesgue measure that

$$\mathcal{P}(B_1(0),\varepsilon)\left(\frac{\varepsilon}{2}\right)^N m(B_1(0)) = \sum_{x\in P} m\big(B_{\varepsilon/2}(x)\big) \leqslant \left(1+\frac{\varepsilon}{2}\right)^N m(B_0(1)).$$

Rearranging reveals that

$$\mathcal{N}(B_1(0),\varepsilon) \leqslant \mathcal{P}(B_1(0),\varepsilon) \leqslant \left(1+\frac{2}{\varepsilon}\right)^N.$$

Choosing $\varepsilon := 1/2$, it is therefore possible to find a set A of size 5^N with the property that for every $x \in \mathbb{R}^N$ with $|x| \leqslant 1$, there exists $y \in A$ such that $|x-y| \leq \frac{1}{2}$. In particular,

$$|Wx| \leqslant \|W\|_*|x-y| + |Wy| \leqslant \frac{1}{2}\|W\|_* + \sup_{y\in A}|Wy|,$$

and taking the supremum in x shows that $\|W\|_* \leqslant 2\sup_{y\in A}|Wy|$. Together with the pointwise upper bound (2.173), this implies that

$$\mathbb{P}\big\{\|W\|_*^2 \geqslant aN\big\} \leqslant \mathbb{P}\big\{4\sup_{y\in A}|Wy|^2 \geqslant aN\big\} \leqslant 5^N\exp\Big(\Big(C-\frac{a}{C}\Big)N\Big).$$

Redefining the constant C establishes the concentration inequality (2.171).

Step 2: covering argument. To run a covering argument, first, the Hölder continuity of the free energy $\overline{F}_N$ is established. Fix $t,t',h,h' \in [0,M]$ as well as x in the support of P_N, and observe that by the Cauchy-Schwarz inequality and the boundedness of the support of P_1,

$$\begin{aligned}|H_N(t,h,x)-H_N(t',h',x)| &\leqslant \left|\sqrt{\frac{2t}{N}}-\sqrt{\frac{2t'}{N}}\right||x\cdot Wx|+\frac{1}{N}|t-t'|\big(2(x\cdot\bar{x})^2+|x|^4\big)\\ &\qquad+\big|\sqrt{2h}-\sqrt{2h'}\big||x\cdot z|+|h-h'|\big(2|x\cdot\bar{x}|+|x|^2\big)\\ &\leqslant C\sqrt{N}\Big(\big|\sqrt{t}-\sqrt{t'}\big|\|W\|_*+\big|\sqrt{h}-\sqrt{h'}\big||z|\Big)+CN\big(|t-t'|+|h-h'|\big).\end{aligned}$$

Together with the observation that $\sqrt{y}-\sqrt{y'}\leqslant|y-y'|^{1/2}$ and $|y-y'|\leqslant 2|y||y-y'|^{1/2}$ for $y,y'\in\mathbb{R}_{\geqslant 0}$ with $y\geqslant y'$, this bound on the Hamiltonian implies that the free energy (2.58) is Hölder continuous on $[0,M]^2$ with

$$|F_N(t,h)-F_N(t',h')|\leqslant C\big(|t-t'|^{1/2}+|h-h'|^{1/2}\big)X$$

for the random variable

$$X:=1+\frac{\|W\|_*}{\sqrt{N}}+\frac{|z|}{\sqrt{N}}.$$

Averaging this inequality also shows that the free energy (2.59) is Hölder continuous with

$$|\overline{F}_N(t,h)-\overline{F}_N(t',h')|\leqslant C\big(|t-t'|^{1/2}+|h-h'|^{1/2}\big)\mathbb{E}X.$$

These two bounds imply that for any $\lambda>0$ and $\varepsilon>0$,

$$\mathbb{P}\Bigg\{\sup_{[0,M]^2}|F_N(t,h)-\overline{F}_N(t,h)|\geqslant\lambda\Bigg\}\leqslant\mathbb{P}\Bigg\{\sup_{A_\varepsilon}|F_N(t,h)-\overline{F}_N(t,h)|\geqslant\lambda/2\Bigg\}+\mathbb{P}\big\{X\geqslant\varepsilon^{-1/2}\lambda/C\big\}$$

for the set $A_\varepsilon:=\varepsilon\mathbb{Z}^2\cap[0,M]^2$. Indeed, every point $(t,h)\in[0,M]^2$ is at distance at most ε from a point in A_ε. A union bound and the fact that the cardinality of A_ε is bounded by $C\varepsilon^{-2}$ yield that this is further bounded by

$$C\varepsilon^{-2}\sup_{A_\varepsilon}\mathbb{P}\Bigg\{|F_N(t,h)-\overline{F}_N(t,h)|\geqslant\lambda/2\Bigg\}+\mathbb{P}\big\{X\geqslant\varepsilon^{-1/2}\lambda/C\big\}.$$

A simple extension of the proof of the free energy concentration inequality (2.44) reveals that for any $\lambda>0$ and $(t,h)\in[0,M]^2$,

$$\mathbb{P}\big\{|F_N(t,h)-\overline{F}_N(t,h)|>\lambda\big\}\leqslant 2\exp\left(-\frac{N\lambda^2}{C}\right).$$

On the other hand, the Gaussian concentration inequality (Theorem 4.7 in [50]) and the concentration inequality (2.171) imply that for some constant $C'<+\infty$ and any $\varepsilon>0$ and $\lambda>C'\sqrt{\varepsilon}$,

$$\mathbb{P}\{X\geqslant\varepsilon^{-1/2}\lambda/C\}\leqslant\exp\left(-\frac{N\lambda^2}{\varepsilon C}\right).$$

Combining these two bounds and choosing $\varepsilon:=N^{-1}$ reveals that for any $\lambda>C'N^{-1/2}$,

$$\mathbb{P}\Bigg\{\sup_{[0,M]^2}|F_N(t,h)-\overline{F}_N(t,h)|\geqslant\lambda\Bigg\}\leqslant CN^2\exp\left(-\frac{N\lambda^2}{C}\right).$$

Whenever $\lambda>C'N^{-1/2}\sqrt{\log(N)}$ for some sufficiently large constant $C'<+\infty$, the term N^2 can be absorbed into the exponential to obtain (2.170) and complete the proof. ■

It is now possible to verify that any subsequential limit of the enriched free energy (2.59) satisfies the subsolution criterion in the Hamilton-Jacobi equation (2.73).

Lemma 2.29. *Let f be any subsequential limit of the sequence $(\overline{F}_N)_{N\geqslant 1}$ of enriched free energies in the symmetric rank-one matrix estimation problem, and fix $t^*,h^*>0$. If $\phi\in C^\infty(\mathbb{R}_{>0}\times\mathbb{R}_{>0};\mathbb{R})$ is a smooth function with the property that $f-\phi$ has a strict local maximum at $(t^*,h^*)\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}$, then*

$$\big(\partial_t\phi-\overline{\mathsf{H}}(\partial_h\phi)\big)(t^*,h^*)=0 \tag{2.174}$$

Proof. Through a slight abuse of notation, the subsequence along which the convergence of $\overline{F}_N$ to f occurs is not denoted explicitly, in effect pretending that the convergence occurs along the whole sequence. Using Lemma 2.7, find a sequence $(t_N,h_N)_{N\geqslant 1}$ converging to (t^*,h^*) such that $\overline{F}_N-\phi$ has a local maximum at (t_N,h_N); this lemma also guarantees that the neighbourhood over which (t_N,h_N) is a local maximum can be chosen independently of N. To control the right side of the approximate Hamilton-Jacobi equation (2.165) at each contact point (t_N,h_N), essentially, the idea will be to argue that whenever a sequence of convex functions converges to a point of differentiability of its limiting function, the sequence of derivatives also converges to the derivative of the limiting function at this point. Although it is not yet known that the free energy is convex, it will be possible to prove upper and lower bounds on its Hessian. The proof proceeds in three steps. First the upper deviation of $\overline{F}_N$ from its tangent at (t_N,h_N) is controlled by a parabola, then the lower deviation of F_N from its tangent at (t_N,h_N) is controlled by a random parabola, and finally these two ingredients are combined to control the right side of the approximate Hamilton-Jacobi equation (2.165).

Step 1: Hessian of $\overline{F}_N$ upper bound. Since $\overline{F}_N-\phi$ has a local maximum at (t_N,h_N) and ϕ is smooth, there exists $C<+\infty$ such that for all $h'\in\mathbb{R}$ with $|h'|\leqslant C^{-1}$,

$$\overline{F}_N(t_N,h_N+h')-\overline{F}_N(t_N,h_N)\leqslant\phi(t_N,h_N+h')-\phi(t_N,h_N)\leqslant h'\partial_h\phi(t_N,h_N)+C|h'|^2.$$

For every N sufficiently large, one has $h_N>0$, and therefore

$$\partial_h(\overline{F}_N-\phi)(t_N,h_N)=0. \tag{2.175}$$

It follows that for every $h'\in\mathbb{R}$ with $|h'|\leqslant C^{-1}$,

$$\overline{F}_N(t_N,h_N+h')-\overline{F}_N(t_N,h_N)\leq h'\partial_h\overline{F}_N(t_N,h_N)+C|h'|^2. \tag{2.176}$$

In particular,

$$\partial_h^2\overline{F}_N(t_N,h_N)\leqslant C. \tag{2.177}$$

Throughout this proof, it is understood that the constant $C<+\infty$ may change from one occurrence to the next, only making sure that it does not depend on N.

Step 2: Hessian of F_N lower bound. Recalling the definition of the Hamiltonian H_N' in (2.167), a direct computation reveals that

$$\partial_h^2F_N(t_N,h_N+h')=\frac{1}{N}\big\langle(H_N')^2\big\rangle-\frac{1}{N}\langle H_N'\rangle^2-\frac{1}{N(2(h_N+h'))^{3/2}}\langle z\cdot x\rangle.$$

Since h_N converges to $h^*>0$, it remains bounded away from zero for N sufficiently large. Using also the non-negativity of the variance, the Cauchy-Schwarz inequality, and the fact that the measure P_1 has bounded support, it is possible to deduce that for every $|h'| \leq C^{-1}$,

$$\partial_h^2 F_N(t_N, h_N + h') \geqslant -\frac{C|z|}{\sqrt{N}}.$$

It follows by Taylor's theorem that for every $|h'| \leq C^{-1}$,

$$F_N(t_N, h_N + h') - F_N(t_N, h_N) - h'\partial_h F_N(t_N, h_N) \geqslant -\frac{C|z|}{\sqrt{N}}|h'|^2 \tag{2.178}$$

which is the sought after control of the lower deviation of F_N from its tangent at (t_N, h_N) by a random parabola.

Step 3: controlling the right side of (2.165). Combining (2.176) and (2.178) with the fact that $(t_N, h_N)_{N\geqslant 1}$ converges to (t^*, h^*) gives a neighborhood V of (t^*, h^*) with

$$h'\big(\partial_h F_N - \partial_h \overline{F}_N\big)(t_N, h_N) \leqslant 2\sup_V |F_N - \overline{F}_N| + C|h'|^2\left(1 + \frac{|z|}{\sqrt{N}}\right). \tag{2.179}$$

In particular, given a deterministic $\lambda \in [0, C^{-1}]$, the bound (2.179) for

$$h' := \lambda \frac{\partial_h F_N - \partial_h \overline{F}_N}{|\partial_h F_N - \partial_h \overline{F}_N|}(t_N, h_N)$$

reads

$$\lambda|\partial_h F_N - \partial_h \overline{F}_N|(t_N, h_N) \leqslant 2\sup_V |F_N - \overline{F}_N| + C\lambda^2\left(1 + \frac{|z|}{\sqrt{N}}\right).$$

Squaring both sides of this inequality, taking expectations, and leveraging the concentration inequality in Lemma 2.28 yields

$$\lambda^2 \mathbb{E}\big(\partial_h F_N - \partial_h \overline{F}_N\big)^2(t_N, h_N) \leqslant 8\mathbb{E}\sup_V \big(F_N - \overline{F}_N\big)^2 + C\lambda^4 \mathbb{E}\left(1 + \frac{|z|}{\sqrt{N}}\right)^2 \leqslant \frac{C}{N^{1/2}} + C\lambda^4,$$

where the fact that $\mathbb{E}|z|^2 = N\mathbb{E}z_1^2 = N$ has also been used. Choosing $\lambda := N^{-1/8}$ leads to the upper bound

$$\mathbb{E}\big(\partial_h F_N - \partial_h \overline{F}_N\big)^2(t_N, h_N) \leqslant \frac{C}{N^{1/4}}. \tag{2.180}$$

Substituting this into the approximate Hamilton-Jacobi equation in Proposition 2.27 and remembering the Hessian bound (2.177) gives

$$0 \leqslant \big(\partial_t \phi - (\partial_h \phi)^2\big)(t_N, h_N) = \big(\partial_t \overline{F}_N - (\partial_h \overline{F}_N)^2\big)(t_N, h_N) \leqslant \frac{C}{N} + \frac{C}{N^{1/4}}.$$

Letting N tend to infinity and using the smoothness of ϕ reveals that

$$\big(\partial_t \phi - (\partial_h \phi)^2\big)(t^*, h^*) = 0.$$

Combining (2.175) and (2.65) with the smoothness of ϕ shows that $\partial_h \phi(t^*, h^*) \geqslant 0$. Remembering the definition of the non-linearity $\overline{\mathsf{H}}$ in (2.71) completes the proof. ■

That any subsequential limit of the enriched free energy (2.59) satisfies the supersolution criterion in the Hamilton-Jacobi equation (2.73) is much easier to verify by simply leveraging the lower bound in Proposition 2.27. Together with the Hopf-Lax formula, this gives a first proof of Theorem 2.5.

Proof of Theorem 2.5. The Arzelà-Ascoli theorem, the derivative computations (2.61) and (2.65), and the boundedness of the support of P_1 imply that the sequence $(\overline{F}_N)_{N\geqslant 1}$ is precompact. Denote by f any subsequential limit of this sequence of enriched free energies. The subsequence along which the convergence of $\overline{F}_N$ to f occurs is not denoted explicitly, in effect pretending that the convergence occurs along the whole sequence. Lemma 2.29 shows that f is a viscosity subsolution to the Hamilton-Jacobi equation (2.73). To verify the supersolution criterion, let $\phi \in C^\infty(\mathbb{R}_{>0}\times\mathbb{R}_{>0};\mathbb{R})$ be a smooth function with the property that $f-\phi$ has a strict local minimum at $(t^*,h^*)\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}$. Using Lemma 2.7, find a sequence $(t_N,h_N)_{N\geqslant 1}$ converging to (t^*,h^*) such that $\overline{F}_N-\phi$ has a local minimum at (t_N,h_N). For N large enough, one has $(t_N,h_N)\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}$, and therefore

$$\partial_t\overline{F}_N(t_N,h_N)=\partial_t\phi(t_N,h_N)\quad\text{and}\quad\partial_h\overline{F}_N(t_N,h_N)=\partial_h\phi(t_N,h_N).$$

It follows by the lower bound in Proposition 2.27 and the definition of the non-linearity $\overline{\mathsf{H}}$ in (2.71) that

$$0\leqslant\big(\partial_t\phi-(\partial_h\phi)^2\big)(t_N,h_N)=\big(\partial_t\phi-\overline{\mathsf{H}}(\partial_h\phi)\big)(t_N,h_N).$$

Letting N tend to infinity, and leveraging the continuity of $\overline{\mathsf{H}}$ and the smoothness of ϕ shows that f is a viscosity supersolution to the Hamilton-Jacobi equation (2.73). Together with the uniqueness result in Proposition 2.9, this implies that the sequence $(\overline{F}_N)_{N\geqslant 1}$ converges to the unique viscosity solution f to the Hamilton-Jacobi equation (2.73). The Hopf-Lax formula (Proposition 2.10), the definition of the non-linearity $\overline{\mathsf{H}}$ in (2.71), and the fact that the square function $x\mapsto\frac{1}{2}|x|^2$ is its own convex dual (Exercise 2.8 in [50]) imply that for all $t,h\geqslant 0$,

$$f(t,h)=\sup_{h'\in\mathbb{R}_{\geqslant 0}}\left(\psi(h')-\frac{(h'-h)^2}{4t}\right).$$

Remembering that the free energy (2.20) can be recovered from the enriched free energy (2.59) by evaluating at $h=0$, that is, $\overline{F}_N^\circ(t)=\overline{F}_N(t,0)$, gives the representation (2.75) and completes the proof. ∎

As was already discussed, together with Propositions 2.1 and 2.3, this result implies Theorem 1.3 on the limit of the mutual information in the dense stochastic block model. For more general models, such as those discussed in [36], or the sparse stochastic block model discussed in later chapters of this thesis, the bound in Proposition 2.27 is of the form

$$\left|\partial_t\overline{F}_N(t,h)-\big(\partial_h\overline{F}_N(t,h)\big)^2\right|\leqslant\frac{1}{N}\partial_h^2\overline{F}_N(t,h)+\mathbb{E}\big(\partial_hF_N-\partial_h\overline{F}_N\big)^2(t,h).\tag{2.181}$$

In such cases, the subsolution criterion can be established as in Lemma 2.29 but the inequality is too weak to prove the supersolution criterion. Indeed, the gradient of the free energy cannot be shown to concentrate as in (2.180) when the free energy is touched from below by a smooth function. The subsolution criterion can be combined with the comparison principle in Proposition 2.9 to show that any subsequential limit of the free energy is bounded from above by the solution to the Hamilton-Jacobi equation (2.73). This is the idea behind the proof of the lower bound of Conjecture 1.4 in Theorem 1.6 — recall that the mutual information is minus the free energy up to an additive constant, so an upper bound on the free energy corresponds to a lower bound on the mutual information. In the simpler setting of [36], the matching lower bound on the free energy is obtained through a selection principle for the Hamilton-Jacobi equation (2.80) that ensures that a convex function that satisfies the equation on a dense set must in fact satisfy the equation everywhere. Unfortunately, the convexity assumptions required to appeal to this selection criterion do not hold in the setting of the sparse stochastic block model [67] — this is why the upper bound in Conjecture 1.4 remains open.

2.6 Leveraging convexity to identify viscosity solutions

In this section a tool known as *the convex selection principle* is introduced to identify when a *convex* function is a viscosity solution to the Hamilton-Jacobi equation (2.80). In Example 2.6, a function that satisfies a Hamilton-Jacobi equation almost everywhere but is not a viscosity solution to this equation was described. Notice that this counterexample had corner singularities "in both directions"; formally, the second derivative was neither bounded from above nor from below. A convex function cannot look like this, since its Hessian must be non-negative. Roughly speaking, the convex selection principle states that imposing the function to be convex completely rules out the emergence of non-viscosity-type singularities, and thus restores uniqueness. This result will be established in the general setting of the Hamilton-Jacobi equation (2.80).

Proposition 2.30 (Convex selection principle)**.** *If* $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ *is a locally Lipschitz and non-decreasing non-linearity, and* $f:\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ *is a jointly convex and jointly Lipschitz continuous function with* $f(0,\cdot)\in C^1(\mathbb{R}^d_{\geqslant 0};\mathbb{R})$ *that satisfies the Hamilton-Jacobi equation* (2.80) *on a dense subset of* $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$*, then* f *is a viscosity solution to the Hamilton-Jacobi equation* (2.80)*.*

The statement that the function f satisfies the Hamilton-Jacobi equation (2.80) on a dense subset of $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$ means that the set

$$\Big\{(t,x)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{>0}\mid f \text{ is differentiable at } (t,x) \text{ and } \big(\partial_t f-\mathsf{H}(\nabla f)\big)(t,x)=0\Big\} \tag{2.182}$$

is dense in $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$. Naturally, Proposition 2.30 also holds if "convex" is replaced by "concave" in its statement. It was already stressed that, in some sense, the notion of viscosity solution is tailored to approximations in which a small *positive* term times the Laplacian of f appears on the right side of (2.80). In particular, the notion of viscosity solution is sensitive to the orientation of time; in general, it is *not* the case that the time-reversed viscosity solution to some equation will be the viscosity solution to the time-reversed equation. However, superficially, the statement of Proposition 2.30 looks invariant under time reversal. The only assumption that breaks this symmetry is that $f(0,\cdot)\in C^1(\mathbb{R}^d_{\geqslant 0};\mathbb{R})$. This already hints at the fact that this assumption is necessary. In other words, Proposition 2.30 states that under the convexity assumption, pathological solutions cannot spontaneously emerge when starting from a smooth initial condition. However, when starting from a Lipschitz function that does not have this form, it may be possible to exploit the singularities of the initial condition to create solutions that differ from the viscosity solution.

To prove the convex selection principle, first, it will be shown that the function f in its statement must actually satisfy (2.80) at all its points of differentiability.

Lemma 2.31. *If* $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ *is a locally Lipschitz non-linearity, and* $f:\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ *is a jointly convex and jointly Lipschitz continuous function that satisfies the Hamilton-Jacobi equation* (2.80) *on a dense subset of* $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$*, then it satisfies the Hamilton-Jacobi equation* (2.80) *at all its points of differentiability in* $\mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$*. Moreover, for every* $(t,x)\in\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$*, there exists* $(b,q)\in\partial f(t,x)$ *such that* $b-\mathsf{H}(q)=0$*.*

Proof. Fix $(t,x) \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$, and let $(t_n,x_n)_{n\geqslant 1} \subseteq \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$ be a sequence of points of differentiability of f converging to (t,x) at which

$$\big(\partial_t f - \mathsf{H}(\nabla f)\big)(t_n,x_n) = 0. \tag{2.183}$$

Since f is differentiable at the interior point $(t_n,x_n) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, Proposition A.9 implies that $\partial f(t_n,x_n) = \{(\partial_t f(t_n,x_n), \nabla f(t_n,x_n))\}$. The joint Lipschitz continuity of f implies that, up to the extraction of a subsequence, the sequence $(\partial_t f(t_n,x_n), \nabla f(t_n,x_n))_{n\geqslant 1}$ of gradients converges to some point $(b,q) \in \mathbb{R} \times \mathbb{R}^d$. By Proposition A.10, it must be that $(b,q) \in \partial f(t,x)$, and by (2.183) and the continuity of H, the pair $(b,q) \in \partial f(t,x)$ is such that $b - \mathsf{H}(q) = 0$. This establishes the second part of the statement. If $(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$ and f is differentiable at (t,x), then Proposition A.9 implies that $\partial f(t,x) = \{(\partial_t f(t,x), \nabla f(t,x))\}$. It must therefore be the case that $(b,q) = (\partial_t f(t,x), \nabla f(t,x))$, and thus that f satisfies the Hamilton-Jacobi equation (2.80) at the point (t,x). This completes the proof. ■

Proof of Proposition 2.30. The proof is decomposed into two steps. First it is shown that f is a viscosity subsolution to the Hamilton-Jacobi equation (2.80), and then that it is a supersolution to this equation. The assumption on the initial condition will only play a part in showing that f is a viscosity supersolution to (2.80).

Step 1: viscosity subsolution. Consider a smooth function $\phi \in C^\infty(\mathbb{R}_{>0} \times \mathbb{R}^d_{>0}; \mathbb{R})$ with the property that $f - \phi$ has a local maximum at the point $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$. It will be shown that f is differentiable at the contact point (t^*,x^*). By Proposition A.8, the subdifferential $\partial f(t^*,x^*)$ contains at least one element, say (a,p). By the definition of the subdifferential and of a local maximum, for every (t',x') sufficiently close to (t^*,x^*),

$$a(t'-t^*) + p\cdot(x'-x^*) \leqslant f(t',x') - f(t^*,x^*) \leqslant \phi(t',x') - \phi(t^*,x^*). \tag{2.184}$$

It follows by smoothness of ϕ that, as (t',x') tends to (t^*,x^*),

$$(t'-t^*)\big(a - \partial_t\phi(t^*,x^*)\big) + (x'-x^*)\big(p - \nabla\phi(t^*,x^*)\big) = o\big(|t'-t^*| + |x'-x^*|\big).$$

This implies that $(a,p) = (\partial_t\phi, \nabla\phi)(t^*,x^*)$. Using (2.184) once more shows that f is differentiable at (t^*,x^*), and that $(\partial_t f, \nabla f)(t^*,x^*) = (\partial_t\phi, \nabla\phi)(t^*,x^*)$. It follows by Lemma 2.31 that

$$\big(\partial_t\phi - \mathsf{H}(\nabla\phi)\big)(t^*,x^*) = \big(\partial_t f - \mathsf{H}(\nabla f)\big)(t^*,x^*) = 0.$$

This completes the verification that f is a viscosity subsolution to the Hamilton-Jacobi equation (2.80).

Step 2: viscosity supersolution. Consider a smooth function $\phi \in C^\infty(\mathbb{R}_{>0} \times \mathbb{R}^d_{>0}; \mathbb{R})$ with the property that $f - \phi$ has a local minimum at the point $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$. Together with the convexity of f, this implies that for every $(t',x') \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$ and $\varepsilon > 0$ small enough,

$$\begin{aligned} f(t',x') - f(t^*,x^*) &\geq \varepsilon^{-1}\Big(f\big((t^*,x^*) + \varepsilon(t'-t^*, x'-x^*)\big) - f(t^*,x^*)\Big) \\ &\geq \varepsilon^{-1}\Big(\phi\big((t^*,x^*) + \varepsilon(t'-t^*, x'-x^*)\big) - \phi(t^*,x^*)\Big). \end{aligned}$$

Letting ε tend to zero shows that $(\partial_t\phi, \nabla\phi)(t^*,x^*) \in \partial f(t^*,x^*)$. It therefore suffices to fix $(a,p) \in \partial f(t^*,x^*)$ and prove that

$$a - \mathsf{H}(p) \geqslant 0. \tag{2.185}$$

Since $(a,p) \in \partial f(t^*,x^*)$, the definition of the subdifferential implies that $f(0,y) \geqslant f(t^*,x^*) - at^* + (y - x^*)\cdot p$

for every $y \in \mathbb{R}^d_{\geqslant 0}$. Rearranging shows that for every $y \in \mathbb{R}^d_{\geqslant 0}$,

$$f(0,y) - y \cdot p \geqslant f(t^*, x^*) - at^* - p \cdot x^*. \tag{2.186}$$

Inspired by the Hopf formula, it is tempting to consider an optimizer of the minimization problem

$$\inf_{y \in \mathbb{R}^d_{\geqslant 0}} \big(f(0,y) - y \cdot p \big). \tag{2.187}$$

Since the existence of such an optimizer cannot be guaranteed, a small parameter $\varepsilon > 0$ is introduced, and the perturbed minimization problem

$$\inf_{y \in \mathbb{R}^d_{\geqslant 0}} \Big(f(0,y) - y \cdot p + \varepsilon \sqrt{1 + |y|^2} \Big) \tag{2.188}$$

is considered instead. From (2.186), it is apparent that the infimum in (2.188) may be restricted to values of y that range in a compact set (which depends on ε). Together with the continuity of the functions involved, this shows that the infimum in (2.188) is achieved, say at $y_\varepsilon \in \mathbb{R}^d_{\geqslant 0}$. Observe that

$$\lim_{\varepsilon \to 0} \inf_{y \in \mathbb{R}^d_{\geqslant 0}} \Big(f(0,y) - y \cdot p + \varepsilon \sqrt{1 + |y|^2} \Big) = \inf_{y \in \mathbb{R}^d_{\geqslant 0}} \big(f(0,y) - y \cdot p \big). \tag{2.189}$$

Indeed, the existence of the limit on the left side of (2.189) and the fact that it is lower bounded by the right side are immediate. Conversely, for each $\delta > 0$, it is possible to find $y^*_\delta \in \mathbb{R}^d_{\geqslant 0}$ such that

$$f(0, y^*_\delta) - y^*_\delta \cdot p \leqslant \inf_{y \in \mathbb{R}^d_{\geqslant 0}} \big(f(0,y) - y \cdot p \big) + \delta,$$

and the upper bound in (2.189) is obtained up to an error of $\delta > 0$ by using y^*_δ as a candidate in the infimum on the left side of (2.189). Letting $\delta > 0$ tend to zero shows shows (2.189). Since

$$\inf_{y \in \mathbb{R}^d_{\geqslant 0}} \big(f(0,y) - y \cdot p \big) + \varepsilon \sqrt{1 + |y_\varepsilon|^2} \big) \leq f(0, y_\varepsilon) - y_\varepsilon \cdot p + \varepsilon \sqrt{1 + |y_\varepsilon|^2} = \inf_{y \in \mathbb{R}^d_{\geqslant 0}} \big(f(0,y) - y \cdot p + \varepsilon \sqrt{1 + |y|^2} \big),$$

the equality (2.189) implies that

$$\lim_{\varepsilon \to 0} \varepsilon |y_\varepsilon| = 0. \tag{2.190}$$

The optimality of y_ε also imposes that for any coordinate $i \in \{1, \ldots, d\}$,

$$\begin{cases} \partial_{x_i} \psi(y_\varepsilon) + \frac{\varepsilon (y_\varepsilon)_i}{\sqrt{1 + |y_\varepsilon|^2}} = p_i & \text{if } (y_\varepsilon)_i > 0 \\ \partial_{x_i} \psi(y_\varepsilon) + \frac{\varepsilon (y_\varepsilon)_i}{\sqrt{1 + |y_\varepsilon|^2}} \geqslant p_i & \text{if } (y_\varepsilon)_i = 0 \end{cases} \tag{2.191}$$

At this point, let $(y_{\varepsilon,n})_{n \geqslant 1} \subseteq \mathbb{R}^d_{>0}$ be a sequence converging to y_ε. For each $n \geqslant 1$, invoke Lemma 2.31 to find $(b_{\varepsilon,n}, q_{\varepsilon,n}) \in \partial f(0, y_{\varepsilon,n})$ with $b_{\varepsilon,n} - \mathsf{H}(q_{\varepsilon,n}) = 0$. The assumption that $\psi \in C^1(\mathbb{R}^d_{\geqslant 0}; \mathbb{R})$, the observation that $q_{\varepsilon,n} \in \mathbb{R}^d_{>0} \cap \partial \psi(y_{\varepsilon,n})$ for each $n \geqslant 1$, and Proposition A.9 reveal that

$$(b_{\varepsilon,n}, q_{\varepsilon,n}) = \big(\mathsf{H}(\nabla \psi(y_{\varepsilon,n})), \nabla \psi(y_{\varepsilon,n}) \big).$$

The joint Lipschitz continuity of f implies that, up to the extraction of a subsequence, the sequence $(b_{\varepsilon,n}, q_{\varepsilon,n})_{n \geqslant 1}$ converges to some point $(b_\varepsilon, q_\varepsilon) \in \mathbb{R} \times \mathbb{R}^d$. By Proposition A.10, it must be that $(b_\varepsilon, q_\varepsilon) \in$

$\partial f(0,y_\varepsilon)$, and by continuity of H and $\nabla\psi$,

$$(b_\varepsilon, q_\varepsilon) = \big(\mathsf{H}(\nabla\psi(y_\varepsilon)), \nabla\psi(y_\varepsilon)\big) \in \partial f(0,y_\varepsilon). \tag{2.192}$$

Together with the fact that $(a,p) \in \partial f(t^*,x^*)$, this shows that

$$f(t^*,x^*) \geqslant f(0,y_\varepsilon) + b_\varepsilon t^* + q_\varepsilon \cdot (x^* - y_\varepsilon) \quad \text{and} \quad f(0,y_\varepsilon) \geqslant f(t^*,x^*) - at^* + p\cdot(y_\varepsilon - x^*).$$

Combining these two inequalities reveals that $at^* \geqslant b_\varepsilon t^* + (q_\varepsilon - p)\cdot(x^* - y_\varepsilon)$. Using (2.190)- (2.192) and the non-decreasingness of H yields that

$$at^* \geqslant \mathsf{H}(q_\varepsilon)t^* + \bigg(\nabla\psi(y_\varepsilon) + \frac{\varepsilon y_\varepsilon}{\sqrt{1+|y_\varepsilon|^2}} - p\bigg)\cdot(x^* - y_\varepsilon) + o_\varepsilon(1) \geqslant \mathsf{H}\bigg(p - \frac{\varepsilon y_\varepsilon}{\sqrt{1+|y_\varepsilon|^2}}\bigg)t^* + o_\varepsilon(1).$$

Since H is continuous, dividing by t^* and letting ε tend to zero establishes (2.185). This completes the proof. ■

Before leveraging the convex selection principle to provide an alternative proof of Theorem 2.5, it is worth noting that the differentiability assumption $\psi \in C^1(\mathbb{R}^d_{\geqslant 0};\mathbb{R})$ in the convex selection principle cannot be dropped in general. For an explicit counterexample, see Example 3.25 in [50].

2.7 Another Hamilton-Jacobi approach to matrix estimation

Using the results of the previous section, it is now possible to give an alternative proof of Theorem 2.5 on the limit of the enriched free energy (2.59) in the symmetric rank-one matrix estimation problem. The approach discussed in this section generalizes to models where the approximate Hamilton-Jacobi equation in Proposition 2.27 is replaced by the weaker approximate Hamilton-Jacobi equation (2.181) provided that the free energy is convex. This is for instance the case for the statistical inference models studied in [36]. The idea will be to apply the convex selection principle (Proposition 2.30) to any subsequential limit f of the sequence $(\overline{F}_N)_{N\geqslant 1}$ of enriched free energies. It must therefore be verified that the enriched free energy $\overline{F}_N$ is jointly convex, Lipschitz continuous, and non-decreasing. Lemma 2.29 can then be combined with the convex selection principle to conclude that f is the unique viscosity solution to the Hamilton-Jacobi equation (2.73).

Proposition 2.32. *The enriched free energy* (2.59) *in the symmetric rank-one matrix estimation problem is jointly convex, non-decreasing, and Lipschitz continuous uniformly over N.*

The proof that the free energy $\overline{F}_N$ is convex involves a somewhat lengthy calculation. This may come as a surprise, since for simpler models such as the Curie-Weiss model, the convexity of the free energy is a consequence of the very general observation that the log-Laplace transform of a random variable is convex. This argument cannot be applied here, however. One may wonder whether this comes from the square roots acting on the parameters t and h in the definition of the problem; indeed, the function $\widetilde{F}_N : (t,h) \mapsto \overline{F}_N(t^2,h^2)$ is convex. So is the question being made uselessly complicated here? The main reason to insist on showing the convexity property is that this is a requirement for the validity of the convex selection principle. The proof of this result relies on the fact that the underlying Hamilton-Jacobi equation does not explicitly depend on the parameters t and h. In the Hamilton-Jacobi equation for $\widetilde{F}_N$, the parameters t and h are explicitly present in the equation; and the convex selection principle is actually false in this more general setting.

In the context of statistical inference, there is however a fundamental information-theoretic reason to expect that the free energy $\overline{F}_N$ defined in (2.59) is convex but not necessarily jointly convex. Denoting by $I_N(Y,\overline{x})$ the mutual information between the signal $\overline{x}$ and the observation Y, a direct computation as in Proposition 2.1 shows that the convexity of $\overline{F}_N(\cdot,0)$ is equivalent to the concavity of the mutual information $I_N(Y,\overline{x})$. Moreover, it is possible to show that when Y is observed, exactly as much information is gained about the signal $\overline{x}$ as that provided by observing, for two independent copies W^1 and W^2 of W, the quantities

$$\sqrt{\frac{t}{N}}\overline{x}\overline{x}^\top + W^1 \quad \text{and} \quad \sqrt{\frac{t}{N}}\overline{x}\overline{x}^\top + W^2. \tag{2.193}$$

Notice that compared with the definition of Y in (2.13), the variable t has been replaced by $t/2$ in each of the two quantities above. Finally, having observed the first of the two quantities in (2.193), one can verify that at most as much information is gained upon subsequently observing the second quantity in (2.193). In other words, the mutual information $I_N(Y,\overline{x})$ satisfies a sort of subadditivity property, and elementary properties of the mutual information allow one to upgrade this to the fact that $I_N(Y,\overline{x})$ is a concave function. The argument just sketched therefore leads to the conclusion that the function $\overline{F}_N(\cdot,0)$ is convex. Minor variants of this argument yield that the function $\overline{F}_N$ is convex in each of the variables separately. It may seem plausible that this argument can be generalized and lead to a conceptual information-theoretic proof that the function $\overline{F}_N$ is convex *jointly* in (t,h). However, pushing this argument to a setting with multiple variables yields instead that every entry of the Hessian of $\overline{F}_N$ is non-negative. Surprisingly, for the sparse stochastic block model, which unlike the situation considered in this chapter is not reducible to a problem with Gaussian noise, one can indeed show that the relevant free energy is *not* convex in general [67].

To sum up this informal discussion, the proof that $\overline{F}_N$ is jointly convex does have to use some of the particular structure of the class of problems considered, and thus at least some calculations do need to be made. The interested reader is referred to Proposition 3.1 of [67] for a somewhat more general view on such calculations.

Proof of Proposition 2.32. The derivative computations (2.61) and (2.65), and the boundedness of the support of P_1 imply that the first order derivatives of the enriched free energy $\overline{F}_N$ are uniformly bounded. This establishes the uniform Lipschitz continuity of $\overline{F}_N$. That $\overline{F}_N$ is non-decreasing follows from the derivative computation (2.65). To prove the convexity of $\overline{F}_N$, its Hessian is shown to be non-negative definite. Differentiating the expression (2.65) in h and recalling (2.62) reveals that

$$\begin{aligned} N\partial_h^2\overline{F}_N(t,h) &= \mathbb{E}\big\langle (x\cdot\overline{x})\partial_h H_N(t,h,x)\big\rangle - \mathbb{E}\big\langle (x\cdot\overline{x})\partial_h H_N(t,h,x')\big\rangle \\ &= \frac{1}{\sqrt{2h}}\mathbb{E}\big\langle (x\cdot\overline{x})(z\cdot x - z\cdot x')\big\rangle + 2\mathbb{E}\big\langle (x\cdot\overline{x})(x\cdot\overline{x} - x'\cdot\overline{x})\big\rangle - \mathbb{E}\big\langle (x\cdot\overline{x})\big(|x|^2 - |x'|^2\big)\big\rangle. \end{aligned}$$

The Gaussian integration by parts formula (Theorem 4.5 in [50]) is now used to integrate out the noise z. Recalling also that when the Gibbs measure is an average over the two variables x and x', the underlying Hamiltonian is $H_N(t,h,x)+H_N(t,h,x')$,

$$\mathbb{E}\big\langle (x\cdot\overline{x})(z\cdot x)\big\rangle = \sqrt{2h}\mathbb{E}\big\langle (x\cdot\overline{x})\big(|x|^2 - x\cdot x'\big)\big\rangle \quad \text{and} \quad \mathbb{E}\big\langle (x\cdot\overline{x})(z\cdot x')\big\rangle = \sqrt{2h}\mathbb{E}\big\langle (x\cdot\overline{x})\big(|x'|^2 + x\cdot x' - 2x'\cdot x''\big)\big\rangle.$$

It follows by the Nishimori identity that

$$N\partial_h^2\overline{F}_N(t,h) = 2\mathbb{E}\big\langle (x\cdot x')^2\big\rangle - 4\mathbb{E}\big\langle (x\cdot x')(x\cdot x'')\big\rangle + 2\mathbb{E}\big\langle (x\cdot x')(x''\cdot x''')\big\rangle. \tag{2.194}$$

At this point, introduce the re-scaled and centred variable

$$y := \frac{1}{\sqrt{N}}(x - \langle x\rangle),$$

with y', y'', y''' denoting independent copies of y under the measure $\langle\cdot\rangle$, so that

$$N^2\mathbb{E}\big\langle (y\cdot y')^2\big\rangle = \mathbb{E}\big\langle (x\cdot x')^2\big\rangle - 2\mathbb{E}\big\langle (x\cdot x')(x\cdot x'')\big\rangle + \mathbb{E}\big\langle (x\cdot x')(x''\cdot x''')\big\rangle = \frac{N}{2}\partial_h^2\overline{F}_N(t,h). \tag{2.195}$$

This already shows that $\overline{F}_N$ is convex in the h variable. To compute the second derivative in t, rewrite (2.61) in the form

$$N^2\partial_t\overline{F}_N(t,h) = \mathbb{E}\big\langle xx^\top\cdot\bar{x}\bar{x}^\top\big\rangle, \tag{2.196}$$

and follow through the same calculation as for the second derivative in h; the only difference is that each occurrence of x is replaced by $xx^\top$, each occurrence of $\bar{x}$ is replaced by $\bar{x}\bar{x}^\top$, and so on, with z being replaced by W. This leads to

$$N^3\partial_t^2\overline{F}_N(t,h) = 2\mathbb{E}\big\langle (xx^\top\cdot x'x'^\top)^2\big\rangle - 4\mathbb{E}\big\langle (xx^\top\cdot x'x'^\top)(xx^\top\cdot x''x''^\top)\big\rangle + 2\mathbb{E}\big\langle (xx^\top\cdot x'x'^\top)(x''x''^\top\cdot x'''x'''^\top)\big\rangle,$$

so in terms of the re-scaled and centred variable

$$\xi := \frac{1}{N}(xx^\top - \langle xx^\top\rangle),$$

this reads

$$\frac{1}{2N}\partial_t^2\overline{F}_N(t,h) = \mathbb{E}\big\langle (\xi\cdot\xi')^2\big\rangle.$$

For the cross-derivative, start from (2.196) and differentiate in h to obtain

$$N^2\partial_h\partial_t\overline{F}_N(t,h) = \frac{1}{\sqrt{2h}}\mathbb{E}\big\langle (xx^\top\cdot\bar{x}\bar{x}^\top)(z\cdot x - z\cdot x')\big\rangle + 2\mathbb{E}\big\langle (xx^\top\cdot\bar{x}\bar{x}^\top)(x\cdot\bar{x} - x'\cdot\bar{x})\big\rangle - \mathbb{E}\big\langle (xx^\top\cdot\bar{x}\bar{x}^\top)\big(|x|^2 - |x'|^2\big)\big\rangle.$$

A Gaussian integration by parts allows to rewrite the first term on the right side above as

$$\mathbb{E}\big\langle (xx^\top\cdot\bar{x}\bar{x}^\top)(|x|^2 - x\cdot x' - (|x'|^2 + x\cdot x' - 2x'\cdot x''))\big\rangle.$$

An application of the Nishimori identity therefore yields that

$$N^2\partial_h\partial_t\overline{F}_N(t,h) = 2\mathbb{E}\big\langle (xx^\top\cdot x'x'^\top)(x\cdot x')\big\rangle - 4\mathbb{E}\big\langle (xx^\top\cdot x'x'^\top)(x\cdot x'')\big\rangle - 2\mathbb{E}\big\langle (xx^\top\cdot x'x'^\top)(x''\cdot x''')\big\rangle,$$

which can be rewritten in terms of the re-scaled and centred variables ξ as

$$\frac{1}{2N}\partial_h\partial_t\overline{F}_N(t,h) = \mathbb{E}\big\langle (\xi\cdot\xi')(y\cdot y')\big\rangle.$$

To see that the Hessian of $\overline{F}_N$ is non-negative definite, take $w = (a,b)\in\mathbb{R}^2$ and observe that

$$\frac{1}{2N}w\cdot\nabla^2\overline{F}_N(t,h)w = a^2\mathbb{E}\big\langle (\xi\cdot\xi')^2\big\rangle + b^2\mathbb{E}\big\langle (y\cdot y')^2\big\rangle + 2ab\mathbb{E}\big\langle (\xi\cdot\xi')(y\cdot y')\big\rangle = \mathbb{E}\big\langle (a\xi\cdot\xi' + by\cdot y')^2\big\rangle \geq 0,$$

as desired. ∎

Alternative proof of Theorem 2.5. The Arzelà-Ascoli theorem, the derivative computations (2.61) and (2.65), and the boundedness of the support of P_1 imply that the sequence $(\overline{F}_N)_{N\geqslant 1}$ is precompact. Denoting by f a subsequential limit, the idea is to apply the convex selection principle (Proposition 2.30) to f and show that it is the unique viscosity solution to the Hamilton-Jacobi equation (2.73). The proof therefore proceeds in two steps. First, it is shown that f satisfies the Hamilton-Jacobi equation (2.73) on a dense subset of $\mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}$, and then the convex selection principle is invoked.

Step 1: f satisfies (2.73) on a dense set. The goal of this step is to show that the set

$$\mathcal{A} := \left\{(t^*,h^*)\in\mathbb{R}_{>0}\times\mathbb{R}_{>0} \mid f \text{ is differentiable at } (t^*,h^*) \text{ and } \big(\partial_t f - \overline{\mathsf{H}}(\partial_h f)\big)(t^*,h^*) = 0\right\}$$

is dense in $\mathbb{R}_{>0}\times\mathbb{R}_{>0}$, and therefore in $\mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}$. Fix $(t_0,h_0)\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}$, and let V be a compact neighbourhood of (t_0,h_0) in $\mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}$. For each $\alpha\geq 1$, consider the mapping

$$\phi_\alpha(t,h) := \frac{\alpha}{2}\big(|t-t_0|^2+|h-h_0|^2\big),$$

and denote by (t_α,h_α) a maximizer of $f-\phi_\alpha$ on V. Writing L for the joint Lipschitz constant of f and rearranging the bound

$$f(t_\alpha,h_\alpha)-\frac{\alpha}{2}\big(|t_\alpha-t_0|^2+|h_\alpha-h_0|^2\big) = \big(f-\phi_\alpha\big)(t_\alpha,h_\alpha)\geqslant \big(f-\phi_\alpha\big)(t_0,h_0) = f(t_0,h_0)$$

reveals that

$$\big(|t_\alpha-t_0|+|h_\alpha-h_0|\big)^2 \leqslant \frac{2}{\alpha}|f(t_\alpha,h_\alpha)-f(t_0,h_0)| \leqslant \frac{2L}{\alpha}\big(|t_\alpha-t_0|+|h_\alpha-h_0|\big).$$

It follows that

$$|t_\alpha-t_0|+|h_\alpha-h_0|\leqslant\frac{2L}{\alpha},$$

so the sequence $(t_\alpha,h_\alpha)_{\alpha\geqslant 1}$ tends to (t_0,h_0) as α tends to infinity. In particular, for α sufficiently large, the point (t_α,h_α) is in the interior of V, and so is a local maximum of $f-\phi_\alpha$ as a function on $\mathbb{R}_{>0}\times\mathbb{R}_{>0}$. It will be a strict local maximum for the mapping $(t,x)\mapsto\phi_\alpha(t,x)-(|t-t_\alpha|^2+|h-h_\alpha|^2)$, so, up to modifying ϕ_α by this small parabola, assume without loss of generality that (t_α,h_α) is a strict local maximum of $f-\phi_\alpha$ as a function on $\mathbb{R}_{>0}\times\mathbb{R}_{>0}$. Arguing as in Step 1 of the proof of the convex selection principle, it is possible to show that f is differentiable at the contact point (t_α,h_α) with $(\partial_t f,\partial_h f)(t_\alpha,h_\alpha)=(\partial_t\phi_\alpha,\partial_h\phi_\alpha)(t_\alpha,h_\alpha)$. Invoking Lemma 2.29 implies that for every α sufficiently large, the point (t_α,h_α) belongs to $\mathcal{A}$. Recalling that $(t_\alpha,h_\alpha)_{\alpha\geqslant 1}$ tends to (t_0,h_0) as α tends to infinity, shows that $\mathcal{A}$ is dense in $\mathbb{R}_{>0}\times\mathbb{R}_{>0}$, as desired.

Step 2: applying the convex selection principle. Combining the previous step with the convex selection principle shows that f is a viscosity solution to the Hamilton-Jacobi equation (2.73). Together with the uniqueness result in Proposition 2.9, this implies that the sequence $(\overline{F}_N)_{N\geqslant 1}$ converges to the unique viscosity solution f to the Hamilton-Jacobi equation (2.73). The rest of the proof is identical to that of the first proof of Theorem 2.5 given in Section 2.5. ■

Having established Theorem 1.3 on the mutual information in the dense stochastic block model, the rest of this thesis is devoted to understanding the limit of the mutual information in the much harder context of the sparse stochastic block model using the Hamilton-Jacobi approach developed in this chapter.

Chapter 3

A Hamilton-Jacobi equation for the sparse stochastic block model

In this chapter, the first steps to study the sparse stochastic block model using the Hamilton-Jacobi approach are taken. Just like in the dense stochastic block model, it will be shown that the mutual information in the sparse stochastic block model can be identified with the free energy in a statistical inference problem up to an explicit additive constant and that this finite-volume free energy satisfies a Hamilton-Jacobi equation up to an error term that is expected to vanish in the limit of large system size. The main differences with the dense stochastic block model are that the Hamilton-Jacobi equation will now be infinite-dimensional and that the error term will be considerably harder to control. Establishing a well-posedness theory for the infinite-dimensional Hamilton-Jacobi equation will be the content of Chapter 4, and controlling the error term will be done in Chapter 6 leveraging the multioverlap concentration result developed in Chapter 5. The focus of this chapter is simply to derive the approximate infinite-dimensional Hamilton-Jacobi equation satisfied by the finite-volume free energy. In Section 3.1, the problem of computing the limit of the mutual information in the sparse stochastic block model is reformulated using the language of statistical mechanics by introducing a relevant Gibbs measure and free energy. In Section 3.2, this free energy is modified without changing its limiting value by introducing a Poisson random variable that depends on a continuous time parameter $t \geqslant 0$. This allows one to compute the time derivative of the free energy. To be able to close the equation, the free energy is also enriched by adding a term that depends on a non-negative measure μ to its Hamiltonian. The original free energy in the sparse stochastic block model will be given by the enriched free energy evaluated at the point $(t, \mu) = (1, 0)$. Notice that this is precisely the point appearing in Conjecture 1.4. In Section 3.3, the time derivative and the Gateaux derivative of the enriched free energy are computed. Finally, in Section 3.4 the limit of the enriched free energy at the initial time is computed, and an approximate Hamilton-Jacobi equation for the enriched free energy is derived. This leads to Conjecture 1.4. This chapter parallels Sections 2.1 – 2.3 of Chapter 2, and its contents rely heavily on Sections 2 and 3 of [49].

3.1 From statistical inference to statistical mechanics

The community detection problem associated with the sparse stochastic block model consists in recovering the assignment vector

$$\sigma^* := (\sigma_1^*, \dots, \sigma_N^*) \in \Sigma_N := \{-1, +1\}^N \tag{3.1}$$

given the random undirected graph $\mathbf{G}_N := (G_{ij})_{i,j\leqslant N}$ with vertex set $\{1,\dots,N\}$ constructed by stipulating that an edge between node i and node j is present with conditional probability

$$\mathbb{P}\{G_{ij}=1|\sigma^*\} := \frac{c+\Delta\sigma_i^*\sigma_j^*}{N} \tag{3.2}$$

for some $c>0$ and $\Delta\in(-c,c)$. Recall that the labels $\sigma_i^*\sim P^*$ are taken to be i.i.d. Bernoulli random variables with probability of success $p\in(0,1)$ and expectation $\overline{m}$,

$$p := P^*(1) = \mathbb{P}\{\sigma_i^*=1\} \quad \text{and} \quad \overline{m} := \mathbb{E}\sigma_1^* = 2p-1. \tag{3.3}$$

This means that the assignment vector σ^* follows a product distribution

$$\sigma^* \sim P_N^* := (P^*)^{\otimes N}. \tag{3.4}$$

To understand the mutual information

$$I(\mathbf{G}_N;\sigma^*) := \mathbb{E}\log\frac{\mathbb{P}(\mathbf{G}_N,\sigma^*)}{\mathbb{P}(\mathbf{G}_N)\mathbb{P}(\sigma^*)} = \mathbb{E}\int_{\mathbb{R}^N}\log\left(\frac{\mathrm{d}P_{\sigma^*|\mathbf{G}_N}}{\mathrm{d}P_N^*}(\sigma)\right)\mathrm{d}P_{\sigma^*|\mathbf{G}_N}(\sigma) \tag{3.5}$$

between the assignment vector σ^* and the graph $\mathbf{G}_N$, it will be useful to get a better grasp on the conditional law $P_{\sigma^*|\mathbf{G}_N}$ of the assignment vector σ^* given the graph $\mathbf{G}_N$. Observing that

$$\mathbb{P}\{\mathbf{G}_N=(G_{ij})|\sigma^*=\sigma\} = \prod_{i<j}\Big(\frac{c+\Delta\sigma_i\sigma_j}{N}\Big)^{G_{ij}}\Big(1-\frac{c+\Delta\sigma_i\sigma_j}{N}\Big)^{1-G_{ij}}, \tag{3.6}$$

Bayes' formula can be used to obtain the law of the assignment vector σ^* conditionally on the observation of $\mathbf{G}_N$. It can be written in the form of a Gibbs measure,

$$\mathbb{P}\big\{\sigma^*=\sigma|\mathbf{G}_N=(G_{ij})\big\} = \frac{\exp\big(H_N^\circ(\sigma)\big)P_N^*(\sigma)}{\int_{\Sigma_N}\exp\big(H_N^\circ(\tau)\big)\,\mathrm{d}P_N^*(\tau)}, \tag{3.7}$$

for the Hamiltonian

$$H_N^\circ(\sigma) := \sum_{i<j}\log\bigg[(c+\Delta\sigma_i\sigma_j)^{G_{ij}}\Big(1-\frac{c+\Delta\sigma_i\sigma_j}{N}\Big)^{1-G_{ij}}\bigg]. \tag{3.8}$$

Denoting its associated average free energy by

$$\overline{F}_N^\circ := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp H_N^\circ(\sigma)\,\mathrm{d}P_N^*(\sigma), \tag{3.9}$$

in the limit of large N, this average free energy coincides with the mutual information (3.5) up to an explicit additive constant. Indeed, (3.6) and Bayes' formula imply that

$$I(\mathbf{G}_N;\sigma^*) = \binom{N}{2}\mathbb{E}\log(c+\Delta\sigma_1^*\sigma_2^*)^{G_{12}}\Big(1-\frac{c+\Delta\sigma_1^*\sigma_2^*}{N}\Big)^{1-G_{12}} - N\overline{F}_N^\circ. \tag{3.10}$$

Averaging with respect to the randomness of G_{12}, and Taylor expanding the logarithm reveals that

$$\frac{1}{N}I(\mathbf{G}_N;\sigma^*) = \frac{1}{2}\mathbb{E}(c+\Delta\sigma_1^*\sigma_2^*)\log(c+\Delta\sigma_1^*\sigma_2^*) - \frac{c}{2} - \frac{\Delta\overline{m}^2}{2} - \overline{F}_N^\circ + \mathcal{O}\big(N^{-1}\big). \tag{3.11}$$

Just like in the dense stochastic block model, this observation reduces the task of understanding the limit of the mutual information (3.5), an information-theoretic quantity, to computing the limit of the free energy (3.9), a statistical mechanics quantity. Unlike for the dense stochastic block model, there is no universality property of the free energy (3.9) that allows one to map the sparse stochastic block model to a statistical inference problem with Gaussian noise such as the rank-one matrix estimation problem. The sparse stochastic block model itself therefore has to be understood using the Hamilton-Jacobi approach.

3.2 Modifying and enriching the sparse stochastic block model

To apply the Hamilton-Jacobi approach, the free energy (3.9) will be enriched so that it depends on a temporal variable and a spatial variable with respect to which derivatives can be taken. Before enriching the free energy, it will be convenient to modify it without changing its limiting value. Conditionally on the assignment vector σ^*, the modified Hamiltonian will be a sum of independent random variables, and the sum will be over a Poisson number of terms. The main advantage of this construction is that it will allow for the introduction of a continuous parameter $t \geqslant 0$ tuning the mean of the Poisson random variable. This will constitute the temporal enrichment of the free energy. The spatial enrichment is more sophisticated and will be discussed in due course.

To define the modified free energy more precisely, introduce a random variable $\Pi_1 \sim \mathrm{Poi}\binom{N}{2}$ as well as an independent family of i.i.d. random matrices $(G^k)_{k\geqslant 1}$ each having conditionally independent entries $(G^k_{i,j})_{i,j\leqslant N}$ taking values in $\{0,1\}$ with conditional distribution

$$\mathbb{P}\{G^k_{i,j} = 1 | \sigma^*\} := \frac{c + \Delta\sigma_i^* \sigma_j^*}{N}. \tag{3.12}$$

Given a collection of random indices $(i_k, j_k)_{k\geqslant 1}$ sampled uniformly at random from $\{1,\ldots,N\}^2$, independently of the other random variables, define the Hamiltonian H_N on Σ_N by

$$H_N(\sigma) := \sum_{k\leqslant \Pi_1} \log\left[(c+\Delta\sigma_{i_k}\sigma_{j_k})^{G^k_{i_k,j_k}}\left(1 - \frac{c+\Delta\sigma_{i_k}\sigma_{j_k}}{N}\right)^{1-G^k_{i_k,j_k}}\right], \tag{3.13}$$

and write

$$\overline{F}_N := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N} \exp H_N(\sigma)\,\mathrm{d}P_N^*(\sigma) \tag{3.14}$$

for its associated free energy. Through a slight abuse of terminology, the modified Hamiltonian (3.13) and the modified free energy (3.14) will often be termed the Hamiltonian and the free energy, respectively. This is justified because the free energies (3.9) and (3.14) are asymptotically equivalent by the Binomial-Poisson approximation theorem (Proposition A.15).

Proposition 3.1. *The free energies* (3.9) *and* (3.14) *are asymptotically equivalent,*

$$\lim_{N\to+\infty} \left|\overline{F}_N - \overline{F}_N^\circ\right| = 0. \tag{3.15}$$

Proof. Introduce the Hamiltonians

$$\widetilde{H}_N^\circ(\sigma) := \sum_{i<j}\left(G_{ij}\log(c+\Delta\sigma_i\sigma_j) - \frac{c+\Delta\sigma_i\sigma_j}{N}\right) \quad \text{and} \quad \widetilde{H}_N(\sigma) := \sum_{k\leqslant \Pi_1}\left(G^k_{i_k,j_k}\log(c+\Delta\sigma_{i_k}\sigma_{j_k}) - \frac{c+\Delta\sigma_{i_k}\sigma_{j_k}}{N}\right)$$

on Σ_N, and denote by

$$\widetilde{F}_N^\circ := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp\widetilde{H}_N^\circ(\sigma)\,\mathrm{d}P_N^*(\sigma) \quad \text{and} \quad \widetilde{F}_N := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp\widetilde{H}_N(\sigma)\,\mathrm{d}P_N^*(\sigma)$$

their associated free energy functionals. The proof proceeds in three steps. First, a Taylor expansion is used to show that $\overline{F}_N^\circ$ and $\widetilde{F}_N^\circ$ are asymptotically equivalent and that $\overline{F}_N$ and $\widetilde{F}_N$ are asymptotically equivalent. This reduces the task of showing that $\overline{F}_N^\circ$ and $\overline{F}_N$ are asymptotically equivalent to that of showing that $\widetilde{F}_N^\circ$ and $\widetilde{F}_N$ are asymptotically equivalent. The free energy functional $\widetilde{F}_N$ is then shown to be asymptotically equivalent to a functional $\widetilde{F}_N'$ which has the same structure as $\widetilde{F}_N^\circ$. Finally, an interpolation argument leveraging this similarity in structure and the Binomial-Poisson approximation theorem is used to prove that $\widetilde{F}_N^\circ$ and $\widetilde{F}_N'$ are asymptotically equivalent.

Step 1: reducing to the asymptotic equivalence of $\widetilde{F}_N^\circ$ and $\widetilde{F}_N$. A Taylor expansion of the logarithm shows that for any $\sigma \in \Sigma_N$,

$$\left|H_N^\circ(\sigma) - \widetilde{H}_N^\circ(\sigma)\right| \leqslant \sum_{i<j}\left|\frac{c+\Delta\sigma_i\sigma_j}{N} - (1-G_{ij})\Big(\frac{c+\Delta\sigma_i\sigma_j}{N}\Big)\right| + \mathcal{O}(1) \leqslant \frac{c+|\Delta|}{N}\sum_{i<j}G_{ij} + \mathcal{O}(1)$$

$$\left|H_N(\sigma) - \widetilde{H}_N(\sigma)\right| \leqslant \sum_{k\leqslant\Pi_1}\left|\frac{c+\Delta\sigma_{i_k}\sigma_{j_k}}{N} - (1-G_{i_k,j_k}^k)\Big(\frac{c+\Delta\sigma_{i_k}\sigma_{j_k}}{N}\Big)\right| + \mathcal{O}\Big(\frac{\Pi_1}{N^2}\Big) \leqslant \frac{c+|\Delta|}{N}\sum_{k\leqslant\Pi_1}G_{i_k,j_k}^k + \mathcal{O}\Big(\frac{\Pi_1}{N^2}\Big).$$

Since these bounds are uniform in $\sigma \in \Sigma_N$ and $\mathbb{E}\Pi_1 = \binom{N}{2}$, they imply that

$$\left|\overline{F}_N^\circ - \widetilde{F}_N^\circ\right| \leqslant \frac{c+|\Delta|}{N^2}\sum_{i<j}\mathbb{E}G_{ij} + \mathcal{O}\big(N^{-1}\big) \leqslant \frac{(c+|\Delta|)^2}{N} + \mathcal{O}\big(N^{-1}\big) = \mathcal{O}\big(N^{-1}\big),$$

$$\left|\overline{F}_N - \widetilde{F}_N\right| \leqslant \frac{c+|\Delta|}{N^2}\mathbb{E}\sum_{k\leqslant\Pi_1}G_{i_k,j_k}^k + \mathcal{O}\big(\mathbb{E}\Pi_1/N^3\big) \leqslant \frac{(c+|\Delta|)^2}{N^3}\mathbb{E}\Pi_1 + \mathcal{O}\big(N^{-1}\big) = \mathcal{O}\big(N^{-1}\big).$$

By the triangle inequality, it therefore suffices to show that $\widetilde{F}_N^\circ$ and $\widetilde{F}_N$ are asymptotically equivalent.

Step 2: asymptotic equivalence of $\widetilde{F}_N$ and $\widetilde{F}_N'$. The free energy $\widetilde{F}_N$ is now rewritten in a way that more closely resembles $\widetilde{F}_N^\circ$. For each pair $1 \leqslant i \leqslant j \leqslant N$, introduce the random index set $\mathcal{I}_{i,j} := \{k \leqslant \Pi_1 \mid \{i_k, j_k\} = \{i,j\}\}$ in such a way that

$$\widetilde{H}_N(\sigma) = \sum_{i<j}\left(\widetilde{G}_{i,j}\log(c+\Delta\sigma_i\sigma_j) - \frac{c+\Delta\sigma_i\sigma_j}{N}\right) - \sum_{i\leqslant N}\left(\widetilde{G}_{i,i}\log(c+\Delta) - \frac{c+\Delta}{N}\right)$$

for the random variables $\widetilde{G}_{i,j} := \sum_{k\in\mathcal{I}_{i,j}}G_{i,j}^k$. Observe that $\widetilde{G}_{i,j}$ counts the number of indices $k \leqslant \Pi_1$ with $\{i_k, j_k\} = \{i,j\}$ and $G_{i,j}^k = 1$. By independence of the random variables involved and the Poisson colouring theorem, $\widetilde{G}_{i,j}$ is a Poisson random variable with mean

$$\widetilde{\lambda}_{i,j} := \mathbb{E}\Pi_1\mathbb{P}\big\{\{i_1, j_1\} = \{i,j\}\big\}\mathbb{P}\{G_{i,j}^1 = 1\} = \mathcal{O}\big(N^{-1}\big)$$

This motivates the introduction of the Hamiltonian

$$\widetilde{H}_N'(\sigma) := \sum_{i<j}\left(\widetilde{G}_{i,j}\log(c+\Delta\sigma_i\sigma_j) - \frac{c+\Delta\sigma_i\sigma_j}{N}\right)$$

and of its associated free energy

$$\widetilde{F}_N' := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N} \exp\widetilde{H}_N'(\sigma)\,\mathrm{d}P_N^*(\sigma).$$

Indeed, the free energy functionals $\widetilde{F}_N$ and $\widetilde{F}_N'$ are asymptotically equivalent,

$$\left|\widetilde{F}_N' - \widetilde{F}_N\right| \leqslant \frac{|\log(c+\Delta)|}{N}\sum_{i\leqslant N}\mathbb{E}\widetilde{G}_{i,i} + \frac{c+|\Delta|}{N} \leqslant \frac{|\log(c+\Delta)|}{N}\sum_{i\leqslant N}\widetilde{\lambda}_{i,i} + \frac{2c}{N} = \mathcal{O}\left(N^{-1}\right).$$

By the previous step and the triangle inequality, it therefore suffices to show that $\widetilde{F}_N^\circ$ and $\widetilde{F}_N'$ are asymptotically equivalent.

Step 3: interpolating between $\widetilde{F}_N^\circ$ and $\widetilde{F}_N'$. For any random vector $Y := (Y_{i,j})_{i<j}$ introduce the Hamiltonian $\widetilde{H}_N(\cdot,Y)$ and the measure $\widetilde{P}_N^\star$ on Σ_N defined by

$$\widetilde{H}_N(\sigma,Y) := \sum_{i<j} Y_{i,j}\log(c+\Delta\sigma_i\sigma_j) \quad \text{and} \quad \widetilde{P}_N^*(\sigma) := \exp\Big(-\sum_{i<j}\frac{c+\Delta\sigma_i\sigma_j}{N}\Big)P_N^*(\sigma).$$

As usual, write

$$F_N(Y) := \frac{1}{N}\log\int_{\Sigma_N}\exp\widetilde{H}_N(\sigma,Y)\,\mathrm{d}\widetilde{P}_N^*(\sigma) \quad \text{and} \quad \widetilde{F}_N(Y) := \mathbb{E}F_N(Y)$$

for the associated free energy functionals, and $\langle\cdot\rangle$ for the associated Gibbs measure. Conditionally on the randomness of the assignment vector σ^*, for each $1\leqslant i<j\leqslant N$, introduce a Poisson random variable $\Pi_{i,j}$ with mean $\lambda_{i,j} := N^{-1}(c+\Delta\sigma_i^*\sigma_j^*)$. A direct computation shows that there exists a constant $C<+\infty$ such that for any random vector Y,

$$\left|\partial_{Y_{i,j}}F_N(Y)\right| = \frac{1}{N}\left|\left\langle\partial_{Y_{i,j}}\widetilde{H}_N(\sigma,Y)\right\rangle\right| = \frac{1}{N}\left|\langle\log(c+\Delta\sigma_i\sigma_j)\rangle\right| \leqslant \frac{C}{N}$$

It is understood that the constant $C<+\infty$ may change from one occurrence to the next, only making sure that it does not depend on N. It follows by the mean value theorem that

$$\begin{aligned}
\left|\widetilde{F}_N' - \widetilde{F}_N(\Pi)\right| &= \left|\widetilde{F}_N(\widetilde{G}) - \widetilde{F}_N(\Pi)\right| \leqslant \frac{C}{N}\sum_{i<j}\mathbb{E}\left|\widetilde{G}_{i,j} - \Pi_{i,j}\right|,\\
\left|\widetilde{F}_N(\Pi) - \widetilde{F}_N^\circ\right| &= \left|\widetilde{F}_N(\Pi) - \widetilde{F}_N(G)\right| \leqslant \frac{C}{N}\sum_{i<j}\mathbb{E}\left|\Pi_{i,j} - G_{ij}\right|.
\end{aligned}$$

To bound the first expression in this display, recall that $\widetilde{G}_{i,j}$ is a Poisson random variable with mean

$$\widetilde{\lambda}_{i,j} = \frac{N-1}{N}\cdot\frac{c+\Delta\sigma_i^*\sigma_j^*}{N} = \lambda_{i,j} - \frac{\lambda_{i,j}}{N}.$$

This means that, conditionally on the randomness of σ^*, the Poisson random variable $\Pi_{i,j}$ is equal in distribution to $\widetilde{G}_{i,j} + \Pi_{i,j}'$ for a Poisson random variable $\Pi_{i,j}'$ with mean $\lambda_{i,j}' := N^{-1}\lambda_{i,j}$. It follows that

$$\left|\widetilde{F}_N' - \widetilde{F}_N(\Pi)\right| \leqslant \frac{C}{N}\sum_{i<j}\mathbb{E}\Pi_{i,j}' = \frac{C}{N}\sum_{i<j}\lambda_{i,j}' = \mathcal{O}\left(N^{-1}\right). \tag{3.16}$$

To bound the second expression in the previous display, observe that by the triangle inequality,

$$\begin{aligned}\mathbb{E}\left|\Pi_{i,j}-G_{ij}\right| &\leqslant \mathbb{E}\left|\Pi_{i,j}-G_{ij}\right|\mathbf{1}\{\Pi_{i,j}\geqslant 2\}+\mathbb{E}\left|\Pi_{i,j}-G_{ij}\right|\mathbf{1}\{\Pi_{i,j}\leqslant 2\}\mathbf{1}\{\Pi_{i,j}\neq G_{ij}\}\\ &\leqslant \mathbb{E}\Pi_{i,j}\mathbf{1}\{\Pi_{i,j}\geqslant 2\}+\mathbb{P}\{\Pi_{i,j}\geqslant 2\}+3\mathbb{P}\{\Pi_{i,j}\neq G_{ij}\}\\ &\leqslant 3\mathbb{P}\{\Pi_{i,j}\neq G_{ij}\}+\mathcal{O}\left(\lambda_{i,j}^2\right).\end{aligned}$$

Taking the infimum over all couplings of $\Pi_{i,j}$ and G_{ij}, recalling the definition of the total variation distance in (A.28), and invoking the Binomial-Poisson approximation theorem shows that

$$\left|\widetilde{F}_N(\Pi)-\widetilde{F}_N^\circ\right|\leqslant\frac{C}{N}\sum_{i<j}\Big(\mathrm{TV}\big(\Pi_{i,j},G_{ij}\big)+\mathcal{O}\left(\lambda_{i,j}^2\right)\Big)=\mathcal{O}\left(N\lambda_{i,j}^2\right)=\mathcal{O}\left(N^{-1}\right).$$

This establishes the asymptotic equivalence of $\widetilde{F}_N^\circ$ and $\widetilde{F}_N'$, and completes the proof. ■

Together with the relationship (3.11) between the free energy (3.9) and the mutual information (3.5), this result implies that

$$\frac{1}{N}I(\mathbf{G}_N;\sigma^*)=\frac{1}{2}\mathbb{E}\big(c+\Delta\sigma_1^*\sigma_2^*\big)\log\big(c+\Delta\sigma_1^*\sigma_2^*\big)-\frac{c}{2}-\frac{\Delta\overline{m}^2}{2}-\overline{F}_N+o(1). \tag{3.17}$$

The problem of finding the asymptotic value of the mutual information (3.5) has therefore been reduced to the task of determining the limit of the free energy (3.14).

The advantage of the free energy (3.14) over the free energy (3.9) is that it can easily be enriched in time. For each $t\geqslant 0$, let $\Pi_t\sim\mathrm{Poi}\,t\binom{N}{2}$ be a Poisson random variable with mean $t\binom{N}{2}$, and consider a time-dependent version of the Hamiltonian (3.13) defined on Σ_N by

$$H_N^t(\sigma):=\sum_{k\leqslant\Pi_t}\log\left[(c+\Delta\sigma_{i_k}\sigma_{j_k})^{G^k_{i_k,j_k}}\left(1-\frac{c+\Delta\sigma_{i_k}\sigma_{j_k}}{N}\right)^{1-G^k_{i_k,j_k}}\right]. \tag{3.18}$$

Notice that this is the Hamiltonian associated with the task of inferring the signal σ^* from the data

$$\mathcal{D}_N^t:=\big(\Pi_t,(i_k,j_k)_{k\leqslant\Pi_t},(G^k_{i_k,j_k})_{k\leqslant\Pi_t}\big) \tag{3.19}$$

in the sense that the Gibbs measure associated with $\widetilde{H}_N^t$ is the conditional law of σ^* given the data $\widetilde{\mathcal{D}}_N^t$.

Although the free energy associated with the Hamiltonian (3.18) can be differentiated in time, there is no way of closing the equation without being able to take derivatives in a spatial variable. Indeed, the situation is analogous to that encountered in Section 2.3 in the context of the symmetric rank-one matrix estimation problem. Once again, the free energy will have to be enriched by adding a spatial component to the Hamiltonian. The main difference between the present setting and that of the symmetric rank-one matrix estimation problem is that the finite-dimensional spatial variable $h\in\mathbb{R}_{\geqslant 0}$ will be replaced by an infinite-dimensional variable μ corresponding to a non-negative measure. The non-negative measure μ will be decomposed as $\mu=s\overline{\mu}$ for $s\geqslant 0$ and a probability measure $\overline{\mu}$. It will be used to consider a situation in which the graph of connections of a simpler setting in which each individual i can form connections with its own set of neighbour candidates is also observed. To be more specific, each individual i will have an independent number $\mathrm{Poi}(sN)$ of neighbour candidates indexed by the pairs (i,k) for $k\leqslant\mathrm{Poi}(sN)$. Each candidate neighbour (i,k) will be independently assigned a random "type" $x_{i,k}$ sampled from the distribution $\overline{\mu}$, and an edge will be present between individual

i and its candidate neighbour (i,k) with probability $N^{-1}(c+\Delta\sigma_i^* x_{i,k})$. In the inference problem, the "types" $x_{i,k}$ will be revealed to the statistician. The lack of interactions between individuals makes this piece of information much simpler to understand than the original community detection problem associated with the sparse stochastic block model. Together, the temporal enrichment previously discussed and this spatial enrichment, will lead to an "enriched" free energy, function of t and μ, and it will be possible to recover the free energy (3.14) by evaluating this enriched free energy at the point $(t,\mu)=(1,0)$.

To define the spatial enrichment of the free energy (3.14) precisely, denote by $\Pr[-1,1]$ the set of probability measures on $[-1,1]$, and given $\mu\in\Pr[-1,1]$, consider a sequence $x=(x_{i,k})_{i,k\geqslant 1}$ of i.i.d. random variables with law μ. For each $s>0$ and $i\geqslant 1$, let $\Pi_{i,s}\sim\mathrm{Poi}(sN)$ be Poisson random variables with mean sN that are independent over $i\geqslant 1$. Introduce the Hamiltonian on Σ_N defined by

$$\widetilde{H}_N^{s,\mu}(\sigma):=\sum_{i\leqslant N}\sum_{k\leqslant\Pi_{i,s}}\log\left[(c+\Delta\sigma_i x_{i,k})^{\widetilde{G}_{i,k}^x}\Big(1-\frac{c+\Delta\sigma_i x_{i,k}}{N}\Big)^{1-\widetilde{G}_{i,k}^x}\right], \tag{3.20}$$

where the random variables $(\widetilde{G}_{i,k}^x)_{i,k\geqslant 1}$ are conditionally independent with conditional distribution

$$\mathbb{P}\big\{\widetilde{G}_{i,k}^x=1|\sigma^*,x\big\}:=\frac{c+\Delta\sigma_i^* x_{i,k}}{N}. \tag{3.21}$$

As alluded to above, this is the Hamiltonian associated with the task of inferring the signal σ^* from the data

$$\widetilde{\mathcal{D}}_N^{s,\mu}:=\Big(\Pi_{i,s},(x_{i,k})_{k\leqslant\Pi_{i,s}},(\widetilde{G}_{i,k})_{k\leqslant\Pi_{i,s}}\Big)_{i\leqslant N}, \tag{3.22}$$

in the sense that the Gibbs measure associated with $\widetilde{H}_N^{s,\mu}$ is the conditional law of σ^* given the data $\widetilde{\mathcal{D}}_N^{s,\mu}$.

The temporal and spatial enrichments of the free energy (3.14) can now be combined to define the enriched free energy. Introduce the enriched Hamiltonian on Σ_N defined by

$$\widetilde{H}_N^{t,s,\mu}(\sigma):=H_N^t(\sigma)+\widetilde{H}_N^{s,\mu}(\sigma), \tag{3.23}$$

and denote by

$$\widetilde{F}_N(t,s,\mu):=\frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp\widetilde{H}_N^{t,s,\mu}(\sigma)\,\mathrm{d}P_N^*(\sigma) \tag{3.24}$$

its associated free energy. Observe that $\widetilde{F}_N(1,0,\mu)=\overline{F}_N$ and that (3.23) is the Hamiltonian associated with the task of inferring the signal σ^* from the data

$$\widetilde{\mathcal{D}}_N^{t,s,\mu}:=(\mathcal{D}_N^t,\widetilde{\mathcal{D}}_N^{s,\mu}), \tag{3.25}$$

where the randomness in these two data sets is taken to be independent conditionally on σ^*. To obtain a Hamilton-Jacobi equation, it will be convenient to reinterpret the enriched free energy (3.24) as a function of the time-parameter $t>0$ and a finite measure μ; the parameter s will become the total mass of this finite measure. Denote by $\mathcal{M}_s$ the space of signed measures on $[-1,1]$,

$$\mathcal{M}_s:=\big\{\mu\mid\mu\text{ is a signed measure on }[-1,1]\big\}, \tag{3.26}$$

and by $\mathcal{M}_+$ the cone of non-negative measures on this interval,

$$\mathcal{M}_+ := \{\mu \in \mathcal{M}_s \mid \mu \text{ is a non-negative measure}\}. \tag{3.27}$$

The convention that a signed measure can only take finite values is followed, so every $\mu \in \mathcal{M}_+$ must have finite total mass. This implies that every non-zero measure $\mu \in \mathcal{M}_+$ induces a probability measure,

$$\overline{\mu} := \frac{\mu}{\mu[-1,1]} \in \Pr[-1,1]. \tag{3.28}$$

Given a measure $\mu \in \mathcal{M}_+$, the enriched Hamiltonian $H_N^{t,\mu}$ on Σ_N is defined by

$$H_N^{t,\mu}(\sigma) := \widetilde{H}_N^{t,\mu[-1,1],\overline{\mu}}(\sigma), \tag{3.29}$$

where $\widetilde{H}_N^{0,0} = 0$ for the zero measure by continuity. The free energy associated with this Hamiltonian is given by

$$\overline{F}_N(t,\mu) := \widetilde{F}_N\big(t,\mu[-1,1],\overline{\mu}\big), \tag{3.30}$$

and once again $\overline{F}_N = \overline{F}_N(1,0)$, where 0 denotes the zero measure. The free energy in (3.30) will be termed the enriched free energy, and it will now be shown that, up to an error that is expected to be small, it satisfies an infinite-dimensional Hamilton-Jacobi equation.

3.3 Differentiating the sparse stochastic block model free energy

To derive the approximate infinite-dimensional Hamilton-Jacobi equation satisfied by the enriched free energy (3.30), its temporal and spatial derivatives must first be computed. It will be convenient to write $\langle\cdot\rangle$ for the average with respect to the Gibbs measure associated with the Hamiltonian (3.29). This means that for any bounded measurable function $f = f(\sigma^1,\dots,\sigma^n)$ of finitely many replicas,

$$\langle f(\sigma^1,\dots,\sigma^n)\rangle := \frac{\int_{\Sigma_N^n} f(\sigma^1,\dots,\sigma^n)\prod_{\ell\leq n}\exp H_N^{t,\mu}(\sigma^\ell)\,\mathrm{d}P_N^*(\sigma^\ell)}{\big(\int_{\Sigma_N}\exp H_N^{t,\mu}(\sigma)\,\mathrm{d}P_N^*(\sigma)\big)^n}. \tag{3.31}$$

In this notation, the replicas $\sigma^1,\dots,\sigma^n$ represent i.i.d. samples under the random measure $\langle\cdot\rangle$. By construction,

$$\langle f(\sigma^1)\rangle = \mathbb{E}\big[f(\sigma^*)\big|\mathcal{D}_N^{t,\mu}\big], \tag{3.32}$$

where $\mathcal{D}_N^{t,\mu} := \widetilde{\mathcal{D}}_N^{t,\mu[-1,1],\overline{\mu}}$ is the data defined in (3.25). Just like in the setting of the symmetric rank-one matrix estimation problem, computations will be considerably simplified by the Nishimori identity (Proposition 2.2). This identity allows the interchange of one replica σ^ℓ by the signal σ^* when taking an average with respect to all sources of randomness, thus avoiding a cascade of new replicas as the free energy is differentiated. In the present context, it states that, for every bounded measurable function $f = f(\sigma^1,\dots,\sigma^n,\mathcal{D}_N^{t,\mu})$ of finitely many replicas and the data,

$$\mathbb{E}\big\langle f\big(\sigma^1,\sigma^2,\dots,\sigma^n,\mathcal{D}_N^{t,\mu}\big)\big\rangle = \mathbb{E}\big\langle f\big(\sigma^*,\sigma^2,\dots,\sigma^n,\mathcal{D}_N^{t,\mu}\big)\big\rangle. \tag{3.33}$$

This can be first verified for functions of product form using (3.32), and then extended to all bounded measurable functions using Dynkin's π-λ theorem as in Proposition 2.2.

3.3.1 Computing the time derivative of the enriched free energy

To begin with, the time derivative of the enriched free energy (3.30) is computed with the spatial parameter fixed to some finite measure $\mu \in \mathcal{M}_+$. The calculations of this section and the next closely follow [14] and Lemma 6 of [92]. For each parameter $\lambda > 0$ and every integer $m \geqslant 0$, write

$$\pi(\lambda, m) := \frac{\lambda^m}{m!}\exp(-\lambda) \tag{3.34}$$

for the mass attributed to the atom m by a $\mathrm{Poi}(\lambda)$ distribution, with the convention that $\pi(\lambda,-1) = 0$. Denote by

$$H_{N,m}(\sigma) := \sum_{k\leqslant m}\log\left[\left(c+\Delta\sigma_{i_k}\sigma_{j_k}\right)^{G^k_{i_k,j_k}}\left(1-\frac{c+\Delta\sigma_{i_k}\sigma_{j_k}}{N}\right)^{1-G^k_{i_k,j_k}}\right] \tag{3.35}$$

the Hamiltonian (3.13) conditional on there being m terms in the sum, and introduce the partition function

$$Z_{N,m} := \int_{\Sigma_N}\exp\left(H_{N,m}(\sigma)+\widetilde{H}_N^{\mu[-1,1],\overline{\mu}}(\sigma)\right)\mathrm{d}P_N^*(\sigma). \tag{3.36}$$

In this notation, the enriched free energy (3.30) may be expressed as

$$\overline{F}_N(t,\mu) = \frac{1}{N}\sum_{m\geqslant 0}\pi\left(t\binom{N}{2},m\right)\mathbb{E}\log Z_{N,m}. \tag{3.37}$$

To take the time derivative of this expression, it will be convenient to observe that

$$\partial_\lambda\pi(\lambda,m) = \pi(\lambda,m-1)-\pi(\lambda,m). \tag{3.38}$$

Lemma 3.2. *For any $t > 0$ and $\mu \in \mathcal{M}_+$,*

$$\partial_t\overline{F}_N(t,\mu) = \frac{1}{2}\mathbb{E}\left(c+\Delta\langle\sigma_i\sigma_j\rangle\right)\log\left(c+\Delta\langle\sigma_i\sigma_j\rangle\right)-\frac{\Delta\overline{m}^2}{2}-\frac{c}{2}+\mathcal{O}(N^{-1}), \tag{3.39}$$

where the indices $i, j \in \{1,\dots,N\}$ are uniformly sampled independently of all other sources of randomness.

Proof. To simplify notation, let $\lambda(t) := t\binom{N}{2}$. Leveraging (3.38) to differentiate the right side of (3.37) yields

$$\begin{aligned}\partial_t\overline{F}_N(t,\mu) &= \frac{1}{N}\binom{N}{2}\sum_{m\geqslant 0}\left(\pi(\lambda(t),m-1)-\pi(\lambda(t),m)\right)\mathbb{E}\log Z_{N,m}\\ &= \frac{1}{N}\binom{N}{2}\sum_{m\geqslant 0}\pi(\lambda(t),m)\mathbb{E}\log\frac{Z_{N,m+1}}{Z_{N,m}}.\end{aligned} \tag{3.40}$$

Denote by $i, j \in \{1,\dots,N\}$ uniformly sampled indices, and write $G_{i,j}$ for a random variable with conditional distribution (3.2). These random variables are taken to be independent of all other sources of randomness and of each other. Since

$$Z_{N,m+1} \overset{d}{=} \int_{\Sigma_N}(c+\Delta\sigma_i\sigma_j)^{G_{i,j}}\left(1-\frac{c+\Delta\sigma_i\sigma_j}{N}\right)^{1-G_{i,j}}\exp\left(H_{N,m}(\sigma)+H_N^{\mu[-1,1],\overline{\mu}}(\sigma)\right)\mathrm{d}P_N^*(\sigma),$$

where $\overset{d}{=}$ denotes equality in distribution, it follows by (3.40) and the definition of the Gibbs average in (3.31) that

$$\partial_t \overline{F}_N(t,\mu) = \frac{1}{N}\binom{N}{2}\mathbb{E}\log\left\langle (c+\Delta\sigma_i\sigma_j)^{G_{i,j}}\left(1-\frac{c+\Delta\sigma_i\sigma_j}{N}\right)^{1-G_{i,j}}\right\rangle.$$

Remembering the explicit form of the conditional distribution (3.2), and averaging with respect to the randomness of $G_{i,j}$ reveals that

$$\partial_t \overline{F}_N(t,s,\mu) = \frac{1}{2}\mathbb{E}(c+\Delta\sigma_i^*\sigma_j^*)\log\langle c+\Delta\sigma_i\sigma_j\rangle + \frac{N}{2}\mathbb{E}\left(1-\frac{c+\Delta\sigma_i^*\sigma_j^*}{N}\right)\log\left\langle 1-\frac{c+\Delta\sigma_i\sigma_j}{N}\right\rangle + \mathcal{O}(N^{-1}).$$

Taylor expanding the logarithm and keeping only first-order terms reduces this to

$$\begin{aligned}\partial_t \overline{F}_N(t,\mu) &= \frac{1}{2}\mathbb{E}(c+\Delta\sigma_i^*\sigma_j^*)\log\langle c+\Delta\sigma_i\sigma_j\rangle - \frac{\Delta}{2}\mathbb{E}\langle\sigma_i\sigma_j\rangle - \frac{c}{2} + \mathcal{O}(N^{-1})\\ &= \frac{1}{2}\mathbb{E}(c+\Delta\sigma_i^*\sigma_j^*)\log\langle c+\Delta\sigma_i\sigma_j\rangle - \frac{\Delta}{2}\mathbb{E}\sigma_i^*\sigma_j^* - \frac{c}{2} + \mathcal{O}(N^{-1})\\ &= \frac{1}{2}\mathbb{E}(c+\Delta\sigma_i^*\sigma_j^*)\log(c+\Delta\langle\sigma_i\sigma_j\rangle) - \frac{\Delta\overline{m}^2}{2} - \frac{c}{2} + \mathcal{O}(N^{-1}),\end{aligned}$$

where the second equality uses the Nishimori identity (3.33) and the third equality uses the fact that i and j are distinct with overwhelming probability. Noticing that the Gibbs average $\langle\sigma_i\sigma_j\rangle$ is a measurable function of the data by (3.32), and applying the Nishimori identity (3.33) completes the proof. ∎

To compare (3.39) with the Gateaux derivative of the enriched free energy which will be computed below, it will be convenient to Taylor expand the logarithm. This will make the dependence of the time derivative of the enriched free energy on the multioverlaps

$$R_{\ell_1,\ldots,\ell_n} := \frac{1}{N}\sum_{i\leqslant N}\sigma_i^{\ell_1}\cdots\sigma_i^{\ell_n} \tag{3.41}$$

associated with the enriched Hamiltonian (3.29) explicit. Here $(\sigma^\ell)_{\ell\geqslant 1}$ denotes a sequence of i.i.d. replicas sampled from the Gibbs measure (3.31). To simplify notation, it will be convenient to write $R_{[n]} := R_{1,\ldots,n}$.

Corollary 3.3. *For any $t>0$ and $\mu\in\mathcal{M}_+$,*

$$\partial_t \overline{F}_N(t,\mu) = \frac{1}{2}\left(c+\Delta\overline{m}^2\right)\log(c) + \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\left\langle R_{[n]}^2\right\rangle - \frac{c}{2} + \mathcal{O}(N^{-1}). \tag{3.42}$$

Proof. A Taylor expansion of the logarithm shows that

$$\log(c+\Delta\langle\sigma_i\sigma_j\rangle) = \log(c) + \log\left(1+\frac{\Delta}{c}\langle\sigma_i\sigma_j\rangle\right) = \log(c) - \sum_{n\geqslant 1}\frac{(-\Delta/c)^n}{n}\langle\sigma_i\sigma_j\rangle^n.$$

Together with the Nishimori identity (3.33) this implies that

$$\mathbb{E}(c+\Delta\langle\sigma_i\sigma_j\rangle)\log(c+\Delta\langle\sigma_i\sigma_j\rangle) = \left(c+\Delta\overline{m}^2\right)\log(c) - \sum_{n\geqslant 1}\frac{(-\Delta/c)^n}{n}\mathbb{E}(c+\Delta\langle\sigma_i\sigma_j\rangle)\langle\sigma_i\sigma_j\rangle^n. \tag{3.43}$$

Averaging with respect to the randomness of the uniformly sampled indices $i,j\in\{1,\ldots,N\}$ reveals that

$$\mathbb{E}(c+\Delta\langle\sigma_i\sigma_j\rangle)\langle\sigma_i\sigma_j\rangle^n = c\mathbb{E}\left\langle R_{[n]}^2\right\rangle + \Delta\mathbb{E}\left\langle R_{[n+1]}^2\right\rangle.$$

Remembering that $|\Delta| < c$, and noticing that $\mathbb{E}\langle R_1^2\rangle = \overline{m}^2 + \mathcal{O}(N^{-1})$ by the Nishimori identity, it follows that

$$\sum_{n\geqslant 1}\frac{(-\Delta/c)^n}{n}\mathbb{E}(c+\Delta\langle\sigma_i\sigma_j\rangle)\langle\sigma_i\sigma_j\rangle^n = -\Delta\mathbb{E}\big\langle R_1^2\big\rangle + c\sum_{n\geqslant 2}\bigg(\frac{(-\Delta/c)^n}{n} - \frac{(-\Delta/c)^n}{n-1}\bigg)\mathbb{E}\big\langle R_{[n]}^2\big\rangle + \mathcal{O}(N^{-1})$$
$$= -\Delta\overline{m}^2 - c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\big\langle R_{[n]}^2\big\rangle + \mathcal{O}(N^{-1}).$$

Substituting this into (3.43) and invoking Lemma 3.2 completes the proof. ■

3.3.2 Computing the Gateaux derivative of the enriched free energy

The computation of the Gateaux derivative of the enriched free energy (3.30) at a measure $\mu \in \mathcal{M}_+$ in the direction of a probability measure $\nu \in \Pr[-1,1]$,

$$D_\mu \overline{F}_N(t,\mu;\nu) := \lim_{\varepsilon\to 0}\frac{\overline{F}_N(t,\mu+\varepsilon\nu) - \overline{F}_N(t,\mu)}{\varepsilon}, \tag{3.44}$$

is slightly more involved. It will be useful to compute the derivative of the free energy (3.24) with respect to the parameter $s \geqslant 0$ first. Fix a probability measure $\mu \in \Pr[-1,1]$ and a time $t \geqslant 0$. For each $i \leqslant N$, write

$$\widetilde{H}_{N,m}^{s,i}(\sigma) := \sum_{j\neq i}\sum_{k\leqslant\Pi_{j,s}}\log\bigg[(c+\Delta\sigma_j x_{j,k})^{\widetilde{G}_{j,k}^x}\Big(1-\frac{c+\Delta\sigma_j x_{j,k}}{N}\Big)^{1-\widetilde{G}_{j,k}^x}\bigg]$$
$$+\sum_{k\leqslant m}\log\bigg[(c+\Delta\sigma_i x_{i,k})^{\widetilde{G}_{i,k}^x}\Big(1-\frac{c+\Delta\sigma_i x_{i,k}}{N}\Big)^{1-\widetilde{G}_{i,k}^x}\bigg] \tag{3.45}$$

for the Hamiltonian (3.20) conditional on the i'th Poisson sum containing m terms, and denote by

$$Z_{N,m}^{s,i} := \int_{\Sigma_N}\exp\big(H_N^t(\sigma) + \widetilde{H}_{N,m}^{s,i}(\sigma)\big)\,\mathrm{d}P_N^*(\sigma) \tag{3.46}$$

its associated partition function. In this notation, the free energy (3.24) may be expressed as

$$\widetilde{F}_N(t,s,\mu) = \frac{1}{N}\sum_{m\geqslant 0}\pi(sN,m)\mathbb{E}\log Z_{N,m}^{s,i}. \tag{3.47}$$

Lemma 3.4. *For any $t>0$, $s>0$ and $\mu\in\Pr[-1,1]$,*

$$\partial_s\widetilde{F}_N(t,s,\mu) = \mathbb{E}(c+\Delta\langle\sigma_i\rangle x_i)\log(c+\Delta\langle\sigma_i\rangle x_i) - c - \Delta\overline{m}\mathbb{E}x_1 + \mathcal{O}(N^{-1}), \tag{3.48}$$

where the index $i\in\{1,\dots,N\}$ is uniformly sampled and the random variables $(x_i)_{i\geqslant 1}$ are sampled from the measure μ independently of all other sources of randomness.

Proof. Conditioning on the number of terms in each of the Poisson sums that appear in the definition of the free energy (3.24), and leveraging the product rule as well as equations (3.47) and (3.38), one can show that

$$\partial_s\widetilde{F}_N(t,s,\mu) = \frac{1}{N}\sum_{i\leqslant N}\sum_{m\geqslant 0}\partial_s\pi(sN,m)\mathbb{E}\log Z_{N,m}^{s,i} = \sum_{i\leqslant N}\sum_{m\geqslant 0}\pi(sN,m)\mathbb{E}\log\frac{Z_{N,m+1}^{s,i}}{Z_{N,m}^{s,i}}. \tag{3.49}$$

For each $i\leqslant N$, denote by x_i a sample from the measure μ, and write $\widetilde{G}_i^x$ for a random variable with conditional distribution (3.21). These random variables are taken to be independent for different values of $i \leqslant N$, and independent of all other sources of randomness. Since

$$Z_{N,m+1}^{s,i} \overset{d}{=} \int_{\Sigma_N} (c+\Delta\sigma_i x_i)^{\widetilde{G}_i^x}\Big(1-\frac{c+\Delta\sigma_i x_i}{N}\Big)^{1-\widetilde{G}_i^x} \exp\big(H_N^t(\sigma)+\widetilde{H}_{N,m}^{s,i}(\sigma)\big)\,\mathrm{d}P_N^*(\sigma), \tag{3.50}$$

where $\overset{d}{=}$ denotes equality in distribution, it follows by (3.49) and the definition of the Gibbs average in (3.31) that

$$\partial_s \widetilde{F}_N(t,s,\mu) = \sum_{i\leqslant N}\mathbb{E}\log\Big\langle (c+\Delta\sigma_i x_i)^{\widetilde{G}_i^x}\Big(1-\frac{c+\Delta\sigma_i x_i}{N}\Big)^{1-\widetilde{G}_i^x}\Big\rangle. \tag{3.51}$$

Remembering the explicit form of the conditional distribution (3.21) reveals that

$$\partial_s \widetilde{F}_N(t,s,\mu) = \frac{1}{N}\sum_{i\leqslant N}\mathbb{E}(c+\Delta\sigma_i^* x_i)\log\langle c+\Delta\sigma_i x_i\rangle + \sum_{i\leqslant N}\mathbb{E}\Big(1-\frac{c+\Delta\sigma_i^* x_i}{N}\Big)\log\Big\langle 1-\frac{c+\Delta\sigma_i x_i}{N}\Big\rangle.$$

Taylor expanding the logarithm and keeping only first-order terms reduces this to

$$\begin{aligned}\partial_s \widetilde{F}_N(t,s,\mu) &= \frac{1}{N}\sum_{i\leqslant N}\mathbb{E}(c+\Delta\sigma_i^* x_i)\log\langle c+\Delta\sigma_i x_i\rangle - c - \frac{\Delta}{N}\sum_{i\leqslant N}\mathbb{E}x_i\mathbb{E}\langle\sigma_i\rangle + \mathcal{O}(N^{-1})\\ &= \frac{1}{N}\sum_{i\leqslant N}\mathbb{E}(c+\Delta\sigma_i^* x_i)\log(c+\Delta\langle\sigma_i\rangle x_i) - c - \Delta\overline{m}\mathbb{E}x_1 + \mathcal{O}(N^{-1}),\end{aligned}$$

where the second equality uses the Nishimori identity (3.33). Noticing that the Gibbs average $\langle\sigma_i\rangle$ is a measurable function of the data by (3.32), and applying the Nishimori identity (3.33) completes the proof. ■

Before leveraging this result to compute the Gateaux derivative (3.44), it will be convenient to discuss some distributional identities which will simplify the calculation. Fix a finite measure $\mu \in \mathcal{M}_+$ and a probability measure $\nu \in \Pr[-1,1]$. Let $s := \mu[-1,1]$ and fix $\varepsilon > 0$. Introduce the measure $\lambda := \mu + \varepsilon\nu$, and observe that

$$\overline{\lambda} = \frac{\lambda}{s+\varepsilon} = \frac{s}{s+\varepsilon}\overline{\mu} + \frac{\varepsilon}{s+\varepsilon}\nu. \tag{3.52}$$

Denote by $(x_{i,k})_{i,k\geqslant 1}$ i.i.d. random variables sampled from the measure $\overline{\mu}$, and write $(y_{i,k})_{i,k\geqslant 1}$ for i.i.d. random variables sampled from the measure ν. Given i.i.d. random variables $(w_{i,k})_{i,k\geqslant 1}$ with distribution $\mathrm{Ber}(\frac{s}{s+\varepsilon})$, notice that by (3.52) the random variables

$$z_{i,k} := x_{i,k}^{w_{i,k}} y_{i,k}^{1-w_{i,k}} \tag{3.53}$$

are i.i.d. with distribution $\overline{\lambda}$. In particular, if $(\widetilde{G}_{i,k}^z)_{i,k\geqslant 1}$ are independent random variables with conditional distribution (3.21), the Hamiltonian (3.20) may be expressed as

$$\widetilde{H}_N^{s+\varepsilon,\overline{\lambda}}(\sigma) \overset{d}{=} \sum_{i\leqslant N}\sum_{k\leqslant \Pi_{i,s+\varepsilon}}\log\Big[(c+\Delta\sigma_i z_{i,k})^{\widetilde{G}_{i,k}^z}\Big(1-\frac{c+\Delta\sigma_i z_{i,k}}{N}\Big)^{1-\widetilde{G}_{i,k}^z}\Big], \tag{3.54}$$

where $\overset{d}{=}$ denotes equality in distribution. This identity will allow for the linearization of the enriched free energy (3.30) upon realizing that

$$\overline{F}_N(t,\mu+\varepsilon\nu) = \widetilde{F}_N\big(t,s+\varepsilon,\overline{\lambda}\big). \tag{3.55}$$

To make the computation as clear as possible, it will also be convenient to introduce additional notation. In the same spirit as (3.45), for each $i \leqslant N$, write

$$\widetilde{H}^{s,i,+}_{N,m}(\sigma) := \widetilde{H}^{s,i}_{N,m} + \log\left[(c+\Delta\sigma_i y_i)^{\widetilde{G}^y_i}\Big(1-\frac{c+\Delta\sigma_i y_i}{N}\Big)^{1-\widetilde{G}^y_i}\right] \tag{3.56}$$

for the Hamiltonian (3.20) conditional on the i'th Poisson sum containing $m+1$ terms one of which is sampled from the measure ν. Denote by

$$Z^{s,i,+}_{N,m} := \int_{\Sigma_N} \exp\big(H^t_N(\sigma) + \widetilde{H}^{s,i,+}_{N,m}(\sigma)\big)\,\mathrm{d}P^*_N \tag{3.57}$$

its associated partition function. Finally, it will be useful to record the following consequence of Taylor's theorem,

$$\Big(\frac{s}{s+\varepsilon}\Big)^{\sum_{i\leqslant N} m_i} = 1 - \frac{\varepsilon}{s+\varepsilon}\sum_{i\leqslant N} m_i + o(\varepsilon), \tag{3.58}$$

as well as the elementary identity,

$$m\pi(\lambda,m) = \lambda\pi(\lambda,m-1). \tag{3.59}$$

Lemma 3.5. *For any $t>0$, $\mu \in \mathcal{M}_+$ and $\nu \in \Pr[-1,1]$,*

$$D_\mu \overline{F}_N(t,\mu;\nu) = \mathbb{E}(c+\Delta\langle\sigma_i\rangle y_i)\log(c+\Delta\langle\sigma_i\rangle y_i) + N\mathbb{E}\Big(1-\frac{c+\Delta\langle\sigma_i\rangle y_i}{N}\Big)\log\Big(1-\frac{c+\Delta\langle\sigma_i\rangle y_i}{N}\Big) \tag{3.60}$$

where the index $i \in \{1,\dots,N\}$ is uniformly sampled and the random variables $(y_i)_{i\geqslant 1}$ are sampled from the measure ν independently of all other sources of randomness.

Proof. Leveraging (3.54), conditioning on the number of random variables $(w_{i,k})_{i,k\geqslant 1}$ that are equal to one, and using (3.58), one can show that

$$\begin{aligned}\widetilde{F}_N(t,s+\varepsilon,\overline{\lambda}) = \widetilde{F}_N(t,s+\varepsilon,\overline{\mu}) &- \frac{\varepsilon}{s+\varepsilon}\sum_{i\leqslant N}\sum_{m\geqslant 0} m\pi((s+\varepsilon)N,m)\frac{1}{N}\mathbb{E}\log Z^{s+\varepsilon,i}_{n,m}\\ &+ \frac{\varepsilon}{s+\varepsilon}\sum_{i\leqslant N}\sum_{m\geqslant 0}(m+1)\pi((s+\varepsilon)N,m+1)\frac{1}{N}\mathbb{E}\log Z^{s+\varepsilon,i,+}_{N,m} + o(\varepsilon).\end{aligned}$$

Keeping in mind (3.59), this simplifies to

$$\widetilde{F}_N\big(t,s+\varepsilon,\overline{\lambda}\big) = \widetilde{F}_N(t,s+\varepsilon,\overline{\mu}) + \varepsilon\sum_{i\leqslant N}\sum_{m\geqslant 0}\pi\big((s+\varepsilon)N,m\big)\left(\mathbb{E}\log\frac{Z^{s+\varepsilon,i,+}_{N,m}}{Z^{s+\varepsilon,i}_{N,m}} - \mathbb{E}\log\frac{Z^{s+\varepsilon,i}_{N,m+1}}{Z^{s+\varepsilon,i}_{N,m}}\right) + o(\varepsilon). \tag{3.61}$$

For each $i \leqslant N$, denote by x_i a sample from the measure $\overline{\mu}$ and by y_i a sample from the measure ν. Write $\widetilde{G}^x_i$ and $\widetilde{G}^y_i$ for random variables with conditional distribution (3.21). These random variables are taken to be independent for different values of $i \leqslant N$, and independent of all other sources of randomness. Combining (3.61) with (3.50) and the identity

$$Z^{s+\varepsilon,i,+}_{N,m} \overset{d}{=} \int_{\Sigma_N}(c+\Delta\sigma_i y_i)^{\widetilde{G}^y_i}\Big(1-\frac{c+\Delta\sigma_i y_i}{N}\Big)^{1-\widetilde{G}^y_i}\exp\big(H^t_N(\sigma)+\widetilde{H}^{s+\varepsilon,i}_{N,m}(\sigma)\big)\,\mathrm{d}P^*_N(\sigma),$$

where $\overset{d}{=}$ denotes equality in distribution, yields

$$\widetilde{F}_N(t,s+\varepsilon,\overline{\lambda}) = \widetilde{F}_N(t,s+\varepsilon,\overline{\mu}) + \varepsilon\sum_{i\leqslant N}\mathbb{E}\log\left\langle (c+\Delta\sigma_i y_i)^{\widetilde{G}_i^y}\left(1-\frac{c+\Delta\sigma_i y_i}{N}\right)^{1-\widetilde{G}_i^y}\right\rangle \\ - \varepsilon\sum_{i\leqslant N}\mathbb{E}\log\left\langle (c+\Delta\sigma_i x_i)^{\widetilde{G}_i^x}\left(1-\frac{c+\Delta\sigma_i x_i}{N}\right)^{1-\widetilde{G}_i^x}\right\rangle + o(\varepsilon). \quad (3.62)$$

Together with (3.51), this implies that

$$D_\mu \overline{F}_N(t,\mu;\nu) = \lim_{\varepsilon\to 0}\frac{\widetilde{F}_N(t,s+\varepsilon,\overline{\lambda}) - \widetilde{F}_N(t,s+\varepsilon,\overline{\mu})}{\varepsilon} + \partial_s\overline{F}_N(t,s,\overline{\mu}) \\ = \sum_{i\leqslant N}\mathbb{E}\log\left\langle (c+\Delta\sigma_i y_i)^{\widetilde{G}_i^y}\left(1-\frac{c+\Delta\sigma_i y_i}{N}\right)^{1-\widetilde{G}_i^y}\right\rangle.$$

Notice that the Gibbs averages in (3.62) depend on ε, so in taking this limit it has been implicitly used that this dependence is continuous. Proceeding exactly as in the proof of Lemma 3.4 (after display (3.51)) completes the proof. ■

To compare (3.60) with the time derivative of the enriched free energy in Corollary 3.3 it will be useful to once again Taylor expand the logarithm. The comparison will in fact be between the time derivative of the enriched free energy and the density of its Gateaux derivative,

$$D_\mu \overline{F}_N(t,\mu,x) := \mathbb{E}(c+\Delta\langle\sigma_i\rangle x)\log(c+\Delta\langle\sigma_i\rangle x) + N\mathbb{E}\left(1-\frac{c+\Delta\langle\sigma_i\rangle x}{N}\right)\log\left(1-\frac{c+\Delta\langle\sigma_i\rangle x}{N}\right). \quad (3.63)$$

By density it is meant that for every measure $\nu\in\mathcal{M}_+$, the Gateaux derivative (3.44) may be expressed as

$$D_\mu \overline{F}_N(t,\mu;\nu) = \int_{-1}^{1} D_\mu \overline{F}_N(t,\mu,x)\,\mathrm{d}\nu(x). \quad (3.64)$$

Taylor expanding the logarithm shows that

$$D_\mu \overline{F}_N(t,\mu,x) = \mathbb{E}(c+\Delta\langle\sigma_i\rangle x)\log\left(c+\Delta\langle\sigma_i\rangle x\right) - c - \Delta\overline{m}x + \mathcal{O}(N^{-1}). \quad (3.65)$$

Corollary 3.6. *For every $t>0$ and $\mu\in\mathcal{M}_+$,*

$$D_\mu \overline{F}_N(t,\mu,x) = (c+\Delta\overline{m}x)\log(c) + c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\langle R_{[n]}\rangle x^n - c + \mathcal{O}(N^{-1}). \quad (3.66)$$

Proof. Fix $\nu\in\Pr[-1,1]$, and recall that the random variable y_i appearing in Lemma 3.5 is sampled from ν. A Taylor expansion of the logarithm shows that

$$\mathbb{E}(c+\Delta\langle\sigma_i\rangle y_i)\log(c+\Delta\langle\sigma_i\rangle y_i) = (c+\Delta\overline{m}\mathbb{E}y_1)\log(c) - \sum_{n\geqslant 1}\frac{(-\Delta/c)^n}{n}\mathbb{E}(c+\Delta\langle\sigma_i\rangle y_i)y_i^n\langle\sigma_i\rangle^n. \quad (3.67)$$

Since $|\Delta| < c$ and $\mathbb{E}\langle\sigma_i\rangle = \overline{m}$ by the Nishimori identity,

$$\begin{aligned}\sum_{n\geqslant 1}\frac{(-\Delta/c)^n}{n}\mathbb{E}(c+\Delta\langle\sigma_i\rangle y_i)y_i^n\langle\sigma_i\rangle^n &= \sum_{n\geqslant 1}\frac{(-\Delta/c)^n}{n}\big(c\mathbb{E}y_i^n\mathbb{E}\langle\sigma_i\rangle^n+\Delta\mathbb{E}y_i^{n+1}\mathbb{E}\langle\sigma_i\rangle^{n+1}\big)\\ &= -\Delta\mathbb{E}y_i\mathbb{E}\langle\sigma_i\rangle + c\sum_{n\geqslant 2}\bigg(\frac{(-\Delta/c)^n}{n}-\frac{(-\Delta/c)^n}{n-1}\bigg)\mathbb{E}y_i^n\mathbb{E}\langle\sigma_i\rangle^n\\ &= -\Delta\overline{m}\mathbb{E}y_1 - c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\langle R_{[n]}\rangle\mathbb{E}y_i^n.\end{aligned}$$

Substituting this into (3.67) and recalling (3.65) completes the proof. ■

With the time derivative (3.42) and the Gateaux derivative (3.66) at hand, the infinite-dimensional Hamilton-Jacobi equation for the enriched free energy (3.30) may now be derived by identifying a non-linearity that maps the Gateaux derivative to the time derivative, up to a small error that is expected to vanish in the limit of large system size.

3.4 A sparse stochastic block model Hamilton-Jacobi equation

To relate the time derivative (3.42) and the Gateaux derivative (3.66) of the enriched free energy (3.30) it will be convenient to introduce additional notation. Let $g:[-1,1]\to\mathbb{R}$ denote the function defined by

$$g(z) := (c+\Delta z)\big(\log(c+\Delta z)-1\big) = (c+\Delta z)\log(c) + c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}z^n - c, \tag{3.68}$$

and introduce the cone of functions

$$\mathcal{C}_\infty := \bigg\{G_\mu:[-1,1]\to\mathbb{R}\mid G_\mu(x) := \int_{-1}^1 g(xy)\,\mathrm{d}\mu(y)\text{ for some }\mu\in\mathcal{M}_+\bigg\} \tag{3.69}$$

as well as the non-linearity $\mathsf{C}_\infty:\mathcal{C}_\infty\to\mathbb{R}$ given by

$$\mathsf{C}_\infty(G_\mu) := \frac{1}{2}\int_{-1}^1 G_\mu(x)\,\mathrm{d}\mu(x). \tag{3.70}$$

This non-linearity is well-defined by the Fubini-Tonelli theorem. Indeed, if $G_\mu = G_\nu$ for some measures $\mu,\nu\in\mathcal{M}_+$, then

$$\int_{-1}^1 G_\mu(x)\,\mathrm{d}\mu(x) = \int_{-1}^1 G_\nu(x)\,\mathrm{d}\mu(x) = \int_{-1}^1\int_{-1}^1 g(xy)\,\mathrm{d}\mu(x)\,\mathrm{d}\nu(x), \tag{3.71}$$

while

$$\int_{-1}^1 G_\nu(x)\,\mathrm{d}\nu(x) = \int_{-1}^1 G_\mu(x)\,\mathrm{d}\nu(x) = \int_{-1}^1\int_{-1}^1 g(xy)\,\mathrm{d}\mu(y)\,\mathrm{d}\nu(x), \tag{3.72}$$

and the symmetry of the map $(x,y)\mapsto g(xy)$ implies that these two expressions coincide. Corollary 3.6 implies that the Gateaux derivative density (3.63) is close to an element in the cone of functions (3.69). Indeed, if $\mu^* := \mathcal{L}(\langle\sigma_i\rangle)$ denotes the law of the Gibbs average of a uniformly sampled spin coordinate, then (3.66) may be formally written as

$$D_\mu\overline{F}_N(t,\mu,x) \simeq \int_{-1}^1 g(xy)\,\mathrm{d}\mu^*(y) = G_{\mu^*}(x), \tag{3.73}$$

where the Nishimori identity (3.33) has been used to assert that $\mathbb{E}\langle\sigma_i\rangle = \overline{m}$. It follows by another application of the Nishimori identity that

$$\mathsf{C}_\infty\big(D_\mu \overline{F}_N(t,\mu)\big) \simeq \frac{1}{2}\big(c+\Delta\overline{m}^2\big)\log(c) + \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\big(\mathbb{E}\langle R_{[n]}\rangle\big)^2 - \frac{c}{2}. \tag{3.74}$$

Comparing this with the expression in Corollary 3.3, and assuming the approximate concentration of all the multioverlaps,

$$\mathbb{E}\langle R_{[n]}^2\rangle \simeq \big(\mathbb{E}\langle R_{[n]}\rangle\big)^2, \tag{3.75}$$

reveals that, up to a small error, the enriched free energy (3.30) formally satisfies the infinite-dimensional Hamilton-Jacobi equation

$$\partial_t f(t,\mu) = \mathsf{C}_\infty(D_\mu f(t,\mu)) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathcal{M}_+. \tag{3.76}$$

The difficulty in making this informal derivation rigorous is two-fold. On the one hand, infinite-dimensional Hamilton-Jacobi equations of the form (3.76) are not well-studied in the literature. This will be resolved in the next chapter, where the well-posedness theory for Hamilton-Jacobi equations on positive half-space developed in Section 2.4 is leveraged to establish the well-posedness of infinite-dimensional Hamilton-Jacobi equations of the form (3.76). On the other hand, the concentration of the multioverlaps (3.75) is not expected to be valid for each choice of the parameters t and μ. On the positive side, the arguments in [15] reveal that the concentration of the multioverlaps can be enforced through a small perturbation of the Hamiltonian which does not affect the limit of the free energy for most values of the perturbation parameters. Yet, the solution theory for Hamilton-Jacobi equations is rather sensitive to details, and in particular, this control "for most values" or after a suitable local averaging is not sufficient to run the Hamilton-Jacobi approach. The situation is analogous to that encountered in the context of the symmetric rank-one matrix estimation problem when having to control the right side of the approximate Hamilton-Jacobi equation in Proposition 2.27. There, it was possible to control this error term at contact points when the free energy was touched from above — see Lemma 2.29. However, the author is not aware of any way to control this error term at contact points when the free energy is touched from below. This same phenomenon will occur in the present context of the sparse stochastic block model. The inability to control the error term when the free energy is touched from below is the reason why the upper bound in Conjecture 1.4 remains open. Notice that this control is not required in the symmetric rank-one matrix estimation problem due to the lower bound in the approximate Hamilton-Jacobi equation (2.165). This error term, and more specifically the concentration of the multioverlaps (3.41), will be discussed further in Chapters 5 and 6. This chapter closes with the identification of the initial condition associated with the infinite-dimensional Hamilton-Jacobi equation (3.76).

For each integer $N \geqslant 1$, denote by

$$\psi_N(\mu) := \overline{F}_N(0,\mu) = \widetilde{F}_N(0,\mu[-1,1],\overline{\mu}) \tag{3.77}$$

the initial condition associated with the finite-volume enriched free energy (3.30), and notice that the initial condition $\psi:\mathcal{M}_+\to\mathbb{R}$ associated with the infinite-dimensional Hamilton-Jacobi equation (3.76) should be the limit of the initial conditions $(\psi_N)_{N\geqslant 1}$. Following [14], this limit will first be computed for discrete measures $\mu\in\mathcal{M}_+$. A density argument will then be used to show that the convergence extends to all measures in $\mathcal{M}_+$. Given a measure $\mu\in\mathcal{M}_+$, it will be convenient to write $\Pi_\pm(\mu)$ for the Poisson point process with mean measure $(c \pm \Delta x)\,\mathrm{d}\mu(x)$ on $[-1,1]$. For the definition and basic properties of a Poisson point process, see Chapter 5 in [50]. It turns out that the limit of the initial conditions (3.77) is given by an appropriate average with respect to the randomness of the Poisson point processes $\Pi_\pm(\mu)$.

Lemma 3.7. *For any discrete measure $\mu \in \mathcal{M}_+$, the sequence $(\psi_N(\mu))_{N\geqslant 1}$ converges to*

$$\psi(\mu) := -\mu[-1,1]c + p\mathbb{E}\log\int_{\Sigma_1} \exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1) \prod_{x\in\Pi_+(\mu)} (c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)$$
$$+(1-p)\mathbb{E}\log\int_{\Sigma_1} \exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1) \prod_{x\in\Pi_-(\mu)} (c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma), \tag{3.78}$$

where x_1 has law $\overline{\mu}$.

Remark 3.8. From this lemma and its extension to any $\mu \in \mathcal{M}_+$ proved in Proposition 3.12 below, one can also show that for every $\mu \in \mathcal{M}_+$, the density of the Gateaux derivative $D_\mu\psi(\mu)$ is

$$D_\mu\psi(\mu,x) = p\mathbb{E}\langle c+\Delta\sigma x\rangle_+ \log\langle c+\Delta\sigma x\rangle_+ + (1-p)\mathbb{E}\langle c+\Delta\sigma x\rangle_- \log\langle c+\Delta\sigma x\rangle_- - c - \Delta\overline{m}x, \tag{3.79}$$

where $\langle\cdot\rangle_\pm$ denote the Gibbs averages given by

$$\langle f(\sigma)\rangle_\pm := \frac{\int_{\Sigma_1} f(\sigma)\exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1)\prod_{x\in\Pi_\pm(\mu)}(c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)}{\int_{\Sigma_1} \exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1)\prod_{x\in\Pi_\pm(\mu)}(c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)}. \tag{3.80}$$

Proof of Lemma 3.7. Since $\mu \in \mathcal{M}_+$ is a discrete measure, it may be expressed as

$$\mu := \sum_{\ell\leqslant K} p_\ell \delta_{a_\ell}$$

for some integer $K \geqslant 1$, some atoms $a_\ell \in [-1,1]$, and some weights $p_\ell \geqslant 0$. Let $s := \mu[-1,1]$, and introduce independent Poisson random variables $\Pi_{i,s} \sim \mathrm{Poi}(sN)$ in such a way that

$$\psi_N(\mu) = \frac{1}{N}\sum_{i\leqslant N}\mathbb{E}\log\int_{\Sigma_1}\exp\sum_{k\leqslant\Pi_{i,s}}\log\bigg[(c+\Delta\sigma x_{i,k})^{\widetilde{G}^x_{i,k}}\Big(1-\frac{c+\Delta\sigma x_{i,k}}{N}\Big)^{1-\widetilde{G}^x_{i,k}}\bigg]\,\mathrm{d}P^*(\sigma),$$

where $(x_{i,k})_{i,k\geqslant 1}$ are i.i.d. random variables with law $\overline{\mu}$. Since each of the expectations in this average is the same,

$$\psi_N(\mu) = \mathbb{E}\log\int_{\Sigma_1}\exp\sum_{k\leqslant\Pi_{1,s}}\log\bigg[(c+\Delta\sigma x_k)^{\widetilde{G}^x_k}\Big(1-\frac{c+\Delta\sigma x_k}{N}\Big)^{1-\widetilde{G}^x_k}\bigg]\,\mathrm{d}P^*(\sigma), \tag{3.81}$$

where $(x_k)_{k\geqslant 1}$ are i.i.d. random variables with law $\overline{\mu}$ and $\widetilde{G}^x_k$ has conditional distribution (3.21) for $i = 1$ and $x_{1,k}$ replaced by x_k. To simplify this further, introduce the random index sets

$$\mathcal{I}_0 := \big\{k \leqslant \Pi_{1,s} \mid \widetilde{G}^x_k = 0\big\} \quad \text{and} \quad \mathcal{I}_1 := \big\{k \leqslant \Pi_{1,s} \mid \widetilde{G}^x_k = 1\big\}.$$

Decomposing the sum in (3.81) according to the partition $\{k \leqslant \Pi_{1,s}\} = \mathcal{I}_0 \sqcup \mathcal{I}_1$, and applying Taylor's theorem to the logarithm reveals that

$$\psi_N(\mu) = \mathbb{E}\log\int_{\Sigma_1}\prod_{k\in\mathcal{I}_1}(c+\Delta\sigma x_k)\exp\Big(-\frac{\Delta\sigma}{N}\sum_{k\in\mathcal{I}_0}x_k\Big)\,\mathrm{d}P^*(\sigma) - \Big(\frac{c}{N}+\mathcal{O}\big(N^{-2}\big)\Big)\mathbb{E}|\mathcal{I}_0|.$$

Conditionally on σ^*, $\Pi_{1,s}$ and $(x_k)_{k\geqslant 1}$, the random variable $|\mathcal{I}_0|$ is a sum of Bernoulli random variables with probability of success $1-\frac{c+\Delta\sigma_1^* x_k}{N}$. It therefore has mean

$$\mathbb{E}|\mathcal{I}_0| = \mathbb{E}\Pi_{1,s}\mathbb{E}\Big(1-\frac{c+\Delta\sigma_1^* x_1}{N}\Big) = sN\Big(1-\frac{c+\Delta\overline{m}\mathbb{E}x_1}{N}\Big) \tag{3.82}$$

and variance bounded by

$$\mathrm{Var}|\mathcal{I}_0| = \mathbb{E}\Pi_{1,s}\mathbb{E}\Big(1-\frac{c+\Delta\sigma_1^* x_1}{N}\Big)\Big(\frac{c+\Delta\sigma_1^* x_1}{N}\Big) \leqslant s(c+|\Delta|). \tag{3.83}$$

Using (3.82) and introducing the random index sets $\mathcal{I}_1^\ell := \{k\in\mathcal{I}_1 \mid x_k = a_\ell\}$ reveals that

$$\psi_N(\mu) = \mathbb{E}\log\int_{\Sigma_1}\prod_{\ell\leqslant K}\prod_{k\in\mathcal{I}_1^\ell}(c+\Delta\sigma a_\ell)\exp\Big(-\frac{\Delta\sigma}{N}\sum_{k\in\mathcal{I}_0}x_k\Big)\,\mathrm{d}P^*(\sigma) - cs + \mathcal{O}\big(N^{-1}\big).$$

Observe that for any $\sigma\in\Sigma_1$,

$$\mathbb{E}\Big|-\frac{\Delta\sigma}{N}\sum_{k\in\mathcal{I}_0}x_k + \Delta\sigma s\mathbb{E}x_1\Big| \leqslant |\Delta|s\mathbb{E}\Big|\frac{1}{Ns}\sum_{k\in\mathcal{I}_0}x_k - \mathbb{E}x_1\Big| \leqslant |\Delta|s\mathbb{E}\Big|\frac{1}{Ns}\sum_{k\leqslant Ns}x_k - \mathbb{E}x_1\Big| + \frac{\Delta}{N}\big(\mathrm{Var}|\mathcal{I}_0| + |\mathbb{E}|\mathcal{I}_0| - Ns|\big)$$

where the fact that $|x_k|\leqslant 1$ and Jensen's inequality have been used in the second inequality. Recalling (3.83) and invoking the strong law of large numbers shows that

$$\psi_N(\mu) = \mathbb{E}\log\int_{\Sigma_1}\exp(-\Delta\sigma s\mathbb{E}x_1)\prod_{\ell\leqslant K}\prod_{k\in\mathcal{I}_1^\ell}(c+\Delta\sigma a_\ell)\,\mathrm{d}P^*(\sigma) - cs + o(1). \tag{3.84}$$

The Poisson colouring theorem (Proposition A.16) implies that $|\mathcal{I}_1^\ell|$ is a Poisson random variable with mean

$$\mathbb{E}\Pi_{1,s}\cdot\mathbb{P}\{\widetilde{G}_1^x = 1, x_1 = a_\ell\} = sN\cdot\frac{c+\Delta\sigma_1^* a_\ell}{N}\cdot\overline{\mu}(a_\ell) = \big(c+\Delta\sigma_1^* a_\ell\big)\mu(a_\ell),$$

so averaging (3.84) over the randomness of σ^* yields

$$\begin{aligned}\psi_N(\mu) = -cs + p\mathbb{E}\log\int_{\Sigma_1}\exp(-\Delta\sigma s\mathbb{E}x_1)\prod_{x\in\Pi_+(\mu)}(c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)\\ + (1-p)\mathbb{E}\log\int_{\Sigma_1}\exp(-\Delta\sigma s\mathbb{E}x_1)\prod_{x\in\Pi_-(\mu)}(c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma) + o(1).\end{aligned}$$

This completes the proof. ■

To extend this convergence to all measures in $\mathcal{M}_+$, the continuity of the functional (3.78) with respect to the Wasserstein distance on the space of probability measures will be used. The *Wasserstein distance* between two probability measure $\mathbb{P},\mathbb{Q}\in\mathrm{Pr}[-1,1]$ is defined by

$$W(\mathbb{P},\mathbb{Q}) := \sup\Big\{\Big|\int_{-1}^1 h(x)\,\mathrm{d}\mathbb{P}(x) - \int_{-1}^1 h(x)\,\mathrm{d}\mathbb{Q}(x)\Big| \mid h:[-1,1]\to\mathbb{R} \text{ is Lipschitz with } \|h\|_{\mathrm{Lip}}\leqslant 1\Big\} \tag{3.85}$$

$$= \inf\Big\{\int_{[-1,1]^2}|x-y|\,\mathrm{d}\nu(x,y) \mid \nu\in\mathrm{Pr}\big([-1,1]^2\big) \text{ has marginals } \mathbb{P} \text{ and } \mathbb{Q}\Big\}, \tag{3.86}$$

where $\|\cdot\|_{\mathrm{Lip}}$ denotes the Lipschitz semi-norm defined in (2.90). The equality of the representations (3.85) and (3.86) is guaranteed by the Kantorovich-Rubinstein theorem (Theorem 4.15 in [90]). This continuity will be obtained as a consequence of the following uniform bound on the spatial derivatives of the Gateaux derivative density (3.63).

Lemma 3.9. *For every N large enough (relative to c), $\mu \in \mathcal{M}_+$, $t \geqslant 0$ and $x \in [-1,1]$,*

$$\big|D_\mu \overline{F}_N(t,\mu,x)\big| \leqslant 2c\big(2+|\log(2c)|+|\log(c-|\Delta|)|\big), \tag{3.87}$$

$$\big|\partial_x D_\mu \overline{F}_N(t,\mu,x)\big| \leqslant c\big(1+|\log(2c)|+|\log(c-|\Delta|)|\big). \tag{3.88}$$

Proof. Recall from (3.63) that

$$D_\mu \overline{F}_N(t,\mu,x) = \mathbb{E}\big(c+\Delta\langle\sigma_i\rangle x\big)\log\big(c+\Delta\langle\sigma_i\rangle x\big) + N\mathbb{E}\Big(1-\frac{c+\Delta\langle\sigma_i\rangle x}{N}\Big)\log\Big(1-\frac{c+\Delta\langle\sigma_i\rangle x}{N}\Big).$$

It follows by a direct computation that

$$\partial_x D_\mu \overline{F}_N(t,\mu,x) = \Delta\mathbb{E}\langle\sigma_i\rangle\log\big(c+\Delta\langle\sigma_i\rangle x\big) - \Delta\mathbb{E}\langle\sigma_i\rangle\log\Big(1-\frac{c+\Delta\langle\sigma_i\rangle x}{N}\Big).$$

Since all spin configuration coordinates are bounded by one and $|\Delta|<c$, Taylor's theorem implies that for N large enough,

$$\big|D_\mu \overline{F}_N(t,\mu,x)\big| \leqslant 2c\big(2+|\log(2c)|+|\log(c-|\Delta|)|\big),$$
$$\big|\partial_x D_\mu \overline{F}_N(t,\mu,x)\big| \leqslant c\big(1+|\log(2c)|+|\log(c-|\Delta|)|\big).$$

Notice that the choice of N only depends on c as $x\in[-1,1]$ and $|\Delta|<c$. This completes the proof. ∎

Lemma 3.10. *The initial condition ψ_N satisfies the Lipschitz bound*

$$|\psi_N(\mathbb{P})-\psi_N(\mathbb{Q})| \leqslant c\big(1+|\log(2c)|+|\log(c-|\Delta|)|\big)W(\mathbb{P},\mathbb{Q}) \tag{3.89}$$

for all probability measures $\mathbb{P},\mathbb{Q}\in \Pr[-1,1]$.

Proof. The fundamental theorem of calculus and the definition of the Gateaux derivative in (3.44) imply that

$$\psi_N(\mathbb{P})-\psi_N(\mathbb{Q}) = \int_0^1 \frac{\mathrm{d}}{\mathrm{d}t}\psi_N\big(\mathbb{Q}+t(\mathbb{P}-\mathbb{Q})\big)\,\mathrm{d}t = \int_0^1 D_\mu\psi_N\big(\mathbb{Q}+t(\mathbb{P}-\mathbb{Q});\mathbb{P}-\mathbb{Q}\big)\,\mathrm{d}t.$$

Since the Gateaux derivative of the initial condition admits a continuously differentiable density,

$$|\psi_N(\mathbb{P})-\psi_N(\mathbb{Q})| \leqslant \int_0^1 \Big|\int_{-1}^1 f_t(x)\,\mathrm{d}\mathbb{P}(x) - \int_{-1}^1 f_t(x)\,\mathrm{d}\mathbb{Q}(x)\Big|\,\mathrm{d}t$$

for the continuously differentiable function $f_t(x) := D_\mu\psi_N(\mathbb{Q}+t(\mathbb{P}-\mathbb{Q}),x)$. The mean value theorem and (3.88) reveal that $\|f_t\|_{\mathrm{Lip}} \leqslant c\big(1+|\log(2c)|+|\log(c-|\Delta|)|\big)$. It follows by definition of the Wasserstein distance (3.85) that

$$|\psi_N(\mu)-\psi_N(\nu)| \leqslant c\big(1+|\log(2c)|+|\log(c-|\Delta|)|\big)W(\mathbb{P},\mathbb{Q}).$$

This completes the proof. ∎

Lemma 3.11. *The functional $\psi : \mathcal{M}_+ \to \mathbb{R}$ defined by (3.78) is continuous with respect to the weak convergence of measures. This means that for any sequence of measures $(\mu_n)_{n\geqslant 1} \subseteq \mathcal{M}_+$ converging weakly to a measure $\mu \in \mathcal{M}_+$,*

$$\lim_{N\to+\infty} \psi(\mu_n) = \psi(\mu). \tag{3.90}$$

Proof. To alleviate the exposition, the continuity of the functional

$$\psi^1(\mu) := \mathbb{E}\log \int_{\Sigma_1} \exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1) \prod_{x\in\Pi_+(\mu)} (c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)$$

with respect to the weak convergence of measures will be proved instead. Up to an additive constant, the asymptotic initial condition $\psi(\mu)$ is the weighted average of $\psi^1(\mu)$ and another functional of the same form whose continuity can be established using an identical argument, so this suffices. For each measure $\mu \in \mathcal{M}_+$ introduce the Hamiltonian

$$H(\sigma,\mu) := -\Delta\sigma \int_{-1}^{1} x\,\mathrm{d}\mu(x) + \sum_{x\in\Pi_+(\mu)} (c+\Delta\sigma x)$$

in such a way that the asymptotic initial condition is its associated free energy,

$$\psi^1(\mu) = \mathbb{E}\log \int_{\Sigma_1} \exp H(\sigma,\mu)\,\mathrm{d}P^*(\sigma).$$

Consider a sequence of measures $(\mu_n)_{n\geqslant 1} \subseteq \mathcal{M}_+$ converging weakly to a measure $\mu \in \mathcal{M}_+$, and let Π_n and Π be independent Poisson random variables with means $\mu_n[-1,1]$ and $\mu[-1,1]$, respectively. Introduce a collection $(X_k^n, X_k)_{k\geqslant 1}$ of i.i.d. random vectors with joint law $\nu \in \Pr\big([-1,1]^2\big)$ having marginals $\overline{\mu}_n$ and $\overline{\mu}$. In this way, the coordinates $(X_k^n)_{k\geqslant 1}$ are i.i.d. with law $\overline{\mu}_n$, the coordinates $(X_k)_{k\geqslant 1}$ are i.i.d. with law $\overline{\mu}$, and

$$\sum_{x\in\Pi_+(\mu_n)} (c+\Delta\sigma x) \overset{d}{=} \sum_{k\leqslant \Pi_n} \big(c+\Delta\sigma X_k^n\big) \quad \text{and} \quad \sum_{x\in\Pi_+(\mu)} (c+\Delta x) \overset{d}{=} \sum_{k\leqslant \Pi} \big(c+\Delta\sigma X_k\big),$$

where $\overset{d}{=}$ denotes equality in distribution. It follows that for any $\sigma \in \Sigma_1$,

$$|H(\sigma,\mu_n) - H(\sigma,\mu)| \leqslant |\Delta| \Big| \int_{-1}^{1} x\,\mathrm{d}\mu_n(x) - \int_{-1}^{1} x\,\mathrm{d}\mu(x) \Big| + c|\Pi_n - \Pi| + |\Delta| \Big| \sum_{k\leqslant\Pi_n} X_k^n - \sum_{k\leqslant\Pi} X_k \Big|,$$

and therefore,

$$\big|\psi^1(\mu_n) - \psi^1(\mu)\big| \leqslant |\Delta| \Big| \int_{-1}^{1} x\,\mathrm{d}\mu_n(x) - \int_{-1}^{1} x\,\mathrm{d}\mu(x) \Big| + c\mathbb{E}|\Pi_n - \Pi| + |\Delta|\mathbb{E} \Big| \sum_{k\leqslant\Pi_n} X_k^n - \sum_{k\leqslant\Pi} X_k \Big|.$$

To simplify this further, define the random variable $\Pi_n' := \min(\Pi_n, \Pi)$, introduce a Poisson random variable Π_n'' independent of all other sources of randomness with mean $|\mu_n[-1,1] - \mu[-1,1]|$, and define the collection of random variables $(Z_k^n)_{k\geqslant 1}$ by

$$Z_k^n := \begin{cases} X_k^n & \text{if } \Pi_n' = \Pi \\ X_k & \text{otherwise} \end{cases}.$$

The basic properties of Poisson random variables and the fact that $|Z_k^n| \leqslant 1$ imply that

$$\begin{aligned}\left|\psi^1(\mu_n)-\psi^1(\mu)\right| &\leqslant c|\Delta|\left|\int_{-1}^{1} x\,\mathrm{d}(\mu_n-\mu)(x)\right| + c\mathbb{E}\Pi_n'' + |\Delta|\mathbb{E}\sum_{k\leqslant\Pi_n''}\left|Z_k^n\right| + |\Delta|\mathbb{E}\sum_{k\leqslant\Pi_n'}\left|X_k^n - X_k\right| \\ &\leqslant c\bigg(\left|\int_{-1}^{1} x\,\mathrm{d}(\mu_n-\mu)(x)\right| + 2\mathbb{E}\Pi_n'' + \mathbb{E}\Pi_n'\int_{[-1,1]^2}|x-y|\,\mathrm{d}\nu(x,y)|\bigg).\end{aligned}$$

Taking the infimum over all couplings $\nu \in \Pr\big([-1,1]^2\big)$ with marginals $\overline{\mu}_n$ and $\overline{\mu}$ reveals that for n large enough,

$$\left|\psi^1(\mu_n)-\psi^1(\mu)\right| \leqslant c\bigg(\left|\int_{-1}^{1} x\,\mathrm{d}(\mu_n-\mu)(x)\right| + 2\left|\mu_n[-1,1]-\mu[-1,1]\right| + 3\mu[-1,1]W\big(\overline{\mu}_n,\overline{\mu}\big)\bigg),$$

where it has been used that $\mathbb{E}\Pi_n' \leqslant \mathbb{E}\Pi + \mathbb{E}\Pi_n \leqslant 3\mu[-1,1]$ for n large enough as μ_n converges weakly to μ. Letting n tend to infinity and recalling that the Wasserstein distance (3.86) metrizes the weak convergence of probability measures completes the proof. ■

Proposition 3.12. *For any measure $\mu \in \mathcal{M}_+$, the sequence $(\psi_N(\mu))_{N\geqslant 1}$ converges to $\psi(\mu)$ defined in* (3.78).

Proof. Consider a sequence $(\mu_n)_{n\geqslant 1}$ of discrete measures such that $\mu_n[-1,1] = \mu[-1,1]$ for all $n \geqslant 1$ and $\overline{\mu}_n \to \overline{\mu}$ with respect to the Wasserstein distance (3.85). By the triangle inequality and Theorem 3.10,

$$\begin{aligned}\left|\psi(\mu)-\psi_N(\mu)\right| &\leqslant \left|\psi(\mu)-\psi(\mu_n)\right| + \left|\psi(\mu_n)-\psi_N(\mu_n)\right| + \left|\psi_N(\mu_n)-\psi_N(\mu)\right| \\ &\leqslant \left|\psi(\mu)-\psi(\mu_n)\right| + \left|\psi(\mu_n)-\psi_N(\mu_n)\right| + c\big(1+|\log(2c)|+|\log(c-|\Delta|)|\big)W\big(\overline{\mu}_n,\overline{\mu}\big),\end{aligned}$$

where the choice that $\mu_n[-1,1] = \mu[-1,1]$ has played its part. Invoking Lemmas 3.7 and 3.11 to let N tend to infinity and then n tend to infinity completes the proof. ■

This result identifies the initial condition for the Hamilton-Jacobi equation (3.76), and suggests that the limit of the enriched free energy (3.30) should be the solution, in a sense to be made precise, to the infinite-dimensional Hamilton-Jacobi equation

$$\partial_t f(t,\mu) = \mathsf{C}_\infty(D_\mu f(t,\mu)) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathcal{M}_+ \tag{3.91}$$

subject to the initial condition $\psi : \mathcal{M}_+ \to \mathbb{R}$ defined by

$$\begin{aligned}\psi(\mu) := -\mu[-1,1]c + p\mathbb{E}\log\int_{\Sigma_1}\exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1)\prod_{x\in\Pi_+(\mu)}(c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma) \\ + (1-p)\mathbb{E}\log\int_{\Sigma_1}\exp(-\mu[-1,1]\Delta\sigma\mathbb{E}x_1)\prod_{x\in\Pi_-(\mu)}(c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma),\end{aligned} \tag{3.92}$$

where x_1 has law $\overline{\mu}$. Together with (3.17) and the fact that the free energy (3.14) in the sparse stochastic block model can be obtained by evaluating the enriched free energy (3.30) at the point $(1,0)$, that is $\overline{F}_N = \overline{F}_N(1,0)$, this leads to Conjecture 1.4. Leveraging the well-posedness theory for Hamilton-Jacobi equations on positive half-space discussed in Section 2.4, a well-posedness theory for the infinite-dimensional Hamilton-Jacobi equation (3.91) will be developed in the next chapter.

Chapter 4

Well-posedness of infinite-dimensional Hamilton-Jacobi equations

In this chapter, the well-posedness of the infinite-dimensional Hamilton-Jacobi equation (3.91) is established. Previous attempts to apply the Hamilton-Jacobi approach to study mean-field disordered systems had either led to finite-dimensional Hamilton-Jacobi equations posed on closed convex cones [27, 30, 34, 36, 84, 85], in the context of statistical inference problems, or to infinite-dimensional Hamilton-Jacobi equations of transport type [83, 86, 87], in the context of spin-glass models. A general well-posedness theory for the former was established in [32] while one for the latter was developed in [33]. In Chapter 3, an infinite-dimensional Hamilton-Jacobi equation posed over a space of probability measures, but featuring derivatives of "affine" rather than transport type, was proposed to describe the asymptotic mutual information in the sparse stochastic block model. To strive for generality, as opposed to focusing exclusively on the infinite-dimensional Hamilton-Jacobi equation (3.91), a broader class of equations that is expected to appear in other mean-field problems with sparse interactions is considered. Recall the notation $\mathcal{M}_s$ and $\mathcal{M}_+$ introduced in (3.26) and (3.27) for the space of signed measures on $[-1,1]$ and the cone of non-negative measures on this interval. Fix a continuously differentiable function $g:[-1,1]\to\mathbb{R}$, and for each measure $\mu\in\mathcal{M}_+$, define the function $G_\mu:[-1,1]\to\mathbb{R}$ by

$$G_\mu(x) := \int_{-1}^{1} g(xy)\,\mathrm{d}\mu(y). \tag{4.1}$$

Introduce the cone of functions

$$\mathcal{C}_\infty := \left\{G_\mu \mid \mu\in\mathcal{M}_+\right\} \tag{4.2}$$

as well as the non-linearity $\mathsf{C}_\infty:\mathcal{C}_\infty\to\mathbb{R}$ defined on this cone by

$$\mathsf{C}_\infty(G_\mu) := \frac{1}{2}\int_{-1}^{1} G_\mu(x)\,\mathrm{d}\mu(x) = \frac{1}{2}\int_{-1}^{1}\int_{-1}^{1} g(xy)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(x). \tag{4.3}$$

This non-linearity is well-defined by the Fubini-Tonelli theorem (see equations (3.71)-(3.72)). Given a function $f:\mathbb{R}_{\geqslant 0}\times\mathcal{M}_+\to\mathbb{R}$ and measures $\mu,\nu\in\mathcal{M}_+$, denote by $D_\mu f(t,\mu;\nu)$ the Gateaux derivative of the function $f(t,\cdot)$ at the measure μ in the direction ν,

$$D_\mu f(t,\mu;\nu) := \lim_{\varepsilon\to 0}\frac{f(t,\mu+\varepsilon\nu)-f(t,\mu)}{\varepsilon}. \tag{4.4}$$

As in Section 3.3, the Gateaux derivative of $f(t,\cdot)$ at the measure $\mu \in \mathcal{M}_+$ is said to admit a density if there exists a bounded measurable function $x \mapsto D_\mu f(t,\mu,x)$ defined on the interval $[-1,1]$ such that, for every measure $\nu \in \mathcal{M}_+$,

$$D_\mu f(t,\mu;\nu) = \int_{-1}^{1} D_\mu f(t,\mu,x)\,\mathrm{d}\nu(x). \tag{4.5}$$

Abusing notation, the density $D_\mu f(t,\mu,\cdot)$ will often be identified with the Gateaux derivative $D_\mu f(t,\mu)$. The purpose of this chapter is to establish the well-posedness of the infinite-dimensional Hamilton-Jacobi equation

$$\partial_t f(t,\mu) = \mathsf{C}_\infty\big(D_\mu f(t,\mu)\big) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathcal{M}_+ \tag{4.6}$$

subject to some initial condition $f(0,\cdot) = \psi(\cdot)$ under appropriate assumptions on the kernel g and the initial condition $\psi : \mathcal{M}_+ \to \mathbb{R}$. In particular, these assumptions will imply that, in a suitably weak sense, for all $t \geq 0$ and $\mu \in \mathcal{M}_+$, the Gateaux derivative of the solution $D_\mu f(t,\mu)$ belongs to the cone $\mathcal{C}_\infty$. The sparse stochastic block model will correspond to the choice of g in (3.68) and of initial condition ψ in (3.92).

In Section 4.1, the infinite-dimensional Hamilton-Jacobi equation (4.6) is projected onto a positive half-space with dimension monotone in some integer parameter $K \geqslant 1$ to obtain a family of approximating finite-dimensional Hamilton-Jacobi equations. The expectation is that the sequence of solutions to the approximating equations converges to the solution of the infinite-dimensional equation (4.6) upon letting this parameter K tend to infinity. The assumptions on the kernel g and initial condition ψ taken throughout this chapter are also stated in this section, with one of these assumptions imposing a constraint on the gradient of the projected initial conditions. In Section 4.2, the well-posedness of the infinite-dimensional Hamilton-Jacobi equation (4.6) is established. This is done by extending the non-linearity of each of the approximating Hamilton-Jacobi equations from the projection of the cone (4.2) where it is naturally defined to all of Euclidean space, establishing the well-posedness of this extended finite-dimensional approximating Hamilton-Jacobi equation using the techniques developed in Section 2.4, and finally showing that the sequence of solutions to each of these Hamilton-Jacobi equations converges when the parameter K is sent to infinity. The solution to the infinite-dimensional Hamilton-Jacobi equation (4.6) is defined to be the limit of these approximating solutions. In Section 4.3, under appropriate convexity assumptions, a Hopf-Lax variational formula is established for the infinite-dimensional Hamilton-Jacobi equation (4.6) by finding a Hopf-Lax variational formula for the solution to each of the approximating equations, and sending the parameter K to infinity. It will be convenient to first establish all these results under a positivity assumption on the kernel g. In Section 4.4 this assumption will be lifted at the cost of strengthening the assumption on the gradient of the projected initial conditions, and the well-posedness results previously discussed will be extended to this setting.

It may seem surprising to the reader that a well-posedness theory for the infinite-dimensional Hamilton-Jacobi equation (4.6), or at the very least the equation (3.91) appearing in the context of the sparse stochastic block model, cannot be found in the literature. The study of equations posed on infinite-dimensional Banach spaces was initiated in [40, 41, 42]. The assumptions imposed on the Banach space preclude the possibility of applying the results presented there to the space of bounded measures. There, the existence of solutions is obtained via a connection with differential games. An example is also given in which solutions to natural finite-dimensional approximations fail to converge to the solution of the infinite-dimensional equation. This phenomenon is not expected to occur for the equation considered in this chapter, and in any case, the definition of solution as the limit of finite-dimensional approximations is the one used when studying the sparse stochastic block model in Chapter 6. Moreover, for the equations of transport type appearing in the context of mean-field spin glasses, it was shown in [32] that finite-dimensional approximations do converge to the intrinsic viscosity solution of the infinite-dimensional equation. Equations that are posed over a space of probability measures, or more general metric spaces, have been considered in several works including [9, 24, 25, 26, 54, 55, 58, 59]. These works revolve around equations involving derivatives of transport type for probability measures over $\mathbb{R}^d$. Since transportation of mass over $\mathbb{R}^d$ can be carried without limit, questions of boundary conditions do not arise there, unlike in the more recent works [32, 83, 86] in which probability measures over $\mathbb{R}^d_{\geqslant 0}$ or the space of non-negative definite matrices are considered. The author is not aware of previous works considering equations that involve derivatives of "affine" type, as is done here. In this context, the natural "movements" are different from those appearing for the transport geometry, and the additional constraint that non-negative measures must be considered introduces boundary issues.

The well-posedness results for the infinite-dimensional Hamilton-Jacobi equation (4.6) required to study the sparse stochastic block model are all stated in Section 4.1. The reader eager to return to the study of the sparse stochastic block model may therefore consider reading only Section 4.1, taking the results stated there for granted and skipping the rest of this chapter on first reading. This chapter parallels and draws heavily upon Section 2.4 of Chapter 2, and its contents are taken from [48].

4.1 Projections, assumptions, and key well-posedness results

To study the Hamilton-Jacobi equation (4.6), it will be projected from the infinite-dimensional space of measures $\mathcal{M}_+$ to a family of finite-dimensional spaces of measures $\mathcal{M}_+^{(K)}$ with dimension monotone in some integer parameter $K \geqslant 1$. In previous works [83, 86, 87], derivatives of transport type were the primary focus of investigation, and it was thus natural to discretize the space of measures by restricting to measures of the form $K^{-1}\sum_{k=1}^K \delta_{x_k}$, only allowing the x_k's to vary but keeping the weight of each atom fixed. Due to the nature of the derivatives appearing in (4.6), here it will be convenient to define the finite-dimensional approximating space as the cone of non-negative measures supported on dyadic rationals in the interval $[-1,1]$. That is, the weights will be allowed to vary, provided that they remain non-negative, but positions of the atoms will be kept fixed. Given an integer $K \geqslant 1$, write

$$\mathcal{D}_K := \left\{k = \frac{i}{2^K} \mid -2^K \leqslant i < 2^K\right\} \tag{4.7}$$

for the set of dyadic rationals on $[-1,1]$ at scale K. It will be convenient to index vectors using the set of dyadic rationals, writing $x = (x_k)_{k\in\mathcal{D}_K} \in \mathbb{R}^{\mathcal{D}_K}$. The set of discrete measures supported on the dyadic rationals at scale K in the interval $[-1,1]$ is denoted by

$$\mathcal{M}_+^{(K)} := \left\{\mu \in \mathcal{M}_+ \mid \mu = \frac{1}{|\mathcal{D}_K|}\sum_{k\in\mathcal{D}_K} x_k\delta_k \text{ for some } x = (x_k)_{k\in\mathcal{D}_K} \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}\right\}. \tag{4.8}$$

A natural way to project a general measure $\mu \in \mathcal{M}_+$ onto $\mathcal{M}_+^{(K)}$ is via the mapping

$$x^{(K)}(\mu) := \left(|\mathcal{D}_K|\mu\left[k, k+2^{-K}\right)\right)_{k\in\mathcal{D}_K} \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}. \tag{4.9}$$

For $\mu \in \mathcal{M}_+^{(K)}$, the image of μ is simply the sequence of weights of the measure μ at each point in $\mathcal{D}_K$, up to multiplication by $|\mathcal{D}_K|$. The inverse of this mapping assigns to each $x \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$ the measure

$$\mu_x^{(K)} := \frac{1}{|\mathcal{D}_K|}\sum_{k\in\mathcal{D}_K} x_k\delta_k \in \mathcal{M}_+^{(K)}. \tag{4.10}$$

These projections can be used to devise finite-dimensional approximations to the Hamilton-Jacobi equation (4.6). These will be posed on the cone $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$. Indeed, any real-valued function $f:\mathbb{R}_{\geqslant 0}\times\mathcal{M}_+^{(K)}\to\mathbb{R}$ may be identified with the function

$$f^{(K)}(t,x) := f\big(t,\mu_x^{(K)}\big) \tag{4.11}$$

defined on $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$. Moreover, the Gateaux derivative at the measure $\mu\in\mathcal{M}_+^{(K)}$ may be identified with the gradient $|\mathcal{D}_K|\nabla f^{(K)}(t,x^{(K)}(\mu))$ by duality. Indeed, for any direction $\nu\in\mathcal{M}_+^{(K)}$,

$$D_\mu f(t,\mu;\nu) = \frac{\mathrm{d}}{\mathrm{d}\varepsilon}\Big|_{\varepsilon=0} f^{(K)}(t,x^{(K)}(\mu)+\varepsilon x^{(K)}(\nu)) = \nabla f^{(K)}(t,x^{(K)}(\mu))\cdot x^{(K)}(\nu). \tag{4.12}$$

The additional factor of $|\mathcal{D}_K|$ appears because $x^{(K)}(\nu)$ has ℓ^1-norm $|\mathcal{D}_K|$ whenever ν is a probability measure. The corresponding initial condition becomes the function $\psi^{(K)}:\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}\to\mathbb{R}$ defined by

$$\psi^{(K)}(x) := \psi\big(\mu_x^{(K)}\big). \tag{4.13}$$

The cone (4.2) and the non-linearity (4.3) may be projected similarly. Introduce the symmetric matrix

$$G^{(K)} := \frac{1}{|\mathcal{D}_K|^2}\big(g(kk')\big)_{k,k'\in\mathcal{D}_K}\in\mathbb{R}^{\mathcal{D}_K\times\mathcal{D}_K}, \tag{4.14}$$

and observe that for every $\mu\in\mathcal{M}_+^{(K)}$ and $k\in\mathcal{D}_K$,

$$G_\mu(k) = \sum_{k'\in\mathcal{D}_K} g(kk')\mu(k') = \frac{1}{|\mathcal{D}_K|}\sum_{k'\in\mathcal{D}_K} g(kk')x^{(K)}(\mu)_{k'} = |\mathcal{D}_K|\big(G^{(K)}x^{(K)}(\mu)\big)_k. \tag{4.15}$$

This motivates the definition of the projected cone,

$$\mathcal{C}_K := \Big\{G^{(K)}x^{(K)}(\mu)\in\mathbb{R}^{\mathcal{D}_K}\mid \mu\in\mathcal{M}_+^{(K)}\Big\} = \Big\{G^{(K)}x\in\mathbb{R}^{\mathcal{D}_K}\mid x\in\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}\Big\}, \tag{4.16}$$

and the projected non-linearity $\mathsf{C}_K:\mathcal{C}_K\to\mathbb{R}$ given by

$$\mathsf{C}_K\big(G^{(K)}x\big) := \frac{1}{2}G^{(K)}x\cdot x = \frac{1}{2|\mathcal{D}_K|^2}\sum_{k,k'\in\mathcal{D}_K} g(kk')x_k x_{k'} = \mathsf{C}_\infty\big(G_{\mu_x^{(K)}}\big). \tag{4.17}$$

This projected non-linearity is well-defined by the Fubini-Tonelli theorem (see equations (3.71)-(3.72)). In this notation, the finite-dimensional approximation of the Hamilton-Jacobi equation (4.6) reads

$$\partial_t f^{(K)}(t,x) = \mathsf{C}_K\big(\nabla f^{(K)}(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0} \tag{4.18}$$

subject to the initial condition $f^{(K)}(0,\cdot)=\psi^{(K)}(\cdot)$ on $\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$. Just like the equations studied in Section 2.4, this Hamilton-Jacobi equation is posed on positive half-space; however, its non-linearity C_K is defined on the cone $\mathcal{C}_K$ as opposed to the Euclidean space $\mathbb{R}^{\mathcal{D}_K}$. To overcome this difference, an appropriate extension $\mathsf{H}_K:\mathbb{R}^{\mathcal{D}_K}\to\mathbb{R}$ of the non-linearity C_K will be introduced, and instead the Hamilton-Jacobi equation

$$\partial_t f^{(K)}(t,x) = \mathsf{H}_K\big(\nabla f^{(K)}(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^{\mathcal{D}_K}_{>0} \tag{4.19}$$

subject to the initial condition $f^{(K)}(0,\cdot)=\psi^{(K)}(\cdot)$ on $\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$ will be considered. Notice that the cone $\mathbb{R}^{\mathcal{D}_K}_{>0}$ has been used as opposed to the more intuitive cone $\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$. As in Section 2.4, the non-decreasingness of the projected non-linearity (4.17) will make these two choices equivalent. In particular, it will not be necessary to endow the projected Hamilton-Jacobi equation with a boundary condition. Remembering that this Hamilton-Jacobi equation appears in the context of statistical inference makes this insight rather reassuring. Indeed, the statistical inference model does not suggest an obvious choice of boundary condition—given that ultimately the value of the solution at a point in $\mathbb{R}_{\geqslant 0}\times\{0\}$ is the quantity of interest, at least a Dirichlet boundary condition should not be used! In earlier works, the imposition of a Neumann-type boundary condition was observed to be a workable option [83, 85, 86]; however, in [33], it was shown that this somewhat artificial choice is not necessary, and no boundary condition needs to be specified, because the non-linearity "points in the right direction".

The precise assumptions on the initial condition ψ and the kernel g that will be used to obtain the well-posedness of the infinite-dimensional Hamilton-Jacobi equation (4.6) are now stated. In the same spirit as Section 2.4 and [32, 33, 83, 86], the initial conditions $\psi^{(K)}$ and ψ will need to satisfy a certain number of Lipschitz continuity assumptions. For the theory developed to encapsulate the Hamilton-Jacobi equation (3.91) appearing in the context of the sparse stochastic block model, these Lipschitz conditions will be relative to the normalized-ℓ^1 and normalized-$\ell^{1,*}$ norms as opposed to the Euclidean norm. Given an integer $d\geqslant 1$, introduce the normalized-ℓ^1 and normalized-$\ell^{1,*}$ norms, defined for every $x,y\in\mathbb{R}^d$ by

$$|||x|||_1 := \frac{1}{d}\sum_{k=1}^{d}|x_k| \quad\text{and}\quad |||y|||_{1,*} := \max_{k\leqslant d} d|y_k|. \tag{4.20}$$

The underlying dimension $d\geqslant 1$ will be kept implicit but will always be clear from the context. The normalized-ℓ^1 norm is meant to measure elements of $\mathbb{R}^{\mathcal{D}_k}$ with a scaling that is consistent with the identification of this space with the space of measures $\mathcal{M}_+^{(K)}$. The normalized-$\ell^{1,*}$ norm serves to measure elements of the dual space, and is defined so that the Hölder-type inequality $x\cdot y\leq |||x|||_1\,|||y|||_{1,*}$ is valid.

The key continuity assumption on the projected initial condition $\psi^{(K)}$ that will make it possible to establish the well-posedness of the projected Hamilton-Jacobi equations will be Lipschitz continuity with respect to the normalized-ℓ^1 norm. Another way to encode this property is to require the initial condition $\psi:\mathcal{M}_+\to\mathbb{R}$ to be Lipschitz continuous with respect to the total variation distance on $\mathcal{M}_+$ defined in (A.26). The normalized-$\ell^{1,*}$ norm will play its part when discussing the Lipschitz continuity of the projected non-linearity (4.17). To determine the convergence of the projected solutions, it will be important to assume that the initial condition $\psi:\mathcal{M}_+\to\mathbb{R}$ is Lipschitz continuous with respect to the Wasserstein distance on the set of probability measures $\Pr[-1,1]$ defined in (3.85). The final assumption on the initial condition will ensure that, in a sense to be made precise, the solution to the projected Hamilton-Jacobi equation has a bounded gradient close to the projected cone $\mathcal{C}_K$ defined in (4.16). It would of course be more convenient to assume that the gradient really belongs to $\mathcal{C}_K$, rather than only being close to it, but unlike in earlier works, as shown by (3.73), this stronger property does not hold in the context of the sparse stochastic block model. To impose the boundedness of the gradient, fix $a>0$, and for each integer $K\geqslant 1$ introduce the closed convex set

$$\mathcal{K}_{a,K} := \left\{G^{(K)}x\in\mathbb{R}^{\mathcal{D}_K}\mid x\in\mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}\text{ and }|||x|||_1\leqslant a\right\}\subseteq\mathcal{C}_K. \tag{4.21}$$

Given a closed convex set $\mathcal{K}\subseteq\mathbb{R}^d$, write

$$\mathcal{K}' := \mathcal{K}+B_{d^{-1/2}}(0) \tag{4.22}$$

for the neighbourhood of radius $d^{-1/2}$ around $\mathcal{K}$ in the normalized-$\ell^{1,*}$ norm. Here

$$B_r(x) := \left\{x' \in \mathbb{R}^d \mid \|x'-x\|_{1,*} \leqslant r\right\} \tag{4.23}$$

denotes the closed ball of radius $r > 0$ centred around $x \in \mathbb{R}^d$ relative to the normalized-$\ell^{1,*}$ norm. A Lipschitz continuous function $h : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ is said to have its gradient in $\mathcal{K}$ if

$$\nabla h \in L^\infty\left(\mathbb{R}^d_{\geqslant 0}; \mathcal{K}\right). \tag{4.24}$$

Recall that a Lipschitz continuous function is differentiable almost everywhere by Rademacher's theorem (Theorem 2.10 in [50]), so the spatial gradient ∇h is well-defined as an element of L^∞, and the condition (4.24) requires that this object take values in $\mathcal{K}$ almost everywhere. A non-differential criterion for the gradient of a Lipschitz continuous function to lie in a closed convex set is given in Proposition A.2, and will be used frequently throughout this chapter. As will be shown below, assuming that the initial condition has its gradient in $\mathcal{K}'_{a,K}$ suffices to ensure that the gradient of the solution remains in this set at all times. Notice that this is insufficient to be able to evaluate the non-linearity C_K at the gradient of the solution; however, under suitable Lipschitz continuity properties of the extension H_K, it ensures that the projected Hamilton-Jacobi equation (4.19) should be an adequate replacement for the Hamilton-Jacobi equation (4.18). In particular, it justifies defining the solution to the infinite-dimensional Hamilton-Jacobi (4.6) as the limit of the solutions to the projected Hamilton-Jacobi equation (4.19). Besides some smoothness, the only constraint imposed on the kernel $g : [-1,1] \to \mathbb{R}$ is that it be strictly positive. Among other things, this assumption ensures that a non-negative measure $\mu \in \mathcal{M}_+$ cannot have a large total mass unless the function G_μ takes large values. In summary, the assumptions on the kernel $g : [-1,1] \to \mathbb{R}$ and the initial condition $\psi : \mathcal{M}_+ \to \mathbb{R}$ required for the Hamilton-Jacobi equation (4.6) to be well-posed are the following.

H1 The kernel $g : [-1,1] \to \mathbb{R}$ is continuously differentiable and bounded away from zero by some positive constant $m > 0$,

$$g(x) \geqslant m. \tag{4.25}$$

H2 The initial condition $\psi : \mathcal{M}_+ \to \mathbb{R}$ is Lipschitz continuous with respect to the total variation distance (A.26),

$$|\psi(\mu) - \psi(\nu)| \leqslant \|\psi\|_{\mathrm{Lip,TV}} \mathrm{TV}(\mu,\nu) \tag{4.26}$$

for all measures $\nu, \mu \in \mathcal{M}_+$.

H3 There exists $a > 0$ such that the initial condition $\psi : \mathcal{M}_+ \to \mathbb{R}$ has the property that each of the projected initial conditions (4.13) has its gradient in the set $\mathcal{K}'_{a,K}$,

$$\nabla \psi^{(K)} \in L^\infty\left(\mathbb{R}^d_{\geqslant 0}; \mathcal{K}'_{a,K}\right). \tag{4.27}$$

H4 The initial condition $\psi : \mathrm{Pr}[-1,1] \to \mathbb{R}$ is Lipschitz continuous with respect to the Wasserstein distance (3.85),

$$|\psi(\mathbb{P}) - \psi(\mathbb{Q})| \leqslant \|\psi\|_{\mathrm{Lip},W} W(\mathbb{P},\mathbb{Q}) \tag{4.28}$$

for all probability measures $\mathbb{P}, \mathbb{Q} \in \mathrm{Pr}[-1,1]$.

Observe that the hypothesis (**H2**) on the initial condition implies that the projected initial conditions (4.13) are

Lipschitz continuous with respect to the normalized-ℓ^1 norm,

$$\left|\psi^{(K)}(x)-\psi^{(K)}(x')\right| \leqslant \|\psi\|_{\mathrm{Lip,TV}}\mathrm{TV}\big(\mu_x^{(K)},\mu_{x'}^{(K)}\big) \leqslant \|\psi\|_{\mathrm{Lip,TV}}|\!|\!|x-x'|\!|\!|_1. \tag{4.29}$$

With these assumptions at hand, it is natural to wonder why the main result in [33] cannot simply be invoked to obtain the well-posedness of the projected Hamilton-Jacobi equation (4.19). The setting proposed in [33] is that of a Hamilton-Jacobi equation posed on a cone $\mathcal{C}$ and with a non-linearity that is defined over the cone $\mathcal{C}$ as well; the key assumption to establish well-posedness is that the non-linearity and the initial condition have their gradients in the cone $\mathcal{C}$. In the present context, the non-linearity is initially only well-defined on the cone $\mathcal{C}_\infty$, or $\mathcal{C}_K$ for the projected equations, and it must be ensured that the gradient of the solution remains in this space. This suggests that the results in [33] should be used with $\mathcal{C}=\mathcal{C}_\infty$, or $\mathcal{C}_K$ for the projected equations. However, the problem of interest, say for the projected equations, is naturally posed over $\mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$ rather than $\mathcal{C}_K$, and moreover, the gradient of the non-linearity that appears in the present setting is not in $\mathcal{C}_K$, although it is in $\mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$. To make matters more complicated, the gradient of the finite-dimensional initial condition, and therefore also of the solution, does not quite belong to $\mathcal{C}_K$, although it is in the closed convex set $\mathcal{K}'_{a,K}$. Despite all this, it will be shown that the somewhat richer geometry of the problem at hand can be dealt with using arguments that are similar to those in [33].

The structure of these arguments is now described in more detail, and the main results they lead to are stated. It will first be shown that for any $R>0$, it is possible to define a non-linearity $\mathsf{H}_{K,R}:\mathbb{R}^{\mathcal{D}_K}\to\mathbb{R}$ which agrees with the projected non-linearity C_K on a large enough ball $\mathcal{C}_K\cap B_R(0)$, and is uniformly Lipschitz continuous. The well-posedness of the projected Hamilton-Jacobi equation

$$\partial_t f^{(K)}(t,x)=\mathsf{H}_{K,R}\big(\nabla f^{(K)}(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}_{>0}^{\mathcal{D}_K} \tag{4.30}$$

subject to the initial condition $f^{(K)}(0,\cdot)=\psi^{(K)}(\cdot)$ on $\mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$ will then be obtained. Finally, it will be shown that the solutions to these projected Hamilton-Jacobi equations admit a limit as K tends to infinity. This limit will be verified to not depend on the choice of the extension $\mathsf{H}_{K,R}$, provided that R is chosen sufficiently large, and it will be defined as the solution to the infinite-dimensional Hamilton-Jacobi equation (4.6).

To state the main well-posedness results precisely, it will be convenient to introduce additional notation similar to that in (2.90)-(2.92). Given functions $h:\mathbb{R}_{\geqslant 0}^d\to\mathbb{R}$ and $u:\mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}^d\to\mathbb{R}$, define the semi-norms

$$|\!|\!|h|\!|\!|_{\mathrm{Lip},1}:=\sup_{x\neq x'\in\mathbb{R}_{\geqslant 0}^d}\frac{|h(x)-h(x')|}{|\!|\!|x-x'|\!|\!|_1} \quad \text{and} \quad [u]_0:=\sup_{\substack{t>0\\ x\in\mathbb{R}_{\geqslant 0}^d}}\frac{|u(t,x)-u(0,x)|}{t}, \tag{4.31}$$

and introduce the space of functions with Lipschitz initial condition that grow at most linearly in time,

$$\mathfrak{L}:=\big\{u:\mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}^d\to\mathbb{R}\mid u(0,\cdot)\text{ is Lipschitz continuous and }[u]_0<+\infty\big\}, \tag{4.32}$$

as well as its subset of uniformly Lipschitz continuous functions,

$$\mathfrak{L}_{\mathrm{unif}}:=\Big\{u\in\mathfrak{L}\mid \sup_{t\geqslant 0}|\!|\!|u(t,\cdot)|\!|\!|_{\mathrm{Lip},1}<+\infty\Big\}. \tag{4.33}$$

The main well-posedness results for the projected Hamilton-Jacobi equation (4.30) and the infinite-dimensional Hamilton-Jacobi equation (4.6) now read as follows.

Theorem 4.1. *Suppose* (**H1**)-(**H3**)*, and fix* $R>0$*. The projected Hamilton-Jacobi equation* (4.30) *subject to the initial condition* $\psi^{(K)}$ *admits a unique viscosity solution* $f_R^{(K)} \in \mathfrak{L}_{\mathrm{unif}}$ *which satisfies the Lipschitz bound*

$$\sup_{t>0} |||f_R^{(K)}(t,\cdot)|||_{\mathrm{Lip},1} = |||\psi^{(K)}|||_{\mathrm{Lip},1} \leqslant \|\psi\|_{\mathrm{Lip,TV}}, \tag{4.34}$$

and has its gradient in the set $\mathcal{K}'_{a,K}$*. Moreover, if* $u^{(K)}, v^{(K)} \in \mathfrak{L}_{\mathrm{unif}}$ *are a continuous subsolution and a continuous supersolution to the Hamilton-Jacobi equation* (4.30)*, then*

$$\sup_{\mathbb{R}_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}} \left(u^{(K)}(t,x) - v^{(K)}(t,x)\right) = \sup_{\mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}} \left(u^{(K)}(0,x) - v^{(K)}(0,x)\right). \tag{4.35}$$

Theorem 4.2. *Suppose* (**H1**)-(**H4**)*, and given an integer* $K \geqslant 1$ *and a real number* $R > \|\psi\|_{\mathrm{Lip,TV}}$*, denote by* $f_R^{(K)} \in \mathfrak{L}_{\mathrm{unif}}$ *the unique viscosity solution to the Hamilton-Jacobi equation* (4.30) *constructed in Theorem 4.1. For every* $t \geqslant 0$ *and every measure* $\mu \in \mathcal{M}_+$*, the limit*

$$f(t,\mu) = \lim_{K\to+\infty} f_R^{(K)}(t, x^{(K)}(\mu)) \tag{4.36}$$

exists, is finite, and is independent of R. The value of this limit is defined to be the solution to the infinite-dimensional Hamilton-Jacobi equation (4.6)*.*

Solutions to (4.6) satisfy a comparison principle since a comparison principle holds for solutions to the projected Hamilton-Jacobi equation (4.30) by Theorem 4.1.

As is apparent, and similarly to [83, 86], the analysis of this chapter merely identifies the solution to (4.6) as the limit of its finite-dimensional approximations. This will suffice to understand the limit of the mutual information in the sparse stochastic block model. The question of providing a more intrinsic characterization of the solution to (4.6), as was achieved in [32] in a related context, is left open.

In addition to these well-posedness results, a Hopf-Lax variational representation for the solution to the infinite-dimensional Hamilton-Jacobi equation (4.6) in the case when the non-linearity C_∞ is convex is also obtained. Hopf-Lax formulas for related problems have been explored in [29, 30, 32, 33]. In Chapter 6, this variational representation will make it possible to verify that, as stated in Theorem 1.7, in the disassortative regime, the conjectured asymptotic mutual information for the sparse stochastic block model coincides with the value of the asymptotic mutual information established in [38]. The convexity condition on C_∞ boils down to the requirement that the mapping $(x,y)\mapsto g(xy)$ be non-negative definite, and can be phrased as follows.

H5 The kernel $g:[-1,1]\to\mathbb{R}$ satisfies the property

$$\int_{-1}^{1}\int_{-1}^{1} g(xy)\,\mathrm{d}\mu(x)\,\mathrm{d}\mu(y) \geqslant 0 \tag{4.37}$$

for every signed measure $\mu \in \mathcal{M}_s$.

Theorem 4.3. *If* (**H1**)-(**H5**) *hold, then the unique solution* $f:\mathbb{R}_{\geqslant 0}\times\mathcal{M}_+\to\mathbb{R}$ *to the infinite-dimensional Hamilton-Jacobi equation* (4.6) *constructed in Theorem 4.2 admits the Hopf-Lax variational representation*

$$f(t,\mu) = \sup_{\nu\in\mathcal{M}_+}\left(\psi(\mu+t\nu) - \frac{t}{2}\int_{-1}^{1} G_\nu(y)\,\mathrm{d}\nu(y)\right) \tag{4.38}$$

for every $t>0$ *and* $\mu\in\mathcal{M}_+$*. Moreover, the supremum in* (4.38) *is achieved at some* $\nu^*\in\mathcal{M}_+$*, and whenever* *the initial condition ψ admits a Gateaux derivative at the measure $\mu + t\nu^*$ with a density $x \mapsto D_\mu \psi(\mu + t\nu^*, x)$ belonging to the cone $\mathcal{C}_\infty$,*

$$G_{\nu^*} = D_\mu \psi(\mu + t\nu^*, \cdot). \tag{4.39}$$

To study the sparse stochastic block model, it will also be important to identify solutions to equations of the form (4.6) with a kernel g that does not satisfy the positivity assumption (**H1**). The idea will be to introduce a new kernel that satisfies (**H1**) by translating g, and to deduce the well-posedness of the equation with kernel g from the well-posedness of the equation with the translated kernel. For this strategy to work, the assumption (**H3**) on the initial condition will be replaced by a stronger assumption now described.

For every $a \in \mathbb{R}$, introduce the set of measures with mass a,

$$\mathcal{M}_{a,+} := \{\mu \in \mathcal{M}_+ \mid \mu[-1,1] = a\}, \tag{4.40}$$

as well as the set of functions

$$\mathcal{C}_{a,\infty} := \{G_\mu \mid \mu \in \mathcal{M}_{a,+}\}. \tag{4.41}$$

The assumption (**H3**) on the initial condition will essentially be replaced by the assumption that its Gateaux derivative lies in the set $\mathcal{C}_{a,\infty}$ for some $a \in \mathbb{R}$. As before, it will be convenient to state this as an assumption on the projected initial conditions (4.13). For every integer $K \geqslant 1$, introduce the set of projected measures with mass a,

$$\mathcal{M}^{(K)}_{a,+} := \{\mu \in \mathcal{M}^{(K)}_+ \mid \mu[1,1] = a\}, \tag{4.42}$$

and write

$$\mathcal{K}_{=a,K} := \left\{G^{(K)} x \in \mathbb{R}^{\mathcal{D}_K} \mid x \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0} \text{ and } |||x|||_1 = a\right\} \tag{4.43}$$

for its associated set of functions. The assumption (**H3**) is replaced by the following stronger assumption.

H3' There exists $a > 0$ such that the initial condition $\psi : \mathcal{M}_+ \to \mathbb{R}$ has the property that each of the projected initial conditions (4.13) has its gradient in the set $\mathcal{K}'_{=a,K}$,

$$\nabla \psi^{(K)} \in L^\infty(\mathbb{R}^d_{\geqslant 0}; \mathcal{K}'_{=a,K}). \tag{4.44}$$

Formal calculations now suggest a way to modify the solution to the infinite-dimensional Hamilton-Jacobi equation (4.6) if the kernel g is not assumed to satisfy (**H1**) but is translated by a large enough constant so that it becomes positive. Given a continuously differentiable kernel $g : [-1,1] \to \mathbb{R}$, fix $b \in \mathbb{R}$ such that the modified kernel

$$\widetilde{g}_b(z) := g(z) + b \tag{4.45}$$

is strictly positive. For every $\mu \in \mathcal{M}_+$, define the modified function $\widetilde{G}_{b,\mu} : [-1,1] \to \mathbb{R}$,

$$\widetilde{G}_{b,\mu}(x) := \int_{-1}^{1} \widetilde{g}_b(xy) \, \mathrm{d}\mu(y), \tag{4.46}$$

the modified cone of functions,

$$\widetilde{\mathcal{C}}_{b,\infty} := \{\widetilde{G}_{b,\mu} \mid \mu \in \mathcal{M}_+\}, \tag{4.47}$$

and the modified non-linearity $\widetilde{\mathsf{C}}_{b,\infty} : \widetilde{\mathcal{C}}_{b,\infty} \to \mathbb{R}$,

$$\widetilde{\mathsf{C}}_{b,\infty}\big(\widetilde{G}_{b,\mu}\big) := \frac{1}{2}\int_{-1}^{1} \widetilde{G}_{b,\mu}(x)\,\mathrm{d}\mu(x) = \frac{1}{2}\int_{-1}^{1}\int_{-1}^{1} \widetilde{g}_b(xy)\,\mathrm{d}\mu(y)\,\mathrm{d}\mu(x). \tag{4.48}$$

Notice that the additional constant b in $\widetilde{g}_b$ induces a shift in the expression above that depends only on the total mass of the measure μ. This suggests that, under assumption (**H3'**), if $\widetilde{f}_b$ is a solution to the infinite-dimensional Hamilton-Jacobi equation

$$\partial_t \widetilde{f}(t,\mu) = \widetilde{\mathsf{C}}_{b,\infty}\big(D_\mu \widetilde{f}(t,\mu)\big) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathcal{M}_+ \tag{4.49}$$

subject to the initial condition $\widetilde{\psi}_b : \mathcal{M}_+ \to \mathbb{R}$ defined by

$$\widetilde{\psi}_b(\mu) := \psi(\mu) + ab\int_{-1}^{1} \mathrm{d}\mu, \tag{4.50}$$

then the function

$$f_b(t,\mu) := \widetilde{f}_b(t,\mu) - ab\int_{-1}^{1}\mathrm{d}\mu - \frac{a^2bt}{2} \tag{4.51}$$

should be a solution to the infinite-dimensional Hamilton-Jacobi equation (4.6). The dependence of $\widetilde{f}_b$, f_b and $\widetilde{\psi}_b$ on a is omitted since this constant is given and fixed by (**H3'**). The following result renders this construction precise and ensures that it is independent of the choice of b.

Theorem 4.4. *Fix a continuously differentiable kernel $g : [-1,1] \to \mathbb{R}$, and assume that* (**H2**), (**H3'**), *and* (**H4**) *hold. Let $b \in \mathbb{R}$ be such that the function $\widetilde{g}_b$ defined in* (4.45) *is positive on $[-1,1]$, and let $\widetilde{\psi}_b$ be defined by* (4.50)*. Then, the solution $\widetilde{f}_b$ to the infinite-dimensional Hamilton-Jacobi equation* (4.49) *subject to the initial condition $\widetilde{\psi}_b$ constructed in Theorem 4.2 is well-defined, and the function f_b given by* (4.51) *does not depend on the choice of $b \in \mathbb{R}$. The function f_b is defined to be the solution to the infinite-dimensional Hamilton-Jacobi equation* (4.6).

Combining this well-posedness result with the Hopf-Lax representation formula in Theorem 4.3 shows that, under the additional assumption (**H5**), the function (4.51) admits a variational representation.

Theorem 4.5. *Fix a continuously differentiable kernel $g : [-1,1] \to \mathbb{R}$ satisfying* (**H2**), (**H3'**), *and* (**H4**)*. Suppose that there exists $b \in \mathbb{R}$ such that the translated kernel $\widetilde{g}_b$ in* (4.45) *is strictly positive on $[-1,1]$ and satisfies* (**H5**)*. Suppose moreover that for every $\mu \in \mathcal{M}_+$, the initial condition ψ admits a Gateaux derivative with density $x \mapsto D_\mu\psi(\mu,x)$ belonging to the set $\mathcal{C}_{a,\infty}$. Then, the unique solution $f : \mathbb{R}_{\geqslant 0}\times\mathcal{M}_+ \to \mathbb{R}$ to the infinite-dimensional Hamilton-Jacobi equation* (4.6) *constructed in Theorem 4.4 admits the Hopf-Lax variational representation*

$$f(t,\mu) = \sup_{\nu\in\mathcal{M}_{a,+}}\left(\psi(\mu+t\nu) - \frac{t}{2}\int_{-1}^{1} G_\nu(y)\,\mathrm{d}\nu(y)\right) \tag{4.52}$$

for every $t > 0$ and $\mu \in \mathcal{M}_+$. Moreover, the supremum in (4.52) *is achieved at some $\nu^* \in \mathcal{M}_{a,+}$ with*

$$G_{\nu^*} = D_\mu\psi(\mu + t\nu^*, \cdot). \tag{4.53}$$

As previously mentioned, the reader eager to return to the study of the sparse stochastic block model may consider taking Theorems 4.1 - 4.5 for granted, and skipping the rest of this chapter on first reading.

4.2 Establishing well-posedness of infinite-dimensional equations

To obtain the well-posedness of the Hamilton-Jacobi equation (4.6) and establish Theorem 4.2, the non-linearity associated with each of the approximating Hamilton-Jacobi equations (4.18) will be extended from the projected cone (4.16) where it is naturally defined to all of Euclidean space. Importantly, this extension will preserve the non-decreasingness and Lipschitz continuity of the non-linearity. This extension is defined in Section 4.2.1. In Section 4.2.2, a general well-posedness theory for Hamilton-Jacobi equations on positive half-space analogous to that discussed in Section 2.4 will be developed in the setting where the non-linearity is non-decreasing and locally Lipschitz continuous with respect to the normalized-$\ell^{1,*}$ norm, and where the initial condition is Lipschitz continuous with respect to the normalized-ℓ^1 norm. In particular, this will lead to a proof of Theorem 4.1. Finally, in Section 4.2.3, the sequence of solutions to the approximating Hamilton-Jacobi equations will be shown to converge as the dimension parameter K tends to infinity, and the limit thus obtained will be defined as the solution to the infinite-dimensional Hamilton-Jacobi equation (4.6). This will prove Theorem 4.2, and thereby establish a well-posedness theory for the infinite-dimensional Hamilton-Jacobi equation (4.6).

4.2.1 Extending the approximating Hamilton-Jacobi equations

To alleviate notation and strive for generality, instead of extending the projected non-linearity (4.17) from the projected cone (4.16) to all of Euclidean space, an integer dimension $d \geqslant 1$ will be fixed, and a general version of this non-linearity defined on a cone in $\mathbb{R}^d$ will be extended from said cone to all of $\mathbb{R}^d$. More precisely, fix a symmetric matrix $G \in \mathbb{R}^{d\times d}$ for which there exist positive constants $m, M > 0$ with

$$\frac{m}{d^2} \leqslant G_{kk'} \leqslant \frac{M}{d^2} \tag{4.54}$$

for all $1 \leqslant k, k' \leqslant d$, and consider the cone

$$\mathcal{C} := \left\{ Gx \in \mathbb{R}^d \mid x \in \mathbb{R}^d_{\geqslant 0} \right\}, \tag{4.55}$$

and the non-linearity $\mathsf{C} : \mathcal{C} \to \mathbb{R}$ defined by

$$\mathsf{C}(Gx) := \frac{1}{2} Gx \cdot x. \tag{4.56}$$

This mapping is well-defined by the Fubini-Tonelli theorem (see equations (3.71)-(3.72)). The projected non-linearity (4.17) and the projected cone (4.16) are recovered by choosing $d = |\mathcal{D}_K|$ and $G = G^{(K)}$. Observe that (4.54) is satisfied for these choices by the continuity of the kernel g and its lower bound in (**H1**). As in Section 2.4, to establish the well-posedness of the approximating Hamilton-Jacobi equations (4.19), it will be important that the extended non-linearity $\mathsf{H}_R : \mathbb{R}^d \to \mathbb{R}$ be Lipschitz continuous and non-decreasing. The main result of this section is the definition of a uniformly Lipschitz continuous and non-decreasing non-linearity $\mathsf{H}_R : \mathbb{R}^d \to \mathbb{R}$ which agrees with C on the intersection of the cone $\mathcal{C}$ and a large enough ball. Recall the definition of a non-decreasing function in (2.69), and the notation $\leqslant$ for the partial order defined in (2.70). First, these properties are verified locally on the cone for the original non-linearity (4.56). It will be convenient to note that for all $x \in \mathbb{R}^d_{\geqslant 0}$,

$$|||x|||_1 \leqslant \frac{1}{m} |||Gx|||_{1,*}. \tag{4.57}$$

Lemma 4.6. *The non-linearity (4.56) is locally Lipschitz continuous with respect to the normalized-$\ell^{1,*}$ norm,*

$$|\mathsf{C}(y)-\mathsf{C}(y')| \leqslant \frac{1}{m}\big(|||y|||_{1,*}+|||y'|||_{1,*}\big)|||y-y'|||_{1,*} \tag{4.58}$$

for all $y,y' \in \mathcal{C}$.

Proof. Fix $y,y' \in \mathcal{C}$ with $y = Gx$ and $y' = Gx'$ for some $x,x' \in \mathbb{R}^d_{\geqslant 0}$. The symmetry of G and the Cauchy-Schwarz inequality imply that

$$|\mathsf{C}(y)-\mathsf{C}(y')| \leqslant |G(x-x')\cdot x| + |G(x-x')\cdot x'| \leqslant \big(|||x|||_1+|||x'|||_1\big)|||y-y'|||_{1,*}.$$

It follows by (4.57) that

$$|\mathsf{C}(y)-\mathsf{C}(y')| \leqslant \frac{1}{m}\big(|||y|||_{1,*}+|||y'|||_{1,*}\big)|||y-y'|||_{1,*}.$$

This completes the proof. ∎

Lemma 4.7. *The non-linearity (4.56) is non-decreasing.*

Proof. Fix $y,y' \in \mathcal{C}$ with $y'-y \in \mathbb{R}^d_{\geqslant 0}$, and let $x,x' \in \mathbb{R}^d_{\geqslant 0}$ be such that $y = Gx$ and $y' = Gx'$. Observe that

$$2\mathsf{C}(y) = Gx\cdot x = y\cdot x \leqslant y'\cdot x = Gx'\cdot x = Gx\cdot x' = x'\cdot y \leqslant x'\cdot y' = Gx'\cdot x' = 2\mathsf{C}(y').$$

This completes the proof. ∎

Extending the non-linearity (4.56) to $\mathbb{R}^d$ while preserving these two key properties requires some care. For each $R>0$, a non-decreasing function $\mathsf{H}_R:\mathbb{R}^d\to\mathbb{R}$ which is uniformly Lipschitz continuous with respect to the normalized-$\ell^{1,*}$ norm and agrees with the non-linearity (4.56) on the intersection of the cone (4.55) and the ball $B_R := B_R(0)$ defined in (4.23) will be introduced. The definition of this extension is inspired by Proposition 6.8 in [83] and Lemma 2.5 in [33].

Proposition 4.8. *For every $R>0$, there exists a non-decreasing non-linearity $\mathsf{H}_R:\mathbb{R}^d\to\mathbb{R}$ which agrees with C on $\mathcal{C}\cap B_R$ and satisfies the Lipschitz continuity property*

$$|\mathsf{H}_R(y)-\mathsf{H}_R(y')| \leqslant \frac{8RM}{m^2}|||y-y'|||_{1,*} \tag{4.59}$$

for all $y,y' \in \mathbb{R}^d$.

Proof. The proof proceeds in two steps. First, the non-linearity C is regularized by defining a non-decreasing and uniformly Lipschitz continuous function which agrees with C on $\mathcal{C}\cap B_R$, and then this regularization is extended to $\mathbb{R}^d$.

Step 1: regularizing C. By Lemma 4.6, the non-linearity (4.56) satisfies the Lipschitz bound

$$|\mathsf{C}(y)-\mathsf{C}(y')| \leqslant \frac{4R}{m}|||y-y'|||_{1,*}$$

for all $y,y' \in \mathcal{C}\cap B_{2R}$. With this in mind, let $L := \frac{4R}{m}$, and define the regularized non-linearity $\widetilde{\mathsf{C}}_R:\mathcal{C}\to\mathbb{R}$ by

$$\widetilde{\mathsf{C}}_R(y) := \begin{cases} \max\Big(\mathsf{C}(y), \mathsf{C}(0)+2L\big(|||y|||_{1,*}-R\big)\Big) & \text{if } y\in\mathcal{C}\cap B_{2R}, \\ \mathsf{C}(0)+2L\big(|||y|||_{1,*}-R\big) & \text{if } y\in\mathcal{C}\setminus B_{2R}. \end{cases}$$

To see that $\widetilde{\mathsf{C}}_R$ agrees with C on $\mathcal{C} \cap B_R$, observe that for any $y \in \mathcal{C} \cap B_R$,

$$\mathsf{C}(0) + 2L\big(|||y|||_{1,*} - R\big) \leqslant \mathsf{C}(0) = 0 \leqslant \mathsf{C}(y),$$

where the last inequality uses the non-negativity of the components of G. It will also be convenient to note that by Lipschitz continuity of C on $\mathcal{C} \cap B_{2R}$,

$$\mathsf{C}(0) + 2L\big(|||y|||_{1,*} - R\big) = \mathsf{C}(0) + 2LR = \mathsf{C}(0) + L|||y|||_{1,*} \geqslant \mathsf{C}(y)$$

for any $y \in \mathcal{C} \cap \partial B_{2R}$. This shows that $\widetilde{\mathsf{C}}_R$ is continuous. To establish the non-decreasingness of $\widetilde{\mathsf{C}}_R$, fix $y, y' \in \mathcal{C}$ with $y \leqslant y'$. If $y, y' \in B_{2R}$, then the non-decreasingness of C in Lemma 4.7 implies that $\mathsf{C}(y) \leqslant \mathsf{C}(y')$. Combining this with the fact that $|||y|||_{1,*} \leqslant |||y'|||_{1,*}$ reveals that $\widetilde{\mathsf{C}}_R(y) \leqslant \widetilde{\mathsf{C}}_R(y')$. On the other hand, if $y \in B_{2R}$ and $y' \in \mathcal{C} \setminus B_{2R}$, then

$$\mathsf{C}(y) \leqslant \mathsf{C}(0) + L|||y|||_{1,*} \leqslant \mathsf{C}(0) + L|||y'|||_{1,*} + L\big(|||y'|||_{1,*} - 2R\big) = \widetilde{\mathsf{C}}_R(y'),$$

and

$$\mathsf{C}(0) + 2L\big(|||y|||_{1,*} - R\big) \leqslant \mathsf{C}(0) + 2L\big(|||y'|||_{1,*} - R\big) = \widetilde{\mathsf{C}}_R(y').$$

Once again $\widetilde{\mathsf{C}}_R(y) \leqslant \widetilde{\mathsf{C}}_R(y')$. Finally, if $y \in \mathcal{C} \setminus B_{2R}$, then $2R \leqslant |||y|||_{1,*} \leqslant |||y'|||_{1,*}$ so $y' \in \mathcal{C} \setminus B_{2R}$, and clearly $\widetilde{\mathsf{C}}_R(y) \leqslant \widetilde{\mathsf{C}}_R(y')$. This establishes the non-decreasingness of the regularized non-linearity $\widetilde{\mathsf{C}}_R$. It is now shown that this non-linearity is uniformly Lipschitz continuous. The reverse triangle inequality implies that the map $y \mapsto \mathsf{C}(0) + 2L\big(|||y|||_{1,*} - R\big)$ is Lipschitz continuous with Lipschitz constant at most $2L$. Recall that the maximum of two Lipschitz continuous maps with Lipschitz constants at most L_1 and L_2, respectively, is Lipschitz continuous with Lipschitz constant at most $\max(L_1, L_2)$. This means that $\widetilde{\mathsf{C}}_R$ is Lipschitz continuous with Lipschitz constant at most $2L$ when it is restricted to $\mathcal{C} \cap B_{2R}$ or $\mathcal{C} \setminus B_{2R}$. For $y, y' \in \mathcal{C}$ with $y \in B_{2R}$ and $y' \in \mathcal{C} \setminus B_{2R}$, two cases are distinguished. On the one hand, if $\widetilde{\mathsf{C}}_R(y) = \mathsf{C}(0) + 2L\big(|||y|||_{1,*} - R\big)$, the reverse triangle inequality shows that

$$\big|\widetilde{\mathsf{C}}_R(y) - \widetilde{\mathsf{C}}_R(y')\big| \leqslant 2L\big||||y|||_{1,*} - |||y'|||_{1,*}\big| \leqslant 2L|||y - y'|||_{1,*}.$$

On the other hand, if $\widetilde{\mathsf{C}}_R(y) = \mathsf{C}(y)$, then the reverse triangle inequality reveals that

$$\widetilde{\mathsf{C}}_R(y) - \widetilde{\mathsf{C}}_R(y') \leqslant \mathsf{C}(0) + L|||y|||_{1,*} - \mathsf{C}(0) - 2L\big(|||y'|||_{1,*} - R\big) \leqslant L|||y - y'|||_{1,*} + L\big(2R - |||y'|||_{1,*}\big) \leqslant L|||y - y'|||_{1,*}$$

while the lower bound $\widetilde{\mathsf{C}}_R(y) = \mathsf{C}(y) \geqslant \mathsf{C}(0) + 2L\big(|||y|||_{1,*} - R\big)$ yields

$$\widetilde{\mathsf{C}}_R(y') - \widetilde{\mathsf{C}}_R(y) = 2L\big(|||y'|||_{1,*} - |||y|||_{1,*}\big) \leqslant 2L|||y - y'|||_{1,*}.$$

This shows that $\widetilde{\mathsf{C}}_R$ is a non-decreasing function which agrees with C on $\mathcal{C} \cap B_R$ and satisfies the Lipschitz continuity property

$$\big|\widetilde{\mathsf{C}}_R(y) - \widetilde{\mathsf{C}}_R(y')\big| \leqslant \frac{8R}{m}|||y - y'|||_{1,*} \tag{4.60}$$

for all $y, y' \in \mathcal{C}$.

Step 2: extending to $\mathbb{R}^d$. To extend the regularization of the non-linearity (4.56) to $\mathbb{R}^d$, define the function

$\mathsf{H}_R : \mathbb{R}^d \to \mathbb{R}$ by

$$\mathsf{H}_R(y) := \inf\Big\{\widetilde{\mathsf{C}}_R(w) \mid w \in \mathcal{C} \text{ with } w \geqslant y\Big\}. \tag{4.61}$$

Let $\iota := (1,\ldots,1) \in \mathbb{R}^d$, and observe that the vector $v := \frac{G\iota}{m}$ belongs to the cone $\mathcal{C}$ and satisfies the bounds

$$\frac{1}{d} \leqslant v_k \leqslant \frac{M}{dm} \tag{4.62}$$

for $1 \leqslant k \leqslant d$. In particular, the infimum in (4.61) is never taken over the empty set. Moreover, the non-decreasingness of $\widetilde{\mathsf{C}}_R$ and the fact that this function agrees with C on $\mathcal{C} \cap B_R$ imply that H_R also agrees with C on $\mathcal{C} \cap B_R$. To see that H_R is non-decreasing, fix $y, y' \in \mathbb{R}^d$ with $y' \geqslant y$, and let $w \in \mathcal{C}$ be such that $w \geqslant y'$. Since $w \geqslant y$, the definition of H_R gives $\mathsf{H}_R(y) \leqslant \widetilde{\mathsf{C}}_R(w)$, and taking the infimum over all such w shows that $\mathsf{H}_R(y) \leqslant \mathsf{H}_R(y')$. To establish the Lipschitz continuity of H_R, fix $y, y' \in \mathbb{R}^d$ and let $z := |||y - y'|||_{1,*} v \in \mathcal{C}$. Recalling (4.62) reveals that for any $1 \leqslant k \leqslant d$,

$$y_k - y_k' \leqslant \|y - y'\|_\infty = \frac{1}{d}|||y - y'|||_{1,*} \leqslant v_k |||y - y'|||_{1,*} = z_k.$$

This means that $z \geqslant y - y'$. In particular, if $w \in \mathcal{C}$ is such that $w \geqslant y'$, then $w + z \in \mathcal{C}$ with $w + z \geqslant y$. It follows by (4.60)-(4.62) that

$$\mathsf{H}_R(y) - \widetilde{\mathsf{C}}_R(w) \leqslant \widetilde{\mathsf{C}}_R(w+z) - \widetilde{\mathsf{C}}_R(w) \leqslant \frac{8R}{m}|||z|||_{1,*} \leqslant \frac{8RM}{m^2}|||y - y'|||_{1,*}.$$

Taking the infimum over all such w and reversing the roles of y and y' completes the proof. ∎

This result allows one to use solutions to the projected Hamilton-Jacobi equation (4.19) with non-linearity defined on Euclidean space as opposed to the Hamilton-Jacobi equation (4.18) with non-linearity defined on the projected cone (4.16) to establish a well-posedness theory for the infinite-dimensional Hamilton-Jacobi equation (4.6). However, this first requires a well-posedness theory for Hamilton-Jacobi equations of the form (4.19) on positive half-space. Notice that the well-posedness theory developed in Section 2.4 cannot be applied directly as the non-linearity and the initial condition are not necessarily Lipschitz continuous with respect to the Euclidean norm, but rather relative to the normalized-$\ell^{1,*}$ and normalized-ℓ^1 norms, respectively.

4.2.2 Revisiting Hamilton-Jacobi equations on positive half-space

To alleviate notation and strive for generality, instead of establishing the well-posedness of the extended approximating Hamilton-Jacobi equation (4.30), a non-decreasing non-linearity $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$ that is locally Lipschitz continuous with respect to the normalized-$\ell^{1,*}$ norm, and an initial condition $\psi : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ that is Lipschitz continuous with respect to the normalized-ℓ^1 norm are fixed, and the well-posedness of the Hamilton-Jacobi equation

$$\partial_t f(t,x) = \mathsf{H}\big(\nabla f(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d_{>0} \tag{4.63}$$

subject to the initial condition $f(0,\cdot) = \psi(\cdot)$ on $\mathbb{R}^d_{\geqslant 0}$ is established. Just like in Section 2.4, no boundary condition has to be imposed for this equation since the non-linearity "points in the right direction" in the sense that it is non-decreasing. The notion of solution adopted for the Hamilton-Jacobi equation (4.63) is that of viscosity solution discussed in Definition 2.8. The well-posedness of the Hamilton-Jacobi equation (4.63) will be established over the space $\mathfrak{L}_{\mathrm{unif}}$ of uniformly Lipschitz continuous functions defined in (4.33). In analogy to Section 2.4, the main well-posedness result for the Hamilton-Jacobi equation (4.63) is established by analyzing the Hamilton-Jacobi equation

$$\partial_t f(t,x) = \mathsf{H}\big(\nabla f(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0} \tag{4.64}$$

subject to the initial condition $f(0,\cdot) = \psi(\cdot)$ on $\mathbb{R}^d_{\geqslant 0}$, and showing that solutions to these two equations agree.

Proposition 4.9. *If $\psi : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ is an initial condition that is Lipschitz continuous with respect to the normalized-ℓ^1 norm, and $\mathsf{H} : \mathbb{R}^d \to \mathbb{R}$ is a non-decreasing non-linearity that is locally Lipschitz continuous with respect to the normalized-$\ell^{1,*}$ norm, then the Hamilton-Jacobi equation (4.63) admits a unique viscosity solution $f \in \mathfrak{L}_{\mathrm{unif}}$ with*

$$\sup_{t>0} |||f(t,\cdot)|||_{\mathrm{Lip},1} = |||\psi|||_{\mathrm{Lip},1}. \tag{4.65}$$

Moreover, if $u, v \in \mathfrak{L}_{\mathrm{unif}}$ are a continuous subsolution and a continuous supersolution to (4.63), then

$$\sup_{\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}} \big(u(t,x) - v(t,x)\big) = \sup_{\mathbb{R}^d_{\geqslant 0}} \big(u(0,x) - v(0,x)\big). \tag{4.66}$$

To be more specific, given $\delta_0 > 0$, introduce the Lipschitz constants

$$L := \max\Big(\sup_{t>0} |||u(t,\cdot)|||_{\mathrm{Lip},1}, \sup_{t>0} |||v(t,\cdot)|||_{\mathrm{Lip},1}\Big) \text{ and } V := \sup\left\{\frac{|\mathsf{H}(p') - \mathsf{H}(p)|}{|||p'-p|||_{1,*}} \mid |||p|||_{1,*}, |||p'|||_{1,*} \leqslant L + \delta_0\right\}, \tag{4.67}$$

then for every $R \in \mathbb{R}$ and $M > 2L$, the map

$$(t,x) \mapsto u(t,x) - v(t,x) - M\big(|||x|||_1 + Vt - R\big)_+ \tag{4.68}$$

achieves its supremum on $\{0\} \times \mathbb{R}^d_{\geqslant 0}$.

Proof. The proof follows Section 2.4, and proceeds in four steps. First, a comparison principle is obtained for the Hamilton-Jacobi equation (4.64). Then, the Lipschitz bound (4.65) is established for any of its viscosity solutions, assuming these exist. Subsequently, the Perron method is used to prove the existence of a viscosity solution to the Hamilton-Jacobi equation (4.64). Finally, the monotonicity of the non-linearity is leveraged to prove that solutions to the Hamilton-Jacobi equations (4.63) and (4.64) coincide. The proof closely resembles those in Section 2.4, so instead of providing full details, only the key differences are highlighted. The reader interested in full details is referred to Appendix A in [48].

Step 1: comparison principle for (4.64). Let $u, v \in \mathfrak{L}_{\mathrm{unif}}$ be a viscosity subsolution and a viscosity supersolution to the Hamilton-Jacobi equation (4.64). The purpose of this step is to show that for every $R \in \mathbb{R}$ and $M > 2L$, the map

$$(t,x) \mapsto u(t,x) - v(t,x) - M\big(|||x|||_1 + Vt - R\big)_+ \tag{4.69}$$

achieves its supremum on $\{0\} \times \mathbb{R}^d_{\geqslant 0}$. Suppose for the sake of contradiction that there exists $T > 0$ with

$$\sup_{[0,T] \times \mathbb{R}^d_{\geqslant 0}} \big(u(t,x) - v(t,x) - \varphi(t,x)\big) > \sup_{\mathbb{R}^d_{\geqslant 0}} \big(u(0,x) - v(0,x) - \varphi(0,x)\big),$$

where $\varphi(t,x) := M\big(|||x|||_1 + Vt - R\big)_+$. The only additional difficulty in the present context relative to that in Proposition 2.12 is that the normalized-ℓ^1 norm is not differentiable. To overcome this issue, given $\varepsilon_0 \in (0,1)$

to be determined, introduce the smoothed normalized-ℓ^1 norm,

$$|||x|||_{1,\varepsilon_0} := \frac{1}{d}\sum_{k=1}^{d}\left(x_k^2+\varepsilon_0\right)^{\frac{1}{2}}. \tag{4.70}$$

Arguing as in Step 1 of Proposition 2.12 with the functions Φ, d, and χ_2 redefined to have the Euclidean norm replaced by the smoothed normalized-ℓ^1 norm,

$$\Phi(t,x) := M\theta\left(|||x|||_{1,\varepsilon_0}+Vt-R\right), \qquad \mathrm{d}(x) := \inf_{\substack{|||y|||_{1,*}=1\\ y\in\mathbb{R}^d_{\geqslant 0}}} y\cdot x, \quad \text{and} \quad \chi_2(t,x) := \frac{\delta}{\mathrm{d}(x)}+\delta'|||x|||_{1,\varepsilon_0},$$

it is possible to show that for $\varepsilon_0,\varepsilon,\varepsilon',\delta,\delta'>0$ small enough,

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}}(u-v-\chi_1-\chi_2) > \sup_{\{0\}\times\mathbb{R}^d_{\geqslant 0}}(u-v-\chi_1-\chi_2). \tag{4.71}$$

For each $\alpha\geqslant 1$, define the function $\Psi_\alpha:[0,T]\times\mathbb{R}^d_{\geqslant 0}\times[0,T]\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}\cup\{-\infty\}$ by

$$\Psi_\alpha(t,x,t',x',y) := u(t,x)-v(t',x')-\frac{\alpha}{2}\left(|t-t'|^2+|||x-x'|||_{1,\varepsilon_0}+|||x-y|||_{1,\varepsilon_0}\right)-\chi_1(t,x)-\chi_2(t,x)$$

Arguing as in Step 2 of Proposition 2.12, it is possible to find a sequence of quintuples $(t_\alpha,x_\alpha,t'_\alpha x'_\alpha,y_\alpha)_{\alpha\geqslant 1}$ with the property that Ψ_α achieves its supremum at $(t_\alpha,x_\alpha,t'_\alpha x'_\alpha,y_\alpha)$, and that $t_\alpha,t'_\alpha\in(0,T)$ and $x_\alpha,x'_\alpha,y_\alpha\in\mathbb{R}^d_{>0}$ for α large enough. With this in mind, fix $\alpha\geqslant 1$ large enough, and introduce the functions $\phi,\phi'\in C^\infty((0,T)\times\mathbb{R}^d_{\geqslant 0};\mathbb{R})$ defined by

$$\phi(t,x) := v(t'_\alpha,x'_\alpha)+\frac{\alpha}{2}\left(|t-t'_\alpha|^2+|||x-x'_\alpha|||_{1,\varepsilon_0}+|||x-y_\alpha|||_{1,\varepsilon_0}\right)+\chi_1(t,x)+\chi_2(t,y_\alpha),$$
$$\phi'(t',x') := u(t_\alpha,x_\alpha)-\frac{\alpha}{2}\left(|t'-t_\alpha|^2+|||x'-x_\alpha|||_{1,\varepsilon_0}+|||x_\alpha-y_\alpha|||_{1,\varepsilon_0}\right)-\chi_1(t_\alpha,x_\alpha)-\chi_2(t_\alpha,y_\alpha).$$

Arguing as in Step 3 of Proposition 2.12, it is possible to contradict the inequalities

$$\left(\partial_t\phi-\mathsf{H}(\nabla\phi)\right)(t_\alpha,x_\alpha)\leqslant 0 \quad \text{and} \quad \left(\partial_t\phi'-\mathsf{H}(\nabla\phi')\right)(t'_\alpha,x'_\alpha)\geqslant 0,$$

and conclude that the map (4.69) achieves its supremum on $\{0\}\times\mathbb{R}^d_{\geqslant 0}$.

Step 2: Lipschitz bound for for solutions to (4.64). Let $f\in\mathfrak{L}$ be a viscosity solution to the Hamilton-Jacobi equation (4.64). The purpose of this step is to show that

$$\sup_{t\geqslant 0}|||f(t,\cdot)|||_{\mathrm{Lip},1} = |||f(0,\cdot)|||_{\mathrm{Lip},1}. \tag{4.72}$$

Let $L:=|||f(0,\cdot)|||_{\mathrm{Lip},1}$ denote the Lipschitz semi-norm of the initial condition, and suppose for the sake of contradiction that there exists $T>0$ with

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}}\left(f(t,x)-f(t,x')-L|||x-x'|||_1\right)>0\geqslant \sup_{\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}}\left(f(0,x)-f(0,x')-L|||x-x'|||_1\right). \tag{4.73}$$

Arguing as in Step 1 of Proposition 2.13 with the constant V and the function Φ redefined to be

$$V := \sup\left\{\frac{|\mathsf{H}(p')-\mathsf{H}(p)|}{|||p'-p|||_{1,*}} \mid |||p|||_{1,*}, |||p'|||_{1,*} \leqslant L+\delta_0\right\} \quad \text{and} \quad \Phi(t,x) := \delta_0\theta\big(|||x|||_{1,\varepsilon_0}+Vt-R\big),$$

it is possible to show that for $R>0$ large enough and $\varepsilon_0, \varepsilon, \varepsilon' > 0$ small enough,

$$\sup_{[0,T]\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}} \big(f(t,x)-f(t,x')-L|||x-x'|||_{1,\varepsilon_0}-\chi_1(t,x)-\chi_2(t,x')\big)$$
$$> 0 \geqslant \sup_{\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}} \big(f(0,x)-f(0,x')-L|||x-x'|||_{1,\varepsilon_0}-\chi_1(0,x)-\chi_2(0,x')\big).$$

For each $\alpha \geqslant 1$, define the function $\Psi_\alpha : \mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\times\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0} \to \mathbb{R}\cup\{-\infty\}$ by

$$\Psi_\alpha(t,x,t',x') := f(t,x)-f(t',x')_L|||x-x'|||_{1,\varepsilon_0}-\frac{\alpha}{2}|t-t'|^2-\chi_1(t,x)-\chi_2(t',x').$$

Arguing as in Step 2 of Proposition 2.13, it is possible to find a sequence of quadruples $(t_\alpha, x_\alpha, t'_\alpha x'_\alpha)_{\alpha\geqslant 1}$ with the property that Ψ_α achieves its supremum at $(t_\alpha, x_\alpha, t'_\alpha x'_\alpha)$, and that $t_\alpha, t'_\alpha \in (0,T)$ and $x_\alpha \neq x'_\alpha$ for α large enough. With this in mind, fix $\alpha \geqslant 1$ large enough, and introduce the functions $\phi, \phi' \in C^\infty((0,T)\times\mathbb{R}^d_{\geqslant 0};\mathbb{R})$ defined by

$$\phi(t,x) := f(t'_\alpha, x'_\alpha)+L|||x-x'_\alpha|||_{1,\varepsilon_0}+\frac{\alpha}{2}|t-t'_\alpha|^2+\chi_1(t,x)+\chi_2(t'_\alpha,x'_\alpha),$$
$$\phi'(t',x') := f(t_\alpha,x_\alpha)-L|||x'-x_\alpha|||_{1,\varepsilon_0}-\frac{\alpha}{2}|t'-t_\alpha|^2-\chi_1(t_\alpha,x_\alpha)-\chi_2(t',x').$$

Arguing as in Step 3 of Proposition 2.13, it is possible to contradict the inequalities

$$\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha) \leqslant 0 \quad \text{and} \quad \big(\partial_t\phi'-\mathsf{H}(\nabla\phi')\big)(t'_\alpha,x'_\alpha) \geqslant 0,$$

and establish the Lipschitz bound (4.72). The comparison principle in (4.66) now follows by arguing as in Corollary 2.14, and the uniqueness of solutions to the Hamilton-Jacobi equation (4.64) is obtained as in Corollary 2.15.

Step 3: existence of solutions to (4.64). The purpose of this step is to use the Perron method to prove the existence of solutions to the Hamilton-Jacobi equation (4.64). Fix a positive constant

$$K > \sup\big\{|\mathsf{H}(y)| \mid y\in\mathbb{R}^d \text{ with } |||y|||_{1,*} \leqslant |||\psi|||_{\mathrm{Lip},1}\big\},$$

and define the continuous functions $u_\pm : \mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ by $u_\pm(t,x) := \psi(x) \pm Kt$. Arguing as in Section 2.4.2, it is possible to show that the function $f : \mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ defined by $f(t,x) := \sup_{u\in\mathcal{S}} u(t,x)$ for the set

$$\mathcal{S} := \big\{u : \mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0} \to \mathbb{R} \mid u_- \leqslant u \leqslant u_+ \text{ and } \overline{u} \text{ is a subsolution to (4.64)}\big\}$$

is a viscosity solution to the Hamilton-Jacobi equation (4.64) which belongs to the solution space $\mathfrak{L}_{\mathrm{unif}}$.

Step 4: equivalence of solutions to (4.63) *and* (4.64). The purpose of this step is to show continuous subsolutions and supersolutions to the Hamilton-Jacobi equations (4.63) and (4.64) coincide. The argument for viscosity subsolutions and viscosity supersolutions being almost identical, the focus is exclusively on the case of viscosity subsolutions. Arguing as in Proposition 2.21, it suffices to consider a continuous viscosity subsolution to the Hamilton-Jacobi equation (4.63) and deduce the subsolution criterion for the Hamilton-Jacobi equation (4.64) at a boundary point $(t,x)\in\mathbb{R}_{>0}\times\partial\mathbb{R}^d_{\geqslant 0}$. To be more precise, fix a smooth function $\phi\in C^\infty(\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0};\mathbb{R})$ with the property that $u-\phi$ has a strict local maximum at $(t^*,x^*)\in\mathbb{R}_{>0}\times\partial\mathbb{R}^d_{\geqslant 0}$, and aim to show that $(\partial_t\phi-\mathsf{H}(\nabla\phi))(t^*,x^*)\leqslant 0$. This is done as in the proof of Proposition 2.21 with the functions d and ψ_ε redefined to be

$$\mathrm{d}(x):=\inf_{\substack{|||y|||_{1,*}=1\\ y\in\mathbb{R}^d_{\geqslant 0}}} y\cdot x \quad\text{and}\quad \psi_\varepsilon(s,y):=u(s,y)-\phi(s,y)-\frac{\varepsilon}{\mathrm{d}(y)},$$

and the function $\Psi_{\varepsilon,\delta,\theta}:\overline{\mathcal{O}}_r\times\overline{\mathcal{O}}_r\to\mathbb{R}$ redefined to be

$$\Psi_{\varepsilon,\delta,\theta}(t,x,s,y):=u(t,x)-\phi(s,y)-\frac{\varepsilon}{\mathrm{d}(y)}-\frac{2M_u}{\theta^2}|t-s|^2-\frac{2M_u}{\theta^2}|||x-y|||_{1,\varepsilon_0}+\delta\zeta_\varepsilon(s,y)$$

for $M_u:=\sup_{(t,x)\in\overline{\mathcal{O}}_r}|u(t,x)|$. Together with the previous three steps, this completes the proof. ■

To study the sparse stochastic block model, it will also be important to know that whenever the gradient of the initial condition ψ lies in a closed convex set, then so too does the gradient of the solution to the Hamilton-Jacobi equation (4.63) at all future points in time. This is first established for closed convex cones and then extended to arbitrary closed convex sets. Given a set $\mathcal{D}\subseteq\mathbb{R}^d$, recall from Corollary A.3 that a Lipschitz function $h:\mathcal{D}\to\mathbb{R}$ has its gradient in a closed convex cone $\mathcal{K}$ if and only if it is $\mathcal{K}^*$-non-decreasing. A function $h:\mathcal{D}\to\mathbb{R}$ is said to be $\mathcal{K}^*$*-non-decreasing* if, for all $x,x'\in\mathcal{D}$,

$$x'-x\in\mathcal{K}^* \quad\Longrightarrow\quad h(x)\leqslant h(x'). \tag{4.74}$$

Here $\mathcal{K}^*$ denotes the dual of the cone $\mathcal{K}$,

$$\mathcal{K}^*:=\{x\in\mathbb{R}^d \mid x\cdot y\geqslant 0 \text{ for all } y\in\mathcal{K}\}. \tag{4.75}$$

Notice that a function is non-decreasing in the sense defined in (2.69) if and only if it is $(\mathbb{R}^d_{\geqslant 0})^*$-non-decreasing. To show that the gradient of the initial condition stays in a closed convex cone it therefore suffices to prove that the Hamilton-Jacobi equation (4.63) preserves the monotonicity of its initial condition. This will be done through a doubling argument similar to those used in the proof of Proposition 4.9. It will be convenient to adapt the distance-like function (2.106) to the normalized-ℓ^1 and normalized-$\ell^{1,*}$ norms by redefining it to be

$$\mathrm{d}(x):=\inf_{\substack{|||y|||_{1,*}=1\\ y\in\mathbb{R}^d_{\geqslant 0}}} y\cdot x. \tag{4.76}$$

The basic properties of this function stated in Lemma 2.11 still hold with the appropriate modifications. In particular, property (iii) reads that d is Lipschitz continuous with respect to the normalized-ℓ^1 norm, and property (v) states that the normalized-$\ell^{1,*}$ norm of any element in the superdifferential is at most one.

Lemma 4.10. *Fix a closed convex cone $\mathcal{K}$, an initial condition $\psi:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ that is Lipschitz continuous with respect to the normalized-ℓ^1 norm, and a non-decreasing non-linearity $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ that is locally Lipschitz continuous with respect to the normalized-$\ell^{1,*}$ norm. Denote by $f\in\mathfrak{L}_{\mathrm{unif}}$ the unique viscosity solution to the*

Hamilton-Jacobi equation (4.64) *constructed in Proposition 4.9. If ψ is $\mathcal{K}^*$-non-decreasing, then for all $t \geqslant 0$, the function $f(t,\cdot)$ is also $\mathcal{K}^*$-non-decreasing.*

Proof. Let $L := |||\psi|||_{\mathrm{Lip},1}$ be the Lipschitz semi-norm of the initial condition. Introduce the closed set

$$\Omega := \left\{(x,x') \in \mathbb{R}^d_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \mid x' - x \in \mathcal{K}^*\right\},$$

and suppose for the sake of contradiction that there exists $T > 0$ with

$$\sup_{[0,T]\times\Omega} \big(f(t,x) - f(t,x')\big) > 0 \geqslant \sup_{\Omega} \big(f(0,x) - f(0,x')\big). \tag{4.77}$$

The proof proceeds in three steps. First (4.77) is perturbed, then a variable doubling argument is used to obtain a system of inequalities, and finally, this system of inequalities is contradicted.

Step 1: perturbing. Let $\delta_0 > 0$, $x^* \in \mathbb{R}^d_{>0}$ and $y^* \in \mathcal{K}^*$ with $x^* + y^* \in \mathbb{R}^d_{>0}$ be fixed, and let $\theta \in C^\infty(\mathbb{R};\mathbb{R})$ be an increasing function such that, for every $r \in \mathbb{R}$,

$$(r-1)_+ \leqslant \theta(r) \leqslant r_+ \quad \text{and} \quad |\theta'(r)| \leqslant 1.$$

Recall the definition of the smoothed normalized-ℓ^1 norm in (4.70), and for a constant $R > 0$ to be chosen and the local Lipschitz constant

$$V := \sup\left\{\frac{|\mathsf{H}(p') - \mathsf{H}(p)|}{|||p'-p|||_{1,*}} \mid |||p|||_{1,*}, |||p'|||_{1,*} \leqslant L + \delta_0\right\},$$

introduce the function

$$\Phi(t,x) := \delta_0 \theta\big(|||x|||_{1,\varepsilon_0} + Vt - R\big)$$

defined on $\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d$. For small parameters $\varepsilon, \delta, \delta' > 0$ to be determined, introduce the functions

$$\chi_1(t,x) := \Phi(t,x) + \frac{\varepsilon}{T-t} + \varepsilon t, \quad \chi_2(x,x') := \varepsilon' |||x - x'|||^2_{1,\varepsilon_0}, \quad \text{and} \quad \chi_3(y) := \frac{\delta}{\mathrm{d}(y)} + \delta' |||y|||_{1,\varepsilon_0}$$

defined on $\mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$, $\mathbb{R}^d_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$, and $\mathbb{R}^d_{\geqslant 0}$, respectively. Choosing $R > 0$ large enough and $\varepsilon, \varepsilon', \delta, \delta' > 0$ small enough ensures that

$$\begin{aligned}\sup_{[0,T]\times\Omega} &\big(f(t,x) - f(t,x') - \chi_1(t,x) - \chi_2(x,x') - \chi_3(x) - \chi_3(x')\big)\\ &> 0 \geqslant \sup_{\Omega}\big(f(0,x) - f(0,x') - \chi_1(0,x) - \chi_2(x,x') - \chi_3(x) - \chi_3(x')\big).\end{aligned} \tag{4.78}$$

This is the perturbed version of the hypothesis (4.77) that will be used to reach a contradiction.

Step 2: system of inequalities. For each $\alpha \geqslant 1$, define the function $\Psi_\alpha : [0,T] \times [0,T] \times \Omega \times \mathbb{R}^d_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R} \cup \{-\infty\}$ by

$$\begin{aligned}\Psi_\alpha(t,t',x,x',y,y') := f(t,x) - f(t',x') - \chi_1(t,x) - \chi_2(x,x') - \chi_3(y) - \chi_3(y')\\ - \alpha|t-t'|^2 - \alpha|||x-y|||_{1,\varepsilon_0} - \alpha|||x'-y'|||_{1,\varepsilon_0}.\end{aligned} \tag{4.79}$$

It is now argued that the function Ψ_α achieves its supremum at a point $(t_\alpha,t'_\alpha,x_\alpha,x'_\alpha,y_\alpha,y'_\alpha)$ which remains bounded as α tends to infinity. To do so, write $C<+\infty$ for a constant whose value might change throughout the argument, and which may depend on δ_0, L, R, T, V, and $[f]_0$. For every $x\in\mathbb{R}^d_{\geqslant 0}$ with $|||x|||_{1,\varepsilon_0}>R+1$ and $\alpha\geqslant 1$, the bound $\Phi(t,x)\geqslant\delta_0(|||x|||_{1,\varepsilon_0}+Vt-R-1)$ reveals that

$$\begin{aligned}\Psi_\alpha(t,t',x,x',y,y')&\leqslant[f]_0(t+t')+L|||x-x'|||_1-\Phi(t,x)-\varepsilon'|||x-x'|||^2_{1,\varepsilon_0}-\delta'|||y|||_{1,\varepsilon_0}-\delta'|||y'|||_{1,\varepsilon_0}\\&\leqslant L|||x-x'|||_{1,\varepsilon_0}-\varepsilon'|||x-x'|||^2_{1,\varepsilon_0}-\delta_0|||x|||_{1,\varepsilon_0}-\delta'|||y|||_{1,\varepsilon_0}-\delta'|||y'|||_{1,\varepsilon_0}+C.\end{aligned}$$

Observe also that the supremum of (4.79) is bounded from below by $\Psi_\alpha(0,0,x^*,x^*+y^*,x^*,x^*)$, which does not depend on α. This implies that x_α, x'_α, y_α and y'_α remain bounded both with respect to the normalized-ℓ^1 norm and its smoothed counterpart (4.70) as α tends to infinity, and that

$$\alpha|t_\alpha-t'_\alpha|^2+\alpha|||x_\alpha-y_\alpha|||_{1,\varepsilon_0}+\alpha|||x'_\alpha-y'_\alpha|||_{1,\varepsilon_0}+\frac{\varepsilon}{T-t_\alpha}+\frac{\delta}{\mathrm{d}(y_\alpha)}+\frac{\delta}{\mathrm{d}(y'_\alpha)}\leqslant C. \tag{4.80}$$

It follows that, up to the extraction of a subsequence, there exist $t_0\in[0,T]$ and $x_0,x'_0\in\mathbb{R}^d_{\geqslant 0}$ such that $t_\alpha\to t_0$, $t'_\alpha\to t_0$, $x_\alpha\to x_0$, $x'_\alpha\to x'_0$, $y_\alpha\to x_0$ and $y'_\alpha\to x'_0$ as $\alpha\to+\infty$. By (4.80), property (ii) in Lemma 2.11, and the fact that Ω is closed, it must be that $t_0\in[0,T)$, $x_0,x'_0\in\mathbb{R}^d_{>0}$ and $x'_0-x_0\in\mathcal{K}^*$. On the other hand, the continuity of f, χ_1, χ_2 and χ_3 together with the bounds

$$\begin{aligned}&\sup_{[0,T]\times\Omega}\big(f(t,x)-f(t,x')-\chi_1(t,x)-\chi_2(x,x')-\chi_3(x)-\chi_3(x')\big)\\&\qquad\leqslant\Psi_\alpha(t_\alpha,t'_\alpha,x_\alpha,x'_\alpha,y_\alpha,y'_\alpha)\leqslant f(t_\alpha,x_\alpha)-f(t'_\alpha,x'_\alpha)-\chi_1(t_\alpha,x_\alpha)-\chi_2(x_\alpha,x'_\alpha)-\chi_3(y_\alpha)-\chi_3(y'_\alpha)\end{aligned}$$

imply that the supremum on the left side of (4.78) is achieved at (t_0,x_0,x'_0). It must therefore be that $t_0\in(0,T)$. This means that $(t_\alpha,t'_\alpha,x_\alpha,x'_\alpha,y_\alpha,y'_\alpha)_{\alpha\geqslant 1}$ is a sequence such that Ψ_α achieves its supremum at $(t_\alpha,t'_\alpha,x_\alpha,x'_\alpha,y_\alpha,y'_\alpha)$, and with $t_\alpha,t'_\alpha\in(0,T)$ and $x_\alpha,x'_\alpha,y_\alpha,y'_\alpha\in\mathbb{R}^d_{>0}$ for α large enough. With this in mind, fix α large enough, and introduce the functions $\phi,\phi'\in C^\infty((0,T)\times\mathbb{R}^d_{\geqslant 0};\mathbb{R})$ defined by

$$\begin{aligned}\phi(t,x)&:=f(t'_\alpha,x'_\alpha)+\chi_1(t,x)+\chi_2(x,x'_\alpha)+\chi_3(y_\alpha)+\chi_3(y'_\alpha)+\alpha|t-t'_\alpha|^2+\alpha|||x-y_\alpha|||_{1,\varepsilon_0}+\alpha|||x'_\alpha-y'_\alpha|||_{1,\varepsilon_0},\\\phi'(t',x')&:=f(t_\alpha,x_\alpha)-\chi_1(t_\alpha,x_\alpha)-\chi_2(x_\alpha,x')-\chi_3(y_\alpha)-\chi_3(y'_\alpha)-\alpha|t'-t_\alpha|^2-\alpha|||x_\alpha-y_\alpha|||_{1,\varepsilon_0}-\alpha|||x'-y'_\alpha|||_{1,\varepsilon_0}.\end{aligned}$$

Since $(t_\alpha,t'_\alpha,x_\alpha,x'_\alpha,y_\alpha,y'_\alpha)$ maximizes Ψ_α, the function $f-\phi$ achieves a local maximum at the point $(t_\alpha,x_\alpha)\in(0,T)\times\mathbb{R}^d_{>0}$, while the function $f-\phi'$ achieves a local minimum at the point $(t'_\alpha,x'_\alpha)\in(0,T)\times\mathbb{R}^d_{>0}$. It follows by the definition of a viscosity subsolution and supersolution that

$$\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha)\leqslant 0\quad\text{and}\quad\big(\partial_t\phi'-\mathsf{H}(\nabla\phi')\big)(t'_\alpha,x'_\alpha)\geqslant 0. \tag{4.81}$$

This is the system of inequalities that will be contradicted.

Step 3: reaching a contradiction. Define the function $n:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}^d_{\geqslant 0}$ by $n(x):=\nabla|||x|||_{1,\varepsilon_0}$. A direct computation shows that

$$\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha)=\varepsilon+2\alpha(t_\alpha-t'_\alpha)+\partial_t\Phi(t_\alpha,x_\alpha)+\frac{\varepsilon}{(T-t_\alpha)^2}-\mathsf{H}\big(\nabla\phi(t_\alpha,x_\alpha)\big) \tag{4.82}$$

and

$$\big(\partial_t \phi' - \mathsf{H}(\nabla\phi')\big)(t'_\alpha, x'_\alpha) = 2\alpha(t_\alpha - t'_\alpha) - \mathsf{H}\big(\nabla\phi'(t'_\alpha, x'_\alpha)\big). \tag{4.83}$$

To compare these two quantities, the non-decreasingness and local Lipschitz continuity of the non-linearity H will be used to replace the gradient

$$\nabla\phi(t_\alpha, x_\alpha) = \nabla\Phi(t_\alpha, x_\alpha) + 2\varepsilon' n(x_\alpha - x'_\alpha)\|\!|x_\alpha - x'_\alpha|\!\|_{1,\varepsilon_0} + \alpha n(x_\alpha - y_\alpha) \tag{4.84}$$

by the gradient

$$\nabla\phi'(t'_\alpha, x'_\alpha) = 2\varepsilon' n(x_\alpha - x'_\alpha)\|\!|x_\alpha - x'_\alpha|\!\|_{1,\varepsilon_0} - \alpha n(x'_\alpha - y'_\alpha). \tag{4.85}$$

With the definition of V in mind, it is first shown that

$$\|\!|\nabla\phi(t_\alpha, x_\alpha)|\!\|_{1,*} \leqslant L \quad \text{and} \quad \|\!|\nabla\phi'(t'_\alpha, x'_\alpha)|\!\|_{1,*} \leqslant L. \tag{4.86}$$

Fix $z \in \mathbb{R}^d$ and $\eta > 0$. Since $f - \phi$ achieves a local maximum at $(t_\alpha, x_\alpha) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, and f is uniformly Lipschitz continuous with Lipschitz constant L by Proposition 4.9,

$$\phi(t_\alpha, x_\alpha + \eta z) - \phi(t_\alpha, x_\alpha) \geqslant f(t_\alpha, x_\alpha + \eta z) - f(t_\alpha, x_\alpha) \geqslant -\eta L \|\!|z|\!\|_1.$$

Dividing by η and letting η tend to zero reveals that

$$\nabla\phi(t_\alpha, x_\alpha) \cdot z \geqslant -L\|\!|z|\!\|_1.$$

Choosing $z_k := -d\,\mathrm{sgn}(\partial_{x_k}\phi(t_\alpha, x_\alpha))$ gives the first inequality in (4.86); the second inequality is obtained identically. These bounds would be sufficient if the terms $\alpha n(x_\alpha - y_\alpha)$ and $\alpha n(x'_\alpha - y'_\alpha)$ could be made arbitrarily small. To overcome this issue, the non-decreasingness of H will be leveraged. Since the function $y \mapsto \Psi_\alpha(t_\alpha, t'_\alpha, x_\alpha, x'_\alpha, y, y'_\alpha)$ achieves its maximum at $y_\alpha \in \mathbb{R}^d_{>0}$, properties (v) and (vi) in Lemma 2.11 imply that

$$\frac{1}{\delta}\,\mathrm{d}(y_\alpha)^2\big(\alpha n(y_\alpha - x_\alpha) + \delta' n(y_\alpha)\big) \in \partial\,\mathrm{d}(y_\alpha) \quad \text{and} \quad \mathrm{d}(y_\alpha)^2 \|\!|\alpha n(y_\alpha - x_\alpha) + \delta' n(y_\alpha)|\!\|_{1,*} \leqslant \delta.$$

Decreasing δ if necessary and setting $p_\alpha := \delta' n(y_\alpha)$ gives a vector $p_\alpha \in \mathbb{R}^d_{\geqslant 0}$ with $\|\!|p_\alpha|\!\|_{1,*} \leqslant \delta'$ as well as

$$p_\alpha - \alpha n(x_\alpha - y_\alpha) \in \mathbb{R}^d_{\geqslant 0} \quad \text{and} \quad \|\!|p_\alpha - \alpha n(x_\alpha - y_\alpha)|\!\|_{1,*} \leqslant \frac{\delta_0}{2}.$$

The symmetry $n(x) = n(-x)$ of the function n has been used implicitly as well as the fact that $(y_\alpha)_{\alpha \geqslant 1}$ is uniformly bounded away from zero by property (ii) in Lemma 2.11 and its convergence to $y_0 \in \mathbb{R}^d_{>0}$. A similar argument gives a vector $p'_\alpha \in \mathbb{R}^d_{\geqslant 0}$ with $\|\!|p'_\alpha|\!\|_{1,*} \leqslant \delta'$ as well as

$$p'_\alpha - \alpha n(x'_\alpha - y'_\alpha) \in \mathbb{R}^d_{\geqslant 0} \quad \text{and} \quad \|\!|p'_\alpha - \alpha n(x'_\alpha - y'_\alpha)|\!\|_{1,*} \leqslant \frac{\delta_0}{2}.$$

Remembering (4.84) and (4.85), and combining the non-decreasingness of the non-linearity H with the previous two displays yields

$$\mathsf{H}\big(\nabla\phi(t_\alpha, x_\alpha)\big) \leqslant \mathsf{H}\big(\nabla\phi'(t'_\alpha, x'_\alpha) + p_\alpha + p'_\alpha + \nabla\Phi(t_\alpha, x_\alpha)\big) \leqslant \mathsf{H}\big(\nabla\phi'(t'_\alpha, x'_\alpha)\big) + 2V\delta' + V\|\!|\nabla\Phi(t_\alpha, x_\alpha)|\!\|_{1,*}.$$

The second inequality implicitly uses that by (4.84)-(4.86),

$$\begin{aligned}|||\nabla\phi'(t'_\alpha,x'_\alpha)+p_\alpha+p'_\alpha+\nabla\Phi(t_\alpha,x_\alpha)|||_{1,*} &\leqslant |||\nabla\phi(t_\alpha,x_\alpha)|||_{1,*}+|||p_\alpha-\alpha n(x_\alpha-y_\alpha)|||_{1,*}+|||p'_\alpha-\alpha n(x'_\alpha-y'_\alpha)|||_{1,*}\\ &\leqslant L+\delta_0.\end{aligned}$$

It follows by (4.82) that

$$\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha)>\varepsilon+2\alpha(t_\alpha-t'_\alpha)+\partial_t\Phi(t_\alpha,x_\alpha)-\mathsf{H}\big(\nabla\phi'(t'_\alpha,x'_\alpha)\big)-2V\delta'-V|||\nabla\Phi(t_\alpha,x_\alpha)|||_{1,*}.$$

A direct computation shows that $V|||\nabla\Phi(t_\alpha,x_\alpha)|||_{1,*}\leqslant\partial_t\Phi(t_\alpha,x_\alpha)$, so in fact

$$\begin{aligned}\big(\partial_t\phi-\mathsf{H}(\nabla\phi)\big)(t_\alpha,x_\alpha)&>\varepsilon+2\alpha(t_\alpha-t'_\alpha)-\mathsf{H}\big(\nabla\phi'(t'_\alpha,x'_\alpha)\big)-2V\delta'\\ &=\big(\partial_t\phi'-\mathsf{H}(\nabla\phi')\big)(t'_\alpha,x'_\alpha)+\varepsilon-V\delta'\\ &\geqslant\varepsilon-V\delta',\end{aligned}$$

where (4.83) and the second inequality in (4.81) have been used. Choosing $\delta'<\varepsilon'/V$ contradicts the first inequality in (4.81) and completes the proof. ■

Proposition 4.11. *Fix a closed convex set $\mathcal{K}$, an initial condition $\psi:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ that is Lipschitz continuous with respect to the normalized-ℓ^1 norm, and a non-decreasing non-linearity $\mathsf{H}:\mathbb{R}^d\to\mathbb{R}$ that is locally Lipschitz continuous with respect to the normalized-$\ell^{1,*}$ norm. Denote by $f\in\mathfrak{L}_{\mathrm{unif}}$ the unique viscosity solution to the Hamilton-Jacobi equation (4.64) constructed in Proposition 4.9. If ψ has its gradient in $\mathcal{K}$, then for all $t\geqslant 0$, the function $f(t,\cdot)$ also has its gradient in $\mathcal{K}$.*

Proof. Define the set $\mathcal{A}:=\{(v,c)\in\mathbb{R}^{d+1}\mid x\cdot v\geqslant c$ for all $x\in\mathcal{K}$ and $|v|=1\}$, and recall from Proposition A.4 that $\mathcal{K}=\{x\in\mathbb{R}^d\mid x\cdot v\geqslant c$ for all $(v,c)\in\mathcal{A}\}$. For each $(v,c)\in\mathcal{A}$ introduce the closed convex cone

$$\mathcal{H}_v:=\{x\in\mathbb{R}^d\mid x\cdot v\geqslant 0\}$$

and the function $g_{v,c}(t,x):=f(t,x)-cx\cdot v$ on $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$. Define the non-linearity $\widetilde{\mathsf{H}}(p):=\mathsf{H}(p+cv\iota)$ on $\mathbb{R}^d$, where $\iota:=(1,\dots,1)\in\mathbb{R}^d$, and observe that $g_{v,c}$ satisfies the Hamilton-Jacobi equation

$$\partial_t g(t,x)=\widetilde{\mathsf{H}}\big(\nabla g(t,x)\big)\quad\text{on}\quad\mathbb{R}_{>0}\times\mathbb{R}^d_{\geqslant 0}$$

subject to the initial condition $\widetilde{\psi}(x):=\psi(x)-cx\cdot v$. Moreover, this initial condition is $\mathcal{H}_v^*$-non-decreasing. Indeed, the bi-duality result in Proposition A.1 implies that $\mathcal{H}_v^*=\mathbb{R}v$, and for any $x,x'\in\mathbb{R}^d_{\geqslant 0}$ with $x'-x=tv$ for some $t\in\mathbb{R}$,

$$g_{v,c}(x')-g_{v,c}(x)=\psi(x')-\psi(x)-c(x'-x)\cdot v\geqslant tc-tcv\cdot v=0.$$

The fact that $(x'-x)\cdot z\geqslant tc$ for all $z\in\mathcal{K}$ and the characterization of ψ having its gradient in the set $\mathcal{K}$ given in Proposition A.2 have been used. It follows by Lemma 4.10 that $g_{v,c}$ is $\mathcal{H}_v^*$-non-decreasing. At this point, fix $x,x'\in\mathbb{R}^d_{\geqslant 0}$ with the property that for all $z\in\mathcal{K}$, one has $(x'-x)\cdot z\geqslant c$. Define

$$v:=\frac{x'-x}{|x'-x|}\quad\text{and}\quad c':=\frac{c}{|x'-x|},$$

and notice that $(v,c') \in \mathcal{A}$. This means that $g_{v,c'}$ is $\mathcal{H}_v^*$-non-decreasing, and therefore

$$f(t,x') - f(t,x) = g_{v,c'}\big(t, x + |x'-x|v\big) - g_{v,c'}(t,x) + c'|x'-x|v\cdot v \geqslant c.$$

Invoking Proposition A.2 shows that f has its gradient in $\mathcal{K}$ and completes the proof. ∎

Combining Propositions 4.8, 4.9, and 4.11, it is now possible to prove Theorem 4.1.

Proof of Theorem 4.1. Denote by $\mathsf{H}_{K,R}$ the extension of the non-linearity C_K constructed in Proposition 4.8. By Proposition 4.9, the Hamilton-Jacobi equation (4.30) satisfies the comparison principle and admits a unique viscosity solution $f_R^{(K)} \in \mathfrak{L}_{\mathrm{unif}}$ subject to the initial condition $\psi^{(K)}$. Moreover, this solution satisfies the Lipschitz bound

$$\sup_{t>0} |||f_R^{(K)}(t,\cdot)|||_{\mathrm{Lip},1} = |||\psi^{(K)}|||_{\mathrm{Lip},1} \leqslant \|\psi\|_{\mathrm{Lip,TV}}, \tag{4.87}$$

where the second inequality follows from (4.29). That the function $f_R^{(K)}$ has its gradient in the set $\mathcal{K}'_{a,K}$ is immediate from assumption (**H3**) and Proposition 4.11. This completes the proof. ∎

4.2.3 Establishing convergence of the approximating solutions

With the well-posedness of the extended approximating Hamilton-Jacobi equation (4.30) at hand, it is now possible to prove the convergence of its solutions as stated in Theorem 4.2. In the notation of Theorem 4.1, given an integer $K \geqslant 1$ and some $R > 0$, write $f_R^{(K)} \in \mathfrak{L}_{\mathrm{unif}}$ for the unique solution to the Hamilton-Jacobi equation (4.30) subject to the initial condition $\psi^{(K)}$. Recall that $f_R^{(K)}$ has its gradient in the set $\mathcal{K}'_{a,K}$, and satisfies the Lipschitz bound

$$\sup_{t\geqslant 0} |||f_R^{(K)}(t,\cdot)|||_{\mathrm{Lip},1} = |||\psi^{(K)}|||_{\mathrm{Lip},1} \leqslant \|\psi\|_{\mathrm{Lip,TV}}. \tag{4.88}$$

To prove the existence of the limit

$$f_R(t,\mu) = \lim_{K\to+\infty} f_R^{(K)}\big(t, x^{(K)}(\mu)\big), \tag{4.89}$$

the arguments in Section 3.2 of [86] and Section 3.3 of [32] will be appropriately adapted. Given two integers $K' > K$, it will be convenient to introduce the projection map $P^{(K,K')} : \mathbb{R}_{\geqslant 0}^{\mathcal{D}_{K'}} \to \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$ defined by

$$P^{(K,K')}x := x^{(K)}\big(\mu_x^{(K')}\big) \tag{4.90}$$

as well as the lifting map $L^{(K,K')} : \mathbb{R}^{\mathcal{D}_K} \to \mathbb{R}^{\mathcal{D}_{K'}}$ given by

$$L^{(K,K')}x := \big(\widetilde{x}_k\big)_{k\in\mathcal{D}_K}, \tag{4.91}$$

where $\widetilde{x}_k = (x_k,\ldots,x_k) \in \mathbb{R}^{2^{K'-K}}$. A key observation in proving the existence of the limit (4.89) is that

$$P^{(K,K')}L^{(K,K')}x = x. \tag{4.92}$$

It will also be helpful to remember that by Proposition A.4, any closed convex set $\mathcal{K}$ may be written as the intersection of the closed hyper-spaces that contain it,

$$\mathcal{K} = \{x \in \mathbb{R}^d \mid x \cdot v \geqslant c \text{ for all } (v,c) \in \mathcal{A}\} \quad \text{for} \quad \mathcal{A} := \{(v,c) \in \mathbb{R}^{d+1} \mid v \cdot x \geqslant c \text{ for all } x \in \mathcal{K} \text{ and } |||v|||_1 = 1\}. \tag{4.93}$$

The following technical lemmas will also play their part. The first two translate non-differential properties of a non-differentiable function into differential properties of a smooth function at any point where the difference of these functions is locally maximal. The third analyzes the transformation of the pairs $(v,c) \in \mathcal{A}$ by the projection map (4.90), and the fourth shows that these pairs can be used to quantify the distance from a point to the closed convex set they define.

Lemma 4.12. *Let $u \in \mathfrak{L}_{\mathrm{unif}}$ be a uniformly Lipschitz function, and let $L := \sup_{t>0} |||u(t,\cdot)|||_{\mathrm{Lip},1} < +\infty$. If $\phi \in C^\infty(\mathbb{R}_{>0} \times \mathbb{R}^d_{>0};\mathbb{R})$ is a smooth function with the property that $u - \phi$ has a local maximum at the point $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, then $|||\nabla\phi(t^*,x^*)|||_{1,*} \leqslant L$. An identical statement holds at a local minimum.*

Proof. Since $u - \phi$ has a local maximum at $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, for every $\varepsilon > 0$ small enough and $x \in \mathbb{R}^d_{\geqslant 0}$,

$$\phi(t^*,x^* + \varepsilon x) - \phi(t^*,x^*) \geqslant u(t^*,x^* + \varepsilon x) - u(t^*,x^*) \geqslant -\varepsilon L |||x|||_1.$$

Dividing by ε and letting ε tend to zero reveals that

$$\nabla\phi(t^*,x^*) \cdot x \geqslant -L|||x|||_1.$$

Choosing $x_k := -d \operatorname{sgn}\big(\partial_{x_k}\phi(t^*,x^*)\big)e_k$ for each $1 \leqslant k \leqslant d$ completes the proof. ∎

Lemma 4.13. *Let $\mathcal{K}$ be a closed convex set, and let $u : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ be a Lipschitz function whose gradient is in $\mathcal{K}$. Any smooth function $\phi \in C^\infty(\mathbb{R}_{>0} \times \mathbb{R}^d_{>0};\mathbb{R})$ with the property that $u - \phi$ has a local maximum at $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$ is such that $\nabla\phi(t^*,x^*) \in \mathcal{K}$. An identical statement holds at a local minimum.*

Proof. Recall the representation (4.93) of the closed convex set $\mathcal{K}$ as the intersection of the closed and affine half-spaces which contain it, and fix $(v,c) \in \mathcal{A}$. Since $u - \phi$ has a local maximum at $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{>0}$, for every $\varepsilon > 0$ small enough,

$$\phi(t^*,x^* + \varepsilon v) - \phi(t^*,x^*) \geqslant u(t^*,x^* + \varepsilon v) - u(t^*,x^*) \geqslant \varepsilon c.$$

The second inequality combines the characterization of u having its gradient in $\mathcal{K}$ given in Proposition A.2 with the fact that $x \cdot \varepsilon v \geqslant \varepsilon c$ for all $x \in \mathcal{K}$. Dividing by ε and letting ε tend to zero reveals that

$$\nabla\phi(t^*,x^*) \cdot v \geqslant c$$

for all $(v,c) \in \mathcal{A}$. It follows that $\nabla\phi(t^*,x^*) \in \mathcal{K}$. This completes the proof. ∎

Lemma 4.14. *Let $K' > K$ be two large enough integers, and let $\mathcal{A}$ be the set associated with $\mathcal{K}'_{a,K'}$ in its representation (4.93). For every $(v,c) \in \mathcal{A}$ and $y \in \mathcal{K}'_{a,K}$,*

$$P^{(K,K')}v \cdot y \geqslant c - \frac{2}{2^{K/2}}. \tag{4.94}$$

Proof. Fix $y \in \mathcal{K}_{a,K'}$, and find vectors $u^{(K)} \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$ and $w^{(K)} \in \mathbb{R}^{\mathcal{D}_K}$ with

$$y = G^{(K)} u^{(K)} + w^{(K)}, \quad |||u^{(K)}|||_1 \leqslant a \quad \text{and} \quad |||w^{(K)}|||_{1,*} \leqslant \frac{1}{2^{K/2}}.$$

Consider the vector

$$u_{k'}^{(K')} := \frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|} u_{k'}^{(K)} \mathbf{1}\{k' \in \mathcal{D}_K\}$$

in $\mathbb{R}^{\mathcal{D}_{K'}}$, and for each $k' \in \mathcal{D}_{K'}$, write $\underline{k}'$ for the unique dyadic $\underline{k}' \in \mathcal{D}_K$ with $k' \in [\underline{k}', \underline{k}' + 2^{-K})$. Observe that

$$P^{(K,K')} v \cdot y = P^{(K,K')} v \cdot G^{(K)} u^{(K)} + P^{(K,K')} v \cdot w^{(K)} = v \cdot \left(G^{(K')} u^{(K')} + \alpha^{(K')}\right) + P^{(K,K')} v \cdot w^{(K)}$$

for the vector $\alpha^{(K')} \in \mathbb{R}^{\mathcal{D}_{K'}}$ defined by

$$\alpha_{k'}^{(K')} := \frac{1}{|\mathcal{D}_{K'}|^2} \sum_{k' \in \mathcal{D}_{K'}} \left(g(\underline{k}k') - g(kk')\right) u_{k'}^{(K')}.$$

Since $G^{(K')} u^{(K')} \in \mathcal{K}'_{a,K'}$, the defining property of v, the fact that $|||v|||_1 = 1$, and Hölder's inequality give the lower bound

$$P^{(K,K')} v \cdot y \geqslant c - |||\alpha^{(K')}|||_{1,*} - |||w^{(K)}|||_{1,*} \geqslant c - |||\alpha^{(K')}|||_{1,*} - \frac{1}{2^{K/2}}, \tag{4.95}$$

where it has been used that $|||P^{(K,K')} v|||_1 \leqslant |||v|||_1$ and $|||w^{(K)}|||_{1,*} \leqslant 2^{-K/2}$. The mean value theorem reveals that

$$|||\alpha^{(K')}|||_{1,*} \leqslant \frac{\|g'\|_{L^\infty}}{2^K} |||u^{(K)}|||_1 \leqslant \frac{a\|g'\|_{L^\infty}}{2^K}.$$

Substituting this into (4.95) and taking K large enough completes the proof. ■

Lemma 4.15. *Let $\mathcal{K}$ be a closed convex set, and let $\mathcal{A}$ be the set associated with $\mathcal{K}$ in its representation* (4.93). *If $x \in \mathbb{R}^d$ and $\varepsilon > 0$ are such that $x \cdot v \geqslant c - \varepsilon$ for all $(v,c) \in \mathcal{A}$, then there exist $y \in \mathcal{K}$ and $z \in \mathbb{R}^d$ with $x = y + z$ and $|||z|||_{1,*} \leqslant \varepsilon$.*

Proof. Let $y \in \mathcal{K}$ denote a projection of $x \in \mathbb{R}^d$ onto the set $\mathcal{K}$ with respect to the normalized-$\ell^{1,*}$ norm. More precisely, let $y \in \mathcal{K}$ be any minimizer of the map $y' \mapsto |||y' - x|||_{1,*}$ over points $y' \in \mathcal{K}$. The existence of such a projection is guaranteed by the fact that $\mathcal{K}$ is closed. If $y = x$, then the desired conclusion is immediate, so from now on assume that $y \neq x$. Introduce the set

$$\mathcal{I} := \left\{k \leqslant d \mid d|x_k - y_k| = |||x - y|||_{1,*}\right\}$$

of indices at which $|||x - y|||_{1,*}$ is achieved, and define the vector $v \in \mathbb{R}^d$ by

$$v_k := \frac{d}{|\mathcal{I}|} \operatorname{sgn}(y_k - x_k) \mathbf{1}\{k \in \mathcal{I}\}.$$

It is now shown that $(v,c) \in \mathcal{A}$ for $c := v \cdot y$. By construction $|||v|||_1 = 1$, so suppose for the sake of contradiction that there exists $y' \in \mathcal{K}$ with $(y' - y) \cdot v = y' \cdot v - c < 0$. This means that

$$(y - y') \cdot v = \frac{d}{|\mathcal{I}|} \sum_{k \in \mathcal{I}} (y_k - y'_k) \operatorname{sgn}(y_k - x_k) > 0.$$

In particular, a coordinate $k^* \in \mathcal{I}$ at which the quantity $(y_k - y'_k)\operatorname{sgn}(y_k - x_k)$ is maximized over $k \in \mathcal{I}$ must satisfy

$$(y_{k^*} - y'_{k^*})\operatorname{sgn}(y_{k^*} - x_{k^*}) > 0.$$

At this point, fix $t \in (0,1)$ small enough so that $\operatorname{sgn}(y_k - x_k + t(y'_k - y_k)) = \operatorname{sgn}(y_k - x_k)$ for every $k \in \mathcal{I}$. For such a value of $t > 0$,

$$|||y - x + t(y' - y)|||_{1,*} = d(y_{k^*} - x_{k^*} + t(y'_{k^*} - y_{k^*}))\operatorname{sgn}(y_{k^*} - x_{k^*}) < |||x - y|||_{1,*}.$$

Since the point $y'' = y + t(y' - y)$ is a convex combination of $y, y' \in \mathcal{K}$, it must lie in the convex set $\mathcal{K}$. This contradicts the fact that y minimizes the map $y'' \mapsto |||y'' - x|||_{1,*}$ over points $y'' \in \mathcal{K}'$, and shows that $(v,c) \in \mathcal{A}$. It follows that

$$\varepsilon \geqslant c - x \cdot v = v \cdot (y - x) = |||x - y|||_{1,*}.$$

Setting $z = x - y$ completes the proof. ■

Proof of Theorem 4.2. To alleviate notation, until otherwise stated, fix $R > \|\psi\|_{\mathrm{Lip,TV}}$ and keep all dependencies on R implicit. The existence of the limit (4.89) will be established by showing that the sequence $(f^{(K)}(t, x^{(K)}(\mu)))_{K \geqslant 1}$ is Cauchy. With this in mind, fix $K' > K$, and introduce the function

$$f^{(K,K')}(t,x) := f^{(K)}\big(t, P^{(K,K')}x\big) \tag{4.96}$$

defined on $\mathbb{R}_{\geqslant 0} \times \mathbb{R}^{\mathcal{D}_{K'}}_{\geqslant 0}$. Since $x^{(K)}(\mu) = P^{(K,K')}x^{(K')}(\mu)$, the Cauchy condition may be expressed in terms of this function as

$$\big|f^{(K')}\big(t, x^{(K')}(\mu)\big) - f^{(K)}\big(t, x^{(K)}(\mu)\big)\big| = \big|f^{(K')}\big(t, x^{(K')}(\mu)\big) - f^{(K,K')}\big(t, x^{(K')}(\mu)\big)\big|. \tag{4.97}$$

The right side of this expression is controlled in two steps. First $f^{(K,K')}$ is shown to be an approximate viscosity solution to the Hamilton-Jacobi equation (4.30) satisfied by $f^{(K')}$, and then the comparison principle in Theorem 4.1 is leveraged. The final step of the proof is to show that the limit does not depend on the choice of $R > \|\psi\|_{\mathrm{Lip,TV}}$.

Step 1: showing $f^{(K,K')}$ is an approximate viscosity solution. Consider a function $\phi_{K'} \in C^\infty\big(\mathbb{R}_{>0} \times \mathbb{R}^{\mathcal{D}_{K'}}_{>0}; \mathbb{R}\big)$ with the property that $f^{(K,K')} - \phi_{K'}$ achieves a local maximum at $(t^*, x^*) \in \mathbb{R}_{>0} \times \mathbb{R}^{\mathcal{D}_{K'}}_{>0}$. To be more precise, suppose that

$$\sup_{B_{K'}(r)} \big(f^{(K,K')} - \phi_{K'}\big) = \big(f^{(K,K')} - \phi_{K'}\big)(t^*, x^*),$$

where

$$B_{K'}(r) := \Big\{(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}^{\mathcal{D}_{K'}}_{\geqslant 0} \mid |t - t^*| + |||x - x^*|||_1 \leqslant r\Big\}$$

is the ball of radius $r > 0$ centred at (t^*, x^*). Decreasing $r > 0$ if necessary, assume without loss of generality that

$$B_{K'}(r) \subseteq \mathbb{R}_{>0} \times \mathbb{R}^{\mathcal{D}_{K'}}_{>0}.$$

Assume also that $\phi_{K'} \in C^\infty\big(\mathbb{R}_{>0} \times \mathbb{R}^{\mathcal{D}_{K'}}; \mathbb{R}\big)$; this can be ensured by replacing $\phi_{K'}$ with $\eta\phi_{K'}$ for some $\eta \in C^\infty\big(\mathbb{R}^{\mathcal{D}_{K'}}; \mathbb{R}\big)$ which is identically one on $B_{K'}(r)$ and vanishes outside $\mathbb{R}^{\mathcal{D}_{K'}}_{\geqslant 0}$. With these simplifications at hand, introduce the smooth function

$$\phi_K(t,y) := \phi_{K'}\big(t, x^* + L^{(K,K')}y - L^{(K,K')}P^{(K,K')}x^*\big)$$

defined on $\mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K}$. It will now be shown that the function ϕ_K admits a local maximum at $(t^*, P^{(K,K')}x^*)$. It will be convenient to notice that by (4.92), for any $y \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$,

$$P^{(K,K')}\big(x^* + L^{(K,K')}y - L^{(K,K')}P^{(K,K')}x^*\big) = P^{(K,K')}x^* + y - P^{(K,K')}x^* = y \in \mathbb{R}_{>0}^{\mathcal{D}_K}. \tag{4.98}$$

To simplify notation, let $y^* := P^{(K,K')}x^* \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$, and introduce the ball

$$B_K(r) := \big\{(t,y) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K} \mid |t-t^*| + |||y-y^*|||_1 \leqslant r\big\} \subseteq \mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K}$$

of radius $r>0$ centred at (t^*, y^*). Given $(t,y) \in B_K(r)$, let $z_y := x^* + L^{(K,K')}y - L^{(K,K')}P^{(K,K')}x^*$ in such a way that by (4.98),

$$f^{(K)}(t,y) - \phi_K(s,y) = f^{(K,K')}(t,z_y) - \phi_{K'}(t,z_y).$$

Observe that

$$|t-t^*| + |||z_y - x^*|||_1 = |t-t^*| + |||L^{(K,K')}y - L^{(K,K')}y^*|||_1 = |t-t^*| + |||y-y^*|||_1 \leqslant r$$

so $(t,z_y) \in B_{K'}(r)$. It follows that

$$\sup_{B_K(r)}\big(f^{(K)} - \phi_K\big) \leqslant \sup_{B_{K'}(r)}\big(f^{(K,K')} - \phi_{K'}\big) = \big(f^{(K,K')} - \phi_{K'}\big)(t^*,x^*) = \big(f^{(K)} - \phi_K\big)(t^*,y^*)$$

which means that ϕ_K admits a local maximum at $(t^*,y^*) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K}$. Since $f^{(K)}$ is a viscosity subsolution to the Hamilton-Jacobi equation (4.30) and its gradient belongs to the closed convex set $\mathcal{K}'_{a,K}$, Lemma 4.13 implies that

$$\nabla\phi_K(t^*,y^*) \in \mathcal{K}'_{a,K} \quad \text{and} \quad \big(\partial_t\phi_K - \mathsf{H}_K(\nabla\phi_K)\big)(t^*,y^*) \leqslant 0.$$

To write this expression in terms of the original test function $\phi_{K'}$, notice that

$$\partial_t\phi_K(t^*,y^*) = \partial_t\phi_{K'}(t^*,x^*) \quad \text{and} \quad \nabla\phi_K(t^*,y^*) = \frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|}P^{(K,K')}\nabla\phi_{K'}(t^*,x^*).$$

This means that

$$\frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|}P^{(K,K')}\nabla\phi_{K'}(t^*,x^*) \in \mathcal{K}'_{a,K} \quad \text{and} \quad \partial_t\phi_{K'}(t^*,x^*) - \mathsf{H}_K\bigg(\frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|}P^{(K,K')}\nabla\phi_{K'}(t^*,x^*)\bigg) \leqslant 0.$$

The first of these conditions gives vectors $u^{(K)} \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$ and $w^{(K)} \in \mathbb{R}^{\mathcal{D}_K}$ with

$$\frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|}P^{(K,K')}\nabla\phi_{K'}(t^*,x^*) = G^{(K)}u^{(K)} + w^{(K)}, \quad |||u^{(K)}|||_1 \leqslant a, \quad \text{and} \quad |||w^{(K)}|||_{1,*} \leqslant \frac{1}{2^{K/2}}. \tag{4.99}$$

Observe that

$$|||G^{(K)}u^{(K)}|||_{1,*} \leqslant \frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|}|||P^{(K,K')}\nabla\phi_{K'}(t^*,x^*)|||_{1,*} + \frac{1}{2^{K/2}} \leqslant |||\nabla\phi_{K'}(t^*,x^*)|||_{1,*} + \frac{1}{2^{K/2}}.$$

Since $f^{(K,K')} - \phi_{K'}$ achieves a local maximum at (t^*,x^*), Lemma 4.12 and (4.88) imply that

$$|||\nabla\phi_{K'}(t^*,x^*)|||_{1,*} \leqslant \|\psi\|_{\mathrm{Lip,TV}}.$$

Recalling that $R > \|\psi\|_{\mathrm{Lip,TV}}$, and taking K large enough ensures that $G^{(K)}u^{(K)} \in \mathcal{C}_K \cap B_R$. It follows by the Lipschitz continuity of H_K established in Proposition 4.8 that

$$\partial_t\phi_{K'}(t^*,x^*) - \mathsf{C}_K\big(G^{(K)}u^{(K)}\big) \leqslant \partial_t\phi_{K'}(t^*,x^*) - \mathsf{H}_K\left(\frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|}P^{(K,K')}\nabla\phi_{K'}(t^*,x^*)\right) + \frac{8RM}{2^{K/2}m^2} \leqslant \frac{8RM}{2^{K/2}m^2}.$$

At this point, introduce the vector $u^{(K')} \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_{K'}}$ defined by

$$u_{k'}^{(K')} := \frac{|\mathcal{D}_{K'}|}{|\mathcal{D}_K|}u_{k'}^{(K)}\mathbf{1}\{k' \in \mathcal{D}_K\}$$

in such a way that

$$\mathsf{C}_{K'}\big(G^{(K')}u^{(K')}\big) = \frac{1}{|\mathcal{D}_K|^2}\sum_{k,k'\in\mathcal{D}_K} g(kk')u_k^{(K)}u_{k'}^{(K)} = \mathsf{C}_K\big(G^{(K)}u^{(K)}\big),$$

and therefore,

$$\partial_t\phi_{K'}(t^*,x^*) - \mathsf{C}_{K'}\big(G^{(K')}u^{(K')}\big) \leqslant \frac{8RM}{2^{K/2}m^2}. \tag{4.100}$$

It is now shown that, up to an error vanishing with K, the term $G^{(K')}u^{(K')}$ in this expression may be replaced by $\nabla\phi_{K'}(t^*,x^*)$. This is where Lemma 4.14 will play its part. Let $\mathcal{A}$ be the set associated with $\mathcal{K}'_{a,K'}$ in its representation (4.93), and fix $(v,c) \in \mathcal{A}$. The characterization of $f^{(K)}$ having its gradient in $\mathcal{K}'_{a,K}$ given in Proposition A.2 and Lemma 4.14 imply that for every $\varepsilon > 0$ small enough,

$$\phi_{K'}(t^*,x^*+\varepsilon v) - \phi_{K'}(t^*,x^*) \geqslant f^{(K)}\big(t^*,P^{(K,K')}x^* + \varepsilon P^{(K,K')}v\big) - f^{(K)}\big(t^*,P^{(K,K')}x^*\big) \geqslant \varepsilon\left(c - \frac{2}{2^{K/2}}\right)$$

Dividing by ε and letting ε tend to zero reveals that $\nabla\phi_{K'}(t^*,x^*)\cdot v \geqslant c - \frac{2}{2^{K/2}}$. Invoking Lemma 4.15 gives $\alpha^{(K')} \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_{K'}}$ and $\beta^{(K')} \in \mathbb{R}^{\mathcal{D}_{K'}}$ with

$$\nabla\phi_{K'}(t^*,x^*) = G^{(K')}\alpha^{(K')} + \beta^{(K')}, \quad |||\alpha^{(K')}|||_1 \leqslant a, \quad \text{and} \quad |||\beta^{(K')}|||_{1,*} \leqslant \frac{2}{2^{K/2}}. \tag{4.101}$$

At this point, fix $k \in \mathcal{D}_K$ and $k' \in [k, k+2^{-K})$. The mean value theorem implies that

$$\begin{aligned}|\mathcal{D}_{K'}|\big|\big(G^{(K')}u^{(K')}\big)_{k'} - \partial_{x_{k'}}\phi_{K'}(t^*,x^*)\big| &= \left|\frac{1}{|\mathcal{D}_K|}\sum_{k''\in\mathcal{D}_K} g(k'k'')u_{k''}^{(K)} - |\mathcal{D}_{K'}|\partial_{x_{k'}}\phi_{K'}(t^*,x^*)\right| \\ &\leq \big||\mathcal{D}_K|\big(G^{(K)}u^{(K)}\big)_k - |\mathcal{D}_{K'}|\partial_{x_{k'}}\phi_{K'}(t^*,x^*)\big| + \frac{\|g'\|_{L^\infty}|||u^{(K)}|||_1}{2^K}.\end{aligned}$$

Remembering (4.99) and (4.101), noticing that $|\mathcal{D}_K| = 2^{K+1}$, and using the mean value theorem once again shows that

$$
\begin{aligned}
|\mathcal{D}_K|\big(G^{(K)}u^{(K)}\big)_k &= |\mathcal{D}_K|\sum_{\ell=0}^{2^{K'-K}-1}\partial_{x_{k+\frac{\ell}{2^{K'}}}}\phi_{K'}(t^*,x^*)-|\mathcal{D}_K|w_k^{(K)}\\
&= \frac{|\mathcal{D}_K|}{|\mathcal{D}_{K'}|^2}\sum_{\ell=0}^{2^{K'-K}-1}\sum_{k''\in\mathcal{D}_{K'}}g\Big(k+\frac{\ell}{2^{K'}}\cdot k''\Big)\alpha_{k''}^{(K')}+|\mathcal{D}_K|\sum_{\ell=0}^{2^{K'-K}-1}\beta^{(K')}_{k+\frac{\ell}{2^{K'}}}-|\mathcal{D}_K|w_k^{(K)}\\
&= \frac{|\mathcal{D}_K|}{|\mathcal{D}_{K'}|^2}\sum_{\ell=0}^{2^{K'-K}-1}\sum_{k''\in\mathcal{D}_{K'}}g(k'k'')\alpha_{k''}^{(K')}+\mathcal{O}_1\left(\frac{\|g'\|_{L^\infty}|||\alpha^{(K')}|||_1}{2^K}+\frac{3}{2^{K/2}}\right)\\
&= |\mathcal{D}_{K'}|\big(G^{(K')}\alpha^{(K')}\big)_{k'}+\mathcal{O}\left(\frac{4}{2^{K/2}}\right)\\
&= |\mathcal{D}_{K'}|\partial_{x_{k'}}\phi_{K'}(t^*,x^*)+\mathcal{O}\left(\frac{5}{2^{K/2}}\right).
\end{aligned}
$$

The third equality used that $|||\beta^{(K')}|||_{1,*}+|||w^{(K)}|||_{1,*}\leqslant 3\cdot 2^{-K/2}$, while the fourth equality used that $|||\alpha^{(K')}|||_1\leqslant a$ and increases K if necessary. Together with the fact that $|||u^{(K)}|||_1\leqslant a$, and increasing K if necessary, this implies that

$$
|||G^{(K')}u^{(K')}-\nabla\phi_{K'}(t^*,x^*)|||_{1,*}\leqslant\frac{5}{2^{K/2}}+\frac{a\|g'\|_{L^\infty}}{2^K}\leqslant\frac{6}{2^{K/2}}.
$$

Combining this with the Lipschitz continuity of $\mathsf{H}_{K'}$ established in Proposition 4.8 and with (4.100) reveals that

$$
\partial_t\phi_{K'}(t^*,x^*)-\mathsf{H}_{K'}\big(\nabla\phi_{K'}(t^*,x^*)\big)\leqslant\mathcal{E}_K\qquad\text{for the error term}\qquad\mathcal{E}_K:=\frac{56RM}{2^{K/2}m^2}.
$$

In particular, the function $(t,x)\mapsto f^{(K,K')}(t,x)-\mathcal{E}_Kt$ is a viscosity subsolution to the Hamilton-Jacobi equation (4.30) satisfied by $f^{(K')}$. An identical argument shows that $(t,x)\mapsto f^{(K,K')}(t,x)+\mathcal{E}_Kt$ is a viscosity supersolution to the Hamilton-Jacobi equation (4.30) satisfied by $f^{(K')}$.

Step 2: leveraging the comparison principle. Using (4.88) and (4.29), it is readily verified that $f^{(K,K')}$ and $f^{(K')}$ are uniformly Lipschitz continuous in the spatial variable relative to the normalized-ℓ^1 norm with Lipschitz constant at most $L:=\|\psi\|_{\mathrm{Lip,TV}}$. Indeed, for any $t>0$ and all $x,x'\in\mathbb{R}_{\geqslant 0}^{\mathcal{D}_{K'}}$,

$$
|f^{(K,K')}(t,x)-f^{(K,K')}(t,x')|\leqslant\|\psi\|_{\mathrm{Lip,TV}}|||P^{(K,K')}x-P^{(K,K')}x'|||_1\leqslant\|\psi\|_{\mathrm{Lip,TV}}|||x-x'|||_1.
$$

If $V:=|||\mathsf{H}_K|||_{\mathrm{Lip},1,*}$, then the comparison principle in Theorem 4.1 implies that for any $R'\in\mathbb{R}$, the map

$$
(t',x')\mapsto f^{(K,K')}(t',x')-f^{(K')}(t',x')-(2L+1)\big(|||x'|||_1+Vt'-R'\big)_+-\mathcal{E}_Kt' \tag{4.102}
$$

achieves its supremum on $\{0\}\times\mathbb{R}_{\geqslant 0}^{\mathcal{D}_{K'}}$. Set $R':=|||x^{(K')}(\mu)|||_1+Vt$ and distinguish two cases. On the one hand, if $t'=0$ and $|||x'|||_1\geqslant(2L+1)R'$, then (4.102) is bounded by

$$
2L|||x'|||_1-(2L+1)(|||x'|||_1-R')=R'-|||x'|||_1\leqslant 0, \tag{4.103}
$$

where it has been used that $f^{(K,K')}(0,0)=f^{(K')}(0,0)$. On the other hand, if $t'=0$ and $|||x'|||_1\leqslant(2L+1)R'$, then (**H4**) implies that (4.102) is bounded by

$$
\big|\psi^{(K')}(x')-\psi^{(K)}(P^{(K,K')}x')\big|\leqslant\|\psi\|_{\mathrm{Lip},W}|||x'|||_1W\big(\overline{\mu}_{x'}^{(K')},\overline{\mu}^{(K)}_{P^{(K,K')}x'}\big). \tag{4.104}
$$

To estimate this Wasserstein distance, fix a Lipschitz function $h:[-1,1]\to\mathbb{R}$ with $\|h\|_{\mathrm{Lip}}\leqslant 1$ and observe that

$$\Big|\int_{-1}^{1} h(y)\,\mathrm{d}\big(\mu_{x'}^{(K')}-\mu_{P(K,K')x'}^{(K)}\big)(y)\Big| \leqslant \frac{1}{|\mathcal{D}_{K'}|}\sum_{k\in\mathcal{D}_K}\sum_{\ell=0}^{2^{K'-K-1}} x'_{k+\frac{\ell}{2^{K'}}}\Big|h\Big(k+\frac{\ell}{2^{K'}}\Big)-h(k)\Big|$$
$$\leqslant \frac{1}{|\mathcal{D}_{K'}|}\sum_{k\in\mathcal{D}_K}\sum_{\ell=0}^{2^{K'-K-1}} x'_{k+\frac{\ell}{2^{K'}}}\frac{\ell}{2^{K'}} \leqslant \frac{|||x'|||_1}{2^K}.$$

Taking the supremum over all such h and recalling (4.104) shows that (4.102) is bounded by

$$\frac{\|\psi\|_{\mathrm{Lip},W}(2L+1)R'}{2^K} \tag{4.105}$$

whenever $t'=0$ and $|||x'|||_1\leqslant(2L+1)R'$. Combining this with (4.103) reveals that the map (4.102) is uniformly bounded by (4.105). Choosing $t'=t$ and $x'=x^{(K')}(\mu)$, and recalling the choice of R' yields

$$f^{(K,K')}\big(t,x^{(K')}(\mu)\big)-f^{(K')}\big(t,x^{(K')}(\mu)\big)\leqslant \frac{\|\psi\|_{\mathrm{Lip},W}(2L+1)}{2^K}\big(|||x^{(K')}(\mu)|||_1+Vt\big)+\mathcal{E}_K t.$$

Together with (4.97) and an identical argument with the roles of $f^{(K,K')}$ and $f^{(K')}$ reversed, this implies that

$$\big|f^{(K')}\big(t,x^{(K')}(\mu)\big)-f^{(K)}\big(t,x^{(K)}(\mu)\big)\big|\leqslant \frac{\|\psi\|_{\mathrm{Lip},W}(2L+1)}{2^K}\big(\mu[-1,1]+Vt\big)+\mathcal{E}_K t.$$

Since $V=|||\mathsf{H}_K|||_{\mathrm{Lip},1,*}$ is independent of K by Proposition 4.8 and $\mathcal{E}_K$ tends to zero as K tends to infinity, the sequence $(f^{(K)}(t,x^{(K)}(\mu)))_{K\geqslant 1}$ is Cauchy. This establishes the existence of the limit (4.89) for each fixed $R>\|\psi\|_{\mathrm{Lip,TV}}$. All that remains is to show that this limit is independent of R.

Step 3: establishing independence on R. To show that the limit (4.89) is independent of R, fix $R'>R>\|\psi\|_{\mathrm{Lip,TV}}$ as well as $K\geqslant 1$ large enough. The idea will be to show that, up to an error vanishing with K, the function $f_R^{(K)}$ satisfies the Hamilton-Jacobi equation defining $f_{R'}^{(K)}$. The equality of the limit (4.89) associated with R and R' will then follow from the comparison principle in Theorem 4.1. Consider $\phi\in C^\infty\big(\mathbb{R}_{>0}\times\mathbb{R}_{>0}^{\mathcal{D}_K};\mathbb{R}\big)$ with the property that $f_R^{(K)}-\phi$ achieves a local maximum at the point $(t^*,x^*)\in\mathbb{R}_{>0}\times\mathbb{R}_{>0}^{\mathcal{D}_K}$. Since $f_R^{(K)}$ is a viscosity subsolution to the Hamilton-Jacobi equation (4.30) associated with the non-linearity $\mathsf{H}_{K,R}$,

$$\big(\partial_t\phi-\mathsf{H}_{K,R}(\nabla\phi)\big)(t^*,x^*)\leqslant 0.$$

The fact that $f_R^{(K)}$ has its gradient in the set $\mathcal{K}'_{a,K}$ together with (4.88), Lemma 4.12 and Lemma 4.13 implies that

$$\nabla\phi(t^*,x^*)\in\mathcal{K}'_{a,K}\quad\text{and}\quad |||\nabla\phi(t^*,x^*)|||_{1,*}\leqslant\|\psi\|_{\mathrm{Lip,TV}}.$$

It is therefore possible to find $u\in\mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$ and $w\in\mathbb{R}^{\mathcal{D}_K}$ with

$$\nabla\phi(t^*,x^*)=G^{(K)}u+w,\quad |||u|||_1\leqslant a,\quad\text{and}\quad |||w|||_{1,*}\leqslant\frac{1}{2^{K/2}}.$$

Observe that

$$|||G^{(K)}u|||_{1,*}\leqslant|||\nabla\phi(t^*,x^*)|||_{1,*}+|||w|||_{1,*}\leqslant\|\psi\|_{\mathrm{Lip,TV}}+\frac{1}{2^{K/2}},$$

so increasing K if necessary, it is possible to ensure that $G^{(K)}u\in\mathcal{C}_K\cap B_R\subseteq\mathcal{C}_K\cap B_{R'}$. It follows by the Lipschitz continuity of $\mathsf{H}_{K,R}$ established in Proposition 4.8 that

$$\begin{aligned}\big(\partial_t\phi - \mathsf{H}_{K,R'}(\nabla\phi)\big)(t^*,x^*) &\leqslant \partial_t\phi(t^*,x^*) - \mathsf{C}_K\big(G^{(K)}u\big) + \frac{8R'M}{2^{K/2}m^2}\\ &\leqslant \big(\partial_t\phi - \mathsf{H}_{K,R}(\nabla\phi)\big)(t^*,x^*) + \frac{8(R'+R)M}{2^{K/2}m^2} \leqslant \mathcal{E}_K\end{aligned}$$

for the error term

$$\mathcal{E}_K := \frac{8(R'+R)M}{2^{K/2}m^2}.$$

In particular, the function $(t,x)\mapsto f_R^K(t,x) - \mathcal{E}_K t$ is a viscosity subsolution to the Hamilton-Jacobi equation (4.30) defining $f_{R'}^K$. An identical argument shows that $(t,x)\mapsto f_R^K(t,x)+\mathcal{E}_K t$ is a viscosity supersolution to the Hamilton-Jacobi equation (4.30) defining $f_{R'}^K$. It follows by the comparison principle in Theorem 4.1 that for every $\mu\in\mathcal{M}_+$ and $t\geqslant 0$,

$$\big|f_R^{(K)}\big(t,x^{(K)}(\mu)\big) - f_{R'}^{(K)}\big(t,x^{(K)}(\mu)\big)\big| \leqslant \mathcal{E}_K t.$$

Letting K tend to infinity completes the proof. ∎

In addition to this well-posedness result, to recover the variational formula for the limit mutual information in the disassortative sparse stochastic block model stated in Theorem 1.7, a Hopf-Lax variational formula for the unique solution to the infinite-dimensional Hamilton-Jacobi equation (4.6) with convex non-linearity will be required.

4.3 Revisiting the Hopf-Lax formula

The Hopf-Lax variational formula for the infinite-dimensional Hamilton-Jacobi equation (4.6) stated in Theorem 4.3 will be obtained by first establishing an approximate Hopf-Lax variational formula for each of the extended approximating Hamilton-Jacobi equations (4.30), and then letting the dimension parameter $K\geqslant 1$ tend to infinity. The approximate Hopf-Lax formula for the Hamilton-Jacobi equation (4.30) will be obtained in Section 4.3.1 by combining the arguments in Section 2.4.4 with the comparison principle in Proposition 4.9, while the limiting argument leading to the infinite-dimensional Hopf-Lax formula will be presented in Section 4.3.2, and will rely on Theorem 4.2.

4.3.1 Approximate Hopf-Lax formula for approximating Hamilton-Jacobi equations

To alleviate notation and strive for generality, the approximate Hopf-Lax formula for the approximating Hamilton-Jacobi equation (4.30) will be discussed in the context of Section 4.2.1. More precisely, fix a symmetric matrix $G\in\mathbb{R}^{d\times d}$ whose components satisfy the bounds (4.54), and recall the definition of the cone $\mathcal{C}$ in (4.55), and of the non-linearity C in (4.56). Given $R>0$, denote by $\mathsf{H}_R:\mathbb{R}^d\to\mathbb{R}$ the extension of the non-linearity C constructed in Proposition 4.8. For each $a>0$, in analogy to (4.21), define the set

$$\mathcal{K}_a := \{Gx\in\mathbb{R}^d \mid x\in\mathbb{R}^d \text{ and } |||x|||_1\leqslant a\}. \tag{4.106}$$

Assuming that the original non-linearity C is convex, an approximate Hopf-Lax formula will be established for the extended Hamilton-Jacobi equation

$$\partial_t f(t,x) = \mathsf{H}_R\big(\nabla f(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^d_{>0} \tag{4.107}$$

subject to an initial condition $\psi : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ that is Lipschitz continuous with respect to the normalized-ℓ^1 norm and has its gradient in the set $\mathcal{K}'_a$. In this context, an approximate Hopf-Lax formula means a variational formula whose difference with the solution to the Hamilton-Jacobi equation (4.6) constructed in Theorem 4.2 vanishes as the dimension d tends to infinity. The convexity of the non-linearity C will be encoded by the additional assumption that the matrix G is non-negative definite. It will be convenient to introduce the bilinear form

$$(x,y)_G := Gx \cdot y \tag{4.108}$$

associated with the non-negative definite matrix G, as well as its induced semi-norm

$$\|x\|_G := \sqrt{(x,x)_G}. \tag{4.109}$$

In this notation, the non-linearity (4.56) may be written as

$$\mathsf{C}(Gx) = \frac{1}{2} Gx \cdot x = \frac{1}{2}\|x\|_G^2. \tag{4.110}$$

In particular, the non-linearity (4.56) is a convex function. Together with the fact that by Proposition 4.11 the gradient of the solution to the Hamilton-Jacobi equation (4.107) stays in the set $\mathcal{K}'_a$, and therefore that this solution should be close to that of the Hamilton-Jacobi equation (4.107) with the non-linearity H_R replaced by the non-linearity C, this suggests that an approximate Hopf-Lax formula should hold.

To motivate this Hopf-Lax formula, suppose temporarily that the matrix G is invertible. The non-linearity C then admits a natural convex extension $\widetilde{\mathsf{C}} : \mathbb{R}^d \to \mathbb{R}$ defined by

$$\widetilde{\mathsf{C}}(y) := \frac{1}{2} G^{-1} y \cdot y. \tag{4.111}$$

A direct computation reveals that $\widetilde{\mathsf{C}}^*(z) = \frac{1}{2}\|z\|_G^2$ for all $z \in \mathbb{R}^d_{\geqslant 0}$. The Hopf-Lax formula in Proposition 2.10 therefore suggests that the unique solution to the Hamilton-Jacobi equation (4.107) with the non-linearity H_R replaced by the non-linearity $\widetilde{\mathsf{C}}$ should be the function $\widetilde{f} : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ defined by

$$\widetilde{f}(t,x) := \sup_{y \geqslant 0} \left(\psi(x+y) - t\widetilde{\mathsf{C}}^*\left(\frac{y}{t}\right) \right) = \sup_{y \geqslant 0} \left(\psi(x+y) - \frac{\|y\|_G^2}{2t} \right). \tag{4.112}$$

Since the gradients of the solutions to the Hamilton-Jacobi equations (4.107) with non-linearities H_R and $\widetilde{\mathsf{C}}$ are expected to remain in the set $\mathcal{K}'_a$ by Proposition 4.11, and these non-linearities agree on the set $\mathcal{K}_a$, the solutions to these equations should be close. This suggests that the function $\widetilde{f}$ should define an approximate Hopf-Lax formula for the solution to the Hamilton-Jacobi equation (4.107). Of course, this argument is merely formal since the assumptions of Proposition 2.10 are not satisfied by the initial condition ψ, and the matrix G is not invertible. Nonetheless, it motivates the introduction of the Hopf-Lax function $f_{\mathrm{HL}} : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ defined by

$$f_{\mathrm{HL}}(t,x) := \sup_{y \in \mathbb{R}^d_{\geqslant 0}} \left(\psi(x+y) - \frac{\|y\|_G^2}{2t} \right). \tag{4.113}$$

This function will now be shown to give an approximate Hopf-Lax formula for the Hamilton-Jacobi equation (4.107).

Proposition 4.16. *Let $\psi : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ be an initial condition that is Lipschitz continuous with respect to the* *normalized-ℓ^1 norm and has its gradient in the set $\mathcal{K}'_a$. Given $R > |||\psi|||_{\mathrm{Lip},1}$, denote by $f : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0} \to \mathbb{R}$ the unique solution to the Hamilton-Jacobi equation (4.107). If the matrix $G \in \mathbb{R}^{d\times d}$ is non-negative definite, then for all $(t,x) \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$,*

$$|f(t,x) - f_{\mathrm{HL}}(t,x)| \leqslant \frac{t}{\sqrt{d}}\left(R + a + \frac{8RM}{m^2}\right). \tag{4.114}$$

The proof of this result closely follows that of the Hopf-Lax formula in Proposition 2.10. First, it is shown that the convex dual of the norm function $y \mapsto \frac{1}{2}\|y\|_G^2$ is the non-linearity C. Then, it is verified that the Hopf-Lax function (4.113) satisfies the right initial condition and that the supremum in its definition is attained. Subsequently, it is shown that the Hopf-Lax function satisfies a semigroup property from which it is deduced that it belongs to the solution space $\mathfrak{L}_{\mathrm{unif}}$. Finally, it is shown that, in a sense to be made precise, the Hopf-Lax function is an approximate solution to the Hamilton-Jacobi equation (4.107). The comparison principle in Proposition 4.9 is then leveraged to establish Proposition 4.16. It will be convenient to note that for every $z \in \mathbb{R}^d_{\geqslant 0}$,

$$\|z\|_G^2 \geqslant \frac{m}{d^2}\sum_{k,k'=1}^d z_k z_{k'} = m|||z|||_1^2. \tag{4.115}$$

Lemma 4.17. *If $G \in \mathbb{R}^{d\times d}$ is non-negative definite, then for every $z \in \mathcal{C}$,*

$$\mathsf{C}(z) = \sup_{y\in\mathbb{R}^d_{\geqslant 0}}\left(y\cdot z - \frac{\|y\|_G^2}{2}\right). \tag{4.116}$$

Moreover, the supremum is attained at any point $x \in \mathbb{R}^d_{\geqslant 0}$ with $z = Gx$.

Proof. Given $z \in \mathcal{C}$, let $x \in \mathbb{R}^d_{\geqslant 0}$ be such that $z = Gx$. Using that G is non-negative definite, appeal to the Cauchy-Schwarz inequality to assert that

$$y\cdot Gx = (x,y)_G \leq \|x\|_G\,\|y\|_G \leq \frac{1}{2}\|x\|_G^2 + \frac{1}{2}\|y\|_G^2.$$

It follows that

$$\mathsf{C}(Gx) = \frac{1}{2}\|x\|_G^2 \geqslant \sup_{y\in\mathbb{R}^d_{\geqslant 0}}\left(y\cdot Gx - \frac{\|y\|_G^2}{2}\right).$$

For the converse inequality, simply test the supremum with $y = x$. ■

Lemma 4.18. *Under the assumptions of Proposition 4.16, the Hopf-Lax function (4.113) satisfies the right initial condition,*

$$f_{\mathrm{HL}}(0,\cdot) = \psi(\cdot). \tag{4.117}$$

Proof. For $t = 0$, the definition of the Hopf-Lax function is interpreted as

$$f_{\mathrm{HL}}(0,x) = \sup\left\{\psi(x+y) \mid y \in \mathbb{R}^d_{\geqslant 0} \text{ with } \|y\|_G = 0\right\}. \tag{4.118}$$

Recalling (4.115), it becomes apparent that the only $y \in \mathbb{R}^d_{\geqslant 0}$ with $\|y\|_G = 0$ is $y = 0$. Together with (4.118), this completes the proof. ■

Lemma 4.19. *Under the assumptions of Proposition 4.16, for any $(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}^d_{\geqslant 0}$, there exists $y \in \mathbb{R}^d_{\geqslant 0}$ with*

$$f_{\mathrm{HL}}(t,x) = \psi(x+y) - \frac{\|y\|_G^2}{2t}. \tag{4.119}$$

Proof. Combining (4.115) with the Lipschitz continuity of ψ reveals that

$$\psi(x+y)-\frac{\|y\|_G^2}{2t}\leq \psi(x)+|||y|||_1\Big(|||\psi|||_{\mathrm{Lip},1}-\frac{m}{2t}|||y|||_1\Big).$$

The supremum in (4.113) can thus be restricted to the set of y's in $\mathbb{R}^d_{\geqslant 0}$ that satisfy $|||y|||_1\leqslant \frac{2t}{m}|||\psi|||_{\mathrm{Lip},1}$. Since a continuous function is now being optimized over a compact set, it is clear that the supremum is achieved. ■

Lemma 4.20 (Semigroup property)**.** *Under the assumptions of Proposition 4.16, for every pair $t>s>0$ and $x\in\mathbb{R}^d_{\geqslant 0}$,*

$$f_{\mathrm{HL}}(t,x)=\sup_{y\in\mathbb{R}^d_{\geqslant 0}}\bigg(f_{\mathrm{HL}}(s,x+y)-\frac{\|y\|_G^2}{2(t-s)}\bigg). \tag{4.120}$$

Proof. The proof is identical to that of Lemma 2.25. The reader interested in the details is referred to Lemma 4.5 in [48]. ■

Lemma 4.21. *Under the assumptions of Proposition 4.16, the Hopf-Lax function f_{HL} belongs to the solution space $\mathfrak{L}_{\mathrm{unif}}$ with*

$$\sup_{t>0}|||f_{\mathrm{HL}}(t,\cdot)|||_{\mathrm{Lip},1}\leqslant|||\psi|||_{\mathrm{Lip},1}\quad \textit{and}\quad [f_{\mathrm{HL}}]_0\leqslant\frac{|||\psi|||^2_{\mathrm{Lip},1}}{2m}. \tag{4.121}$$

Proof. The spatial Lipschitz continuity of the Hopf-Lax function is established exactly as in Lemma 2.26. The reader interested in the details is referred to Lemma 4.6 in [48]. To obtain Lipschitz continuity in time, fix $x\in\mathbb{R}^d_{\geqslant 0}$ as well as $t>s\geqslant 0$. The semigroup property in Lemma 4.20 with $y=0$ implies that

$$f_{\mathrm{HL}}(t,x)\geqslant f_{\mathrm{HL}}(s,x). \tag{4.122}$$

Using Lemma 4.20 in combination with (4.115) and the first inequality in (4.121) gives

$$f_{\mathrm{HL}}(t,x)\leqslant f_{\mathrm{HL}}(s,x)+\sup_{y\in\mathbb{R}^d_{\geqslant 0}}\bigg(|||\psi|||_{\mathrm{Lip},1}|||y|||_1-\frac{m|||y|||_1^2}{2(t-s)}\bigg)\leq f_{\mathrm{HL}}(s,x)+\frac{|||\psi|||^2_{\mathrm{Lip},1}}{2m}(t-s),$$

where it has been used that $r\mapsto r-\frac{1}{2}ar^2$ achieves its maximum at $r=1/a$. Combining this with (4.122) completes the proof. ■

Proof of Proposition 4.16. Denote by $\mathcal{E}_d$ an error term that will be defined in the course of the argument. The proof proceeds in three steps. First, it is shown that the function $f_+(t,x):=f_{\mathrm{HL}}(t,x)+\mathcal{E}_d t$ is a viscosity supersolution to the Hamilton-Jacobi equation (4.107), then it is verified that the function $f_-(t,x):=f_{\mathrm{HL}}(t,x)-\mathcal{E}_d t$ is a viscosity subsolution to the Hamilton-Jacobi equation (4.107), and finally the comparison principle in Proposition 4.9 is leveraged to conclude.

Step 1: showing f_+ is a viscosity supersolution. Consider a smooth function $\phi\in C^\infty\big(\mathbb{R}_{>0}\times\mathbb{R}^d_{>0};\mathbb{R}\big)$ with the property that $f_+-\phi$ has a local minimum at $(t^*,x^*)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$. Using Proposition A.2, it is readily verified that f_{HL} has its gradient in $\mathcal{K}'_a$ as ψ does. It follows by Lemma 4.13 that $\nabla\phi(t^*,x^*)\in\mathcal{K}'_a$. It is therefore possible to find $u\in\mathbb{R}^d_{\geqslant 0}$ and $w\in\mathbb{R}^d$ with

$$\nabla\phi(t^*,x^*)=Gu+w,\quad |||u|||_1\leqslant a,\quad \text{and}\quad |||w|||_{1,*}\leqslant\frac{1}{\sqrt{d}}.$$

On the one hand, if $s>0$ is sufficiently small that $t^*-s>0$, then

$$f_+(t^*-s,x^*+su)-\phi(t^*-s,x^*+su)\geqslant f_+(t^*,x^*)-\phi(t^*,x^*).$$

On the other hand, taking $su\in\mathbb{R}^d_{\geqslant 0}$ in Lemma 4.20 reveals that

$$f_{\mathrm{HL}}(t^*,x^*)\geqslant f_{\mathrm{HL}}(t^*-s,x^*+su)-s\frac{\|u\|_G^2}{2}.$$

It follows that

$$\phi(t^*,x^*)-\phi(t^*-s,x^*+su)+s\frac{\|u\|_G^2}{2}-\mathcal{E}_d s\geqslant 0.$$

Dividing by $0<s<t^*$ and letting $s\to 0$ yields

$$\partial_t\phi(t^*,x^*)-u\cdot\nabla\phi(t^*,x^*)+\frac{\|u\|_G^2}{2}-\mathcal{E}_d\geqslant 0.$$

Recalling that $\nabla\phi(t^*,x^*)=Gu+w$ and using Lemma 4.17,

$$\partial_t\phi(t^*,x^*)-\mathsf{C}(Gu)-u\cdot w-\mathcal{E}_d\geqslant 0.$$

By Lemma 4.12 and Lemma 4.21,

$$|||Gu|||_{1,*}\leqslant|||\nabla\phi(t^*,x^*)|||_{1,*}+\frac{1}{\sqrt{d}}\leqslant|||\psi|||_{\mathrm{Lip},1}+\frac{1}{\sqrt{d}}\leqslant R,$$

so the Lipschitz continuity of the non-linearity H_R established in Proposition 4.8 implies that

$$\big(\partial_t\phi-\mathsf{H}_R(\nabla\phi)\big)(t^*,x^*)\geqslant\mathcal{E}_d-|||u|||_1|||w|||_{1,*}-\frac{8RM}{m^2\sqrt{d}}\geqslant\mathcal{E}_d-\frac{a}{\sqrt{d}}-\frac{8RM}{m^2\sqrt{d}}.$$

This shows that f_+ is a supersolution to the Hamilton-Jacobi equation (4.107) provided that

$$\mathcal{E}_d\geqslant\frac{a}{\sqrt{d}}+\frac{8RM}{m^2\sqrt{d}}.$$

Step 2: verifying f_- is a viscosity subsolution. Consider a smooth function $\phi\in C^\infty\big(\mathbb{R}_{>0}\times\mathbb{R}^d_{>0};\mathbb{R}\big)$ with the property that $f_--\phi$ has a local maximum at $(t^*,x^*)\in\mathbb{R}_{>0}\times\mathbb{R}^d_{>0}$. Recall that f_{HL} has its gradient in $\mathcal{K}'_a$ as ψ does. It follows by Lemma 4.13 that $\nabla\phi(t^*,x^*)\in\mathcal{K}'_a$. It is therefore possible to find $u\in\mathbb{R}^d_{\geqslant 0}$ and $w\in\mathbb{R}^d$ with

$$\nabla\phi(t^*,x^*)=Gu+w,\quad |||u|||_1\leqslant a,\quad\text{and}\quad |||w|||_{1,*}\leqslant\frac{1}{\sqrt{d}}.$$

Suppose for the sake of contradiction that there exists $\delta>0$ with

$$\big(\partial_t\phi-\mathsf{H}_R(\nabla\phi)\big)(t^*,x^*)\geqslant\delta>0.$$

Arguing as in the previous step, this implies that

$$\partial_t \phi(t^*, x^*) - \mathsf{C}(Gu) \geqslant \delta - \frac{8RM}{m^2\sqrt{d}}.$$

By Lemma 4.17, this may be recast as the assumption that for all $y \in \mathbb{R}^d_{\geqslant 0}$,

$$\partial_t \phi(t^*, x^*) - y \cdot Gu + \frac{\|y\|_G^2}{2} \geqslant \delta - \frac{8RM}{m^2\sqrt{d}}.$$

By continuity, of $\partial_t \phi$ and $\nabla \phi$, up to redefining $\delta > 0$, it is in fact possible to assume that for all $y \in \mathbb{R}^d_{\geqslant 0}$ and (t', x') sufficiently close to (t^*, x^*),

$$\partial_t \phi(t', x') - y \cdot \nabla \phi(t', x') + \frac{\|y\|_G^2}{2} \geqslant \delta - \frac{8RM}{m^2\sqrt{d}} - |||y|||_1 |||w|||_{1,*}. \tag{4.123}$$

Recalling Lemma 4.20, and arguing as in the proof of Lemma 4.19, it is possible to find $R > 0$ such that, for every $s > 0$ sufficiently small, there exists $y_s \in \mathbb{R}^d_{\geqslant 0}$ with $|||y_s|||_1 \leqslant Rs$ and

$$f_{\mathrm{HL}}(t^*, x^*) = f_{\mathrm{HL}}(t^* - s, x^* + y_s) - \frac{\|y_s\|_G^2}{2s}.$$

If $u(r) := (t^* + (r-1)s, x^* + (1-r)y_s)$, then it follows by the fundamental theorem of calculus and the absurd assumption (4.123) with $y := \frac{y_s}{s} \in \mathbb{R}^d_{\geqslant 0}$ that

$$\begin{aligned}
\phi(t^*, x^*) - \phi(t^* - s, x^* + y_s) &= \int_0^1 \frac{\mathrm{d}}{\mathrm{d}r} \phi(u(r))\,\mathrm{d}r \\
&= \int_0^1 \big(s\partial_t \phi - y_s \cdot \nabla \phi\big)(u(r))\,\mathrm{d}r \\
&\geqslant s\delta - \frac{\|y_s\|_G^2}{2s} - s\frac{8RM}{m^2\sqrt{d}} - |||y_s|||_1 |||w|||_{1,*} \\
&\geqslant f_{\mathrm{HL}}(t^*, x^*) - f_{\mathrm{HL}}(t^* - s, x^* + y_s) + s\left(\delta - \frac{8RM}{m^2\sqrt{d}} - \frac{R}{\sqrt{d}}\right).
\end{aligned}$$

Rearranging shows that for s sufficiently small,

$$f_-(t^* - s, x^* + y_s) - \phi(t^* - s, x^* + y_s) \geqslant s\left(\delta - \frac{8RM}{m^2\sqrt{d}} - \frac{R}{\sqrt{d}} + \mathcal{E}_d\right) + f_-(t^*, x^*) - \phi(t^*, x^*).$$

This contradicts the fact that $f - \phi$ admits a local maximum at (t^*, x^*) provided that

$$\mathcal{E}_d \geqslant \frac{R}{\sqrt{d}} + \frac{8RM}{m^2\sqrt{d}}.$$

Step 3: comparison principle. Define the error term $\mathcal{E}_d$ by

$$\mathcal{E}_d := \frac{1}{\sqrt{d}}\left(R + a + \frac{8RM}{m^2}\right).$$

Combining the previous two steps shows that f_+ is a viscosity supersolution to the Hamilton-Jacobi equation (4.107) while f_- is a viscosity subsolution to this equation. Together with Lemmas 4.18 and 4.21, and the comparison principle in Proposition 4.9, this implies that for $(t,x) \in \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$, one has $|f_{\mathrm{HL}}(t,x) - f(t,x)| \leqslant \mathcal{E}_d t$ as required. ∎

This result is now applied to the projected Hamilton-Jacobi equation (4.30) in the setting when the matrix $G^{(K)}$ in (4.14) is non-negative definite. Denote by $f^{(K)} \in \mathfrak{L}_{\mathrm{unif}}$ the unique solution to the Hamilton-Jacobi equation (4.30) provided by Theorem 4.1, and observe that the Hopf-Lax function (4.113) becomes

$$f^{(K)}_{\mathrm{HL}}(t,x) = \sup_{y \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}} \left(\psi^{(K)}\big(\mu_x^{(K)} + \mu_y^{(K)}\big) - \frac{\mathsf{C}_\infty(G_{\mu_y^{(K)}})}{t} \right) = \sup_{\nu \in \mathcal{M}_+^{(K)}} \left(\psi\big(\mu_x^{(K)} + \nu\big) - \frac{\mathsf{C}_\infty(G_\nu)}{t} \right), \tag{4.124}$$

where the first equality uses the relationship (4.13) between the projected initial condition $\psi^{(K)}$ and the initial condition ψ as well as the relationship (4.17) between the projected non-linearity C_K and the non-linearity C_∞, and where the second equality leverages (4.9) and (4.10) to parameterize the Hopf-Lax formula using the space (4.8) of projected measures. The approximate Hopf-Lax variational formula in Proposition 4.16 implies in particular that for any measure $\mu \in \mathcal{M}_+$,

$$f^{(K)}\big(t, x^{(K)}(\mu)\big) = \sup_{\nu \in \mathcal{M}_+^{(K)}} \left(\psi(\mu + \nu) - \frac{\mathsf{C}_\infty(G_\nu)}{t} \right) + \mathcal{O}\big(t|\mathcal{D}_K|^{-1/2}\big). \tag{4.125}$$

Together with Theorem 4.2, this will give the Hopf-Lax variational formula stated in Theorem 4.3 for the infinite-dimensional Hamilton-Jacobi equation (4.6).

4.3.2 Hopf-Lax formula for infinite-dimensional equations

A Hopf-Lax variational formula for the solution to the infinite-dimensional Hamilton-Jacobi equation (4.6) constructed in Theorem 4.2 is now obtained by letting the dimension parameter $K \geqslant 1$ tend to infinity in the approximate Hopf-Lax variational formula (4.125). A proof of the remaining parts of Theorem 4.3 is then provided.

In the context of Theorem 4.3, in addition to the assumptions (**H1**)-(**H4**), the kernel $g : [-1,1] \to \mathbb{R}$ is assumed to be non-negative definite in the sense that it satisfies (**H5**). This assumption is equivalent to the non-negative definiteness of each of the matrices (4.14), and therefore to the convexity of each of the projected non-linearities (4.17). In particular, Proposition 4.16 implies that the unique solution $f^{(K)} : \mathbb{R}_{\geqslant 0} \times \mathbb{R}^{\mathcal{D}_K} \to \mathbb{R}$ to the projected Hamilton-Jacobi equation (4.30) provided by Theorem 4.1 satisfies the approximate Hopf-Lax variational formula (4.125). Using Theorem 4.2 and a simple continuity argument to let K tend to infinity in this expression shows that for any measure $\mu \in \mathcal{M}_+$,

$$f(t,\mu) = \sup_{\nu \in \mathcal{M}_+} \left(\psi(\mu + \nu) - \frac{1}{2t} \int_{-1}^{1} G_\nu(y) \, \mathrm{d}\nu(y) \right) = \sup_{\nu \in \mathcal{M}_+} \left(\psi(\mu + t\nu) - \frac{t}{2} \int_{-1}^{1} G_\nu(y) \, \mathrm{d}\nu(y) \right). \tag{4.126}$$

The second of these expressions follows from the first by setting $\nu' = t\nu$. To establish Theorem 4.3, it remains to verify that the supremum in (4.126) is achieved at some $\nu^* \in \mathcal{M}_+$, and that $G_{\nu^*} = D_\mu(\mu + t\nu^*, \cdot)$ whenever the initial condition admits a Gateaux derivative at the measure $\mu + t\nu^*$ with density $x \mapsto D_\mu \psi(\mu + t\nu^*, x)$ in $\mathcal{C}_\infty$. Putting aside the constraint that the optimizers in (4.126) must be non-negative measures, this latter property is clear from the first-order conditions on a maximizer. The following two technical lemmas will be used to rigorously complete the proof of Theorem 4.3. The first establishes the existence of a maximizer to the variational problem (4.126), and the second is a Cauchy-Schwarz inequality for the kernel $\widetilde{g}(x,y) := g(xy)$.

Lemma 4.22. *For every $t \geqslant 0$ and $\mu \in \mathcal{M}_+$, there exists $\nu^* \in \mathcal{M}_+$ with*

$$f(t,\mu) = \psi(\mu + t\nu^*) - \frac{t}{2}\int_{-1}^{1} G_{\nu^*}(y)\,\mathrm{d}\nu^*(y). \tag{4.127}$$

Proof. Fix a probability measure $\nu \in \mathrm{Pr}[-1,1]$ and a positive constant $\lambda > 0$. The Lipschitz continuity (**H2**) of the initial condition implies that

$$\psi(\mu + \lambda t\nu) \leqslant \psi(\mu) + \|\psi\|_{\mathrm{Lip,TV}}\mathrm{TV}(0,\lambda t\nu) \leqslant \psi(\mu) + 2\lambda t\|\psi\|_{\mathrm{Lip,TV}}.$$

On the other hand,

$$\int_{-1}^{1} G_{\lambda\nu}(y)\,\mathrm{d}(\lambda\nu)(y) = \lambda^2\int_{-1}^{1}\int_{-1}^{1} g(xy)\,\mathrm{d}\nu(x)\,\mathrm{d}\nu(y) \geqslant \lambda^2 m.$$

Combining these two bounds reveals that

$$\psi(\mu + t\nu) - \frac{t}{2}\int_{-1}^{1} G_{\nu}(y)\,\mathrm{d}\nu(y) \leqslant \psi(\mu) + 2\lambda t\|\psi\|_{\mathrm{Lip,TV}} - \frac{\lambda^2 tm}{2}.$$

The supremum in (4.126) can therefore be restricted to measures in $\mathcal{M}_+$ with bounded total mass. The existence of a maximizer is now an immediate consequence of Prokhorov's theorem (Theorem A.20 in [50]). Indeed, if $(\nu_n)_{n\geqslant 1} \subseteq \mathcal{M}_+$ denotes a maximizing sequence, one may assume that each measure in this sequence has total mass bounded by the same constant. It follows by Prokhorov's theorem that this sequence is pre-compact, and therefore admits a subsequential limit with respect to the weak convergence of measures. By continuity of the functional being maximized in (4.126), this weak limit must be a maximizer. ∎

Lemma 4.23. *If g satisfies (**H5**) and $\mu,\nu \in \mathcal{M}_s$ are signed measures, then*

$$\left(\int_{-1}^{1} G_{\nu}(x)\,\mathrm{d}\mu(x)\right)^2 \leqslant \left(\int_{-1}^{1} G_{\mu}(x)\,\mathrm{d}\mu(x)\right)\left(\int_{-1}^{1} G_{\nu}(x)\,\mathrm{d}\nu(x)\right). \tag{4.128}$$

Proof. This is the Cauchy-Schwarz inequality for the non-negative definite kernel $\widetilde{g}(x,y) := g(xy)$, and can be proved in a standard way. Indeed, for every $t \in \mathbb{R}$, let

$$\begin{aligned} P(t) &:= \int_{-1}^{1}\int_{-1}^{1} g(xy)\,\mathrm{d}(\mu + t\nu)(x)\,\mathrm{d}(\mu + t\nu)(y) \\ &= \int_{-1}^{1} G_{\mu}(x)\,\mathrm{d}\mu(x) + 2t\int_{-1}^{1} G_{\nu}(x)\,\mathrm{d}\mu(x) + t^2\int_{-1}^{1} G_{\nu}(x)\,\mathrm{d}\nu(x). \end{aligned}$$

This polynomial is non-negative by (**H5**). In particular, its discriminant cannot be positive. This means that

$$2^2\left(\int_{-1}^{1} G_{\nu}(x)\,\mathrm{d}\mu(x)\right)^2 - 4\left(\int_{-1}^{1} G_{\mu}(x)\,\mathrm{d}\mu(x)\right)\left(\int_{-1}^{1} G_{\nu}(x)\,\mathrm{d}\nu(x)\right) \leqslant 0.$$

Rearranging completes the proof. ∎

Proof of Theorem 4.3. Fix $t > 0$ and $\mu \in \mathcal{M}_+$. Combining Theorem 4.2 with (4.126) shows that the unique solution to the infinite-dimensional Hamilton-Jacobi equation (4.6) admits the Hopf-Lax variational representation (4.38). Moreover, Lemma 4.22 ensures that the supremum in (4.126) is achieved at some $\nu^* \in \mathcal{M}_+$. To establish the final statement in Theorem 4.3, suppose that the initial condition ψ admits a Gateaux derivative at the measure $\mu + t\nu^*$ with density $x \mapsto D_\mu \psi(\mu + t\nu^*, x)$ in $\mathcal{C}_\infty$. For any measure $\eta \in \mathcal{M}_+$, the Gateaux derivative of the functional

$$\nu \mapsto \psi(\mu + t\nu) - \frac{t}{2}\int_{-1}^{1} G_\nu(y)\,\mathrm{d}\nu(y)$$

at the measure ν^* in the direction $\eta - \nu^*$ is

$$D_\mu \psi\big(\mu + t\nu^*; t(\eta - \nu^*)\big) - t\int_{-1}^{1}\int_{-1}^{1} g(xy)\,\mathrm{d}(\eta - \nu^*)(x)\,\mathrm{d}\nu^*(y). \tag{4.129}$$

For every $\varepsilon \in [0,1]$, the measure $\nu^* + \varepsilon(\eta - \nu^*)$ belongs to $\mathcal{M}_+$, and is thus a valid candidate for the optimization problem in (4.126). As a consequence, the quantity in (4.129) must be non-positive. Using also the definition of the Gateaux derivative density in (4.5) reveals that for every $\eta \in \mathcal{M}_+$,

$$t\int_{-1}^{1}\Big(D_\mu \psi(\mu + t\nu^*, x) - \int_{-1}^{1} g(xy)\,\mathrm{d}\nu^*(y)\Big)\,\mathrm{d}(\eta - \nu^*)(x) \leqslant 0. \tag{4.130}$$

The assumption that the density $x \mapsto D_\mu \psi(\mu + t\nu^*, x)$ belongs to the cone $\mathcal{C}_\infty$ gives a measure $\eta^* \in \mathcal{M}_+$ with $G_{\eta^*}(x) = D_\mu \psi(\mu + t\nu^*, x)$. Applying (4.130) to the measure $\eta = \eta^*$ gives

$$\int_{-1}^{1}\int_{-1}^{1} g(xy)\,\mathrm{d}(\eta^* - \nu^*)(y)\,\mathrm{d}(\eta^* - \nu^*)(x) \leqslant 0.$$

Remembering the assumption (**H5**), this inequality is in fact an equality. Applying the Cauchy-Schwarz inequality in Lemma 4.23 to the signed measures $\eta^* - \nu^*$ and δ_x for some $x \in [-1,1]$ shows that

$$\int_{-1}^{1} g(xy)\,\mathrm{d}(\eta^* - \nu^*)(y) = \int_{-1}^{1} g(yz)\,\mathrm{d}(\eta^* - \nu^*)(y)\,\mathrm{d}\delta_x(z) = 0.$$

Rearranging gives $G_{\nu^*}(x) = G_{\eta^*}(x) = D_\mu(\mu + t\nu^*, x)$. Since $x \in [-1,1]$ is arbitrary, this completes the proof. ∎

4.4 Extending to more general infinite-dimensional equations

To study the sparse stochastic block model, it will be important to extend Theorems 4.1 - 4.3 to the setting where the kernel g does not satisfy the positivity assumption (**H1**). This assumption will be replaced by the stronger constraint (**H3'**) on the gradient of the projected initial conditions, and it will lead to Theorems 4.4 and 4.5. Fix $b \in \mathbb{R}$ large enough to that the modified kernel $\widetilde{g}_b$ in (4.45) is strictly positive. Recall the definition of the modified initial condition $\widetilde{\psi}_b$ in (4.50), and of the modified Hamilton-Jacobi equation (4.49). In the spirit of (4.14), (4.16), (4.17) and (4.21), introduce the symmetric matrix

$$\widetilde{G}_b^{(K)} := \frac{1}{|\mathcal{D}_K|^2}\big(\widetilde{g}_b(kk')\big)_{k,k' \in \mathcal{D}_K} \in \mathbb{R}^{\mathcal{D}_K \times \mathcal{D}_K}, \tag{4.131}$$

the projected cone

$$\widetilde{\mathcal{C}}_{b,K} := \left\{ \widetilde{G}_b^{(K)} x^{(K)}(\mu) \in \mathbb{R}^{\mathcal{D}_K} \mid \mu \in \mathcal{M}_+^{(K)} \right\} = \left\{ \widetilde{G}_b^{(K)} x \in \mathbb{R}^{\mathcal{D}_K} \mid x \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K} \right\}, \tag{4.132}$$

the projected non-linearity $\widetilde{\mathsf{C}}_{b,K} : \widetilde{\mathcal{C}}_{b,K} \to \mathbb{R}$ defined by

$$\widetilde{\mathsf{C}}_{b,K}\big(\widetilde{G}_b^{(K)} x\big) := \frac{1}{2} \widetilde{G}_b^{(K)} x \cdot x = \frac{1}{2|\mathcal{D}_K|^2} \sum_{k,k' \in \mathcal{D}_K} \widetilde{g}_b(kk') x_k x_{k'}, \tag{4.133}$$

and the closed convex set

$$\widetilde{\mathcal{K}}_{=a,b,K} := \left\{ \widetilde{G}_b^{(K)} x \in \mathbb{R}^{\mathcal{D}_K} \mid x \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K} \text{ and } |||x|||_1 = a \right\}. \tag{4.134}$$

To prove Theorem 4.4, it will first be verified that, under the assumptions of this result, the kernel $\widetilde{g}_b$ and the initial condition $\widetilde{\psi}_b$ satisfy (**H1**)-(**H4**). Together with Theorems 4.1 and 4.2, this will ensure that the function f_b in (4.51) is well-defined. The comparison principle in Theorem 4.1 will then be used to show that this function is independent of b. Invoking the Hopf-Lax formula in Theorem 4.3 will give the Hopf-Lax formula in Theorem 4.5.

Lemma 4.24. *Under the assumptions of Theorem 4.4, the kernel $\widetilde{g}_b$ in (4.45) and the initial condition $\widetilde{\psi}_b$ in (4.50) satisfy (**H1**)-(**H4**). Moreover, each projected initial condition $\widetilde{\psi}_b^{(K)}$ has its gradient in the closed convex set $\widetilde{\mathcal{K}}_{=a,b,K}$.*

Proof. The kernel $\widetilde{g}_b$ satisfies (**H1**) by the choice of b, while the initial condition $\widetilde{\psi}_b$ satisfies (**H2**) by the triangle inequality and the bound

$$\left| ab \int_{-1}^{1} \mathrm{d}\mu - ab \int_{-1}^{1} \mathrm{d}\nu \right| \leqslant a|b| \big| \mu[-1,1] - \nu[-1,1] \big| \leqslant a|b| \mathrm{TV}(\mu,\nu).$$

An identical argument shows that the initial condition $\widetilde{\psi}_b$ satisfies (**H4**). To verify (**H3**), introduce the closed convex set

$$\widetilde{\mathcal{K}}_{a,b,K} := \left\{ \widetilde{G}_b^{(K)} x \in \mathbb{R}^{\mathcal{D}_K} \mid x \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K} \text{ and } |||x|||_1 \leqslant a \right\},$$

and fix $c \in \mathbb{R}$ as well as $x, x' \in \mathbb{R}^{\mathcal{D}_K}$ with $(x'-x)\cdot z \geqslant c$ for every $z \in \widetilde{\mathcal{K}}'_{a,b,K}$. Given $y \in \mathcal{K}'_{=a,K}$, represent it as $y = G^{(K)} u + w$ for some $u \in \mathbb{R}^d_{\geqslant 0}$ and $w \in \mathbb{R}^d$ with $|||u|||_1 = a$ and $|||w|||_{1,*} \leqslant 2^{-K/2}$. Define $z := \widetilde{G}_b^{(K)} u + w \in \widetilde{\mathcal{K}}'_{=a,b,K}$, and observe that

$$z = G^{(K)} u + b|||u|||_1 \iota_K + w = y + ab\iota_K$$

for the vector $\iota_K := (|\mathcal{D}_K|^{-1})_{k \in \mathcal{D}_K} \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$. Since $z \in \widetilde{\mathcal{K}}'_{=a,b,K} \subseteq \widetilde{\mathcal{K}}'_{a,b,K}$,

$$(x'-x)\cdot y = (x'-x)\cdot z - (x'-x)\cdot ab\iota_K \geqslant c - (x'-x)\cdot ab\iota_K.$$

The assumption (**H3’**) and Proposition A.2 therefore imply that

$$\psi^{(K)}(x') - \psi^{(K)}(x) \geqslant c - (x'-x)\cdot ab\iota_K.$$

Noticing that $x \cdot \iota_K = |||x|||_1$ and rearranging reveals that

$$\widetilde{\psi}_b^{(K)}(x') - \widetilde{\psi}_b^{(K)}(x) \geqslant c.$$

Invoking Proposition A.2 establishes (**H3**). Observe that this argument only uses the assumption that $(x'-x)\cdot z \geqslant c$ for every $z \in \widetilde{\mathcal{K}}_{=a,b,K}$, so it also shows that $\widetilde{\psi}_b^{(K)}$ has its gradient in $\widetilde{\mathcal{K}}_{=a,b,K}$. This completes the proof. ∎

Together with Theorems 4.1 and 4.2, this result ensures that the function f_b in (4.51) is well-defined. Indeed, fix $R > \|\widetilde{\psi}\|_{\mathrm{Lip,TV}}$, and denote by $\widetilde{\mathsf{H}}_{b,K,R} : \mathbb{R}^{\mathcal{D}_K} \to \mathbb{R}$ the extension of the non-linearity $\widetilde{\mathsf{C}}_{b,K}$ given by Proposition 4.8. Invoking Lemma 4.24, Theorem 4.1, and Proposition 4.11 ensures that the Hamilton-Jacobi equation

$$\partial_t \widetilde{f}^{(K)}(t,x) = \widetilde{\mathsf{H}}_{b,K,R}\big(\nabla \widetilde{f}^{(K)}(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K} \tag{4.135}$$

subject to the initial condition $\widetilde{f}^{(K)}(0,\cdot) = \widetilde{\psi}_b(\cdot)$ admits a unique viscosity solution $\widetilde{f}_{b,R}^{(K)} \in \mathcal{L}_{\mathrm{unif}}$ which satisfies the Lipschitz bound

$$\sup_{t\geqslant 0} \big|\!\big|\!\big| \widetilde{f}_{b,R}^{(K)}(t,\cdot) \big|\!\big|\!\big|_{\mathrm{Lip},1} = \big|\!\big|\!\big| \widetilde{\psi}_b^{(K)} \big|\!\big|\!\big|_{\mathrm{Lip},1}, \tag{4.136}$$

and has its gradient in the the closed convex set $\mathcal{K}'_{=a,b,K}$. Appealing to Theorem 4.2 also guarantees that the function

$$\widetilde{f}_b(t,\mu) := \lim_{K\to+\infty} \widetilde{f}_{b,R}^{(K)}\big(t, x^{(K)}(\mu)\big), \tag{4.137}$$

on $\mathbb{R}_{\geqslant 0} \times \mathcal{M}_+$ is well-defined and independent of R. In fact, it is the solution to the infinite-dimensional Hamilton-Jacobi equation (4.49). Using the comparison principle in Theorem 4.1, it will now be shown that the limit defining the function f_b in (4.51),

$$f_b(t,\mu) = \lim_{K\to+\infty}\left(\widetilde{f}_{b,R}^{(K)}\big(t,x^{(K)}(\mu)\big) - ab\big|\!\big|\!\big|x^{(K)}(\mu)\big|\!\big|\!\big|_1 - \frac{a^2bt}{2}\right), \tag{4.138}$$

is independent of b, thus establishing Theorem 4.4. The Hopf-Lax formula in Theorem 4.5 will then be obtained from the Hopf-Lax formula in Theorem 4.3.

Proof of Theorem 4.4. Let $b,b' \in \mathbb{R}$ be such that the kernels $\widetilde{g}_b$ and $\widetilde{g}_{b'}$ are positive on $[-1,1]$, and fix $R > \max(\|\widetilde{\psi}_b\|_{\mathrm{Lip,TV}}, \|\widetilde{\psi}_{b'}\|_{\mathrm{Lip,TV}})$. The idea will be to show that the function

$$f_{b,b'}^{(K)}(t,x) := \widetilde{f}_b^{(K)}(t,x) - a(b-b')\big|\!\big|\!\big|x\big|\!\big|\!\big|_1 - \frac{a^2(b-b')t}{2}$$

satisfies the Hamilton-Jacobi equation defining $\widetilde{f}_{b'}^{(K)}$ up to an error that vanishes with K. The dependence on R has been omitted and will be omitted throughout this proof, since R will remain fixed. The equality of f_b and $f_{b'}$ will then follow from the comparison principle in Theorem 4.1. Consider $\phi \in C^\infty\big(\mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K}; \mathbb{R}\big)$ with the property that $f_{b,b'}^{(K)} - \phi$ achieves a local maximum at the point $(t^*,x^*) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K}$. Since $\widetilde{f}_b^{(K)}$ is a viscosity subsolution to the Hamilton-Jacobi equation (4.135),

$$\frac{a^2(b-b')}{2} + \partial_t\phi(t^*,x^*) - \widetilde{\mathsf{H}}_{b,K}\big(a(b-b')\iota_K + \nabla\phi(t^*,x^*)\big) \leqslant 0$$

for the vector $\iota_K := (|\mathcal{D}_K|^{-1})_{k\in\mathcal{D}_K} \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$. The fact that $\widetilde{f}_b^{(K)}$ has its gradient in $\widetilde{\mathcal{K}}'_{=a,b,K}$ together with (4.88), and Lemmas 4.12 and 4.13 implies that

$$a(b-b')\iota_K + \nabla\phi(t^*,x^*) \in \widetilde{\mathcal{K}}'_{=a,b,K} \quad \text{and} \quad \big|\!\big|\!\big|a(b-b')\iota_K + \nabla\phi(t^*,x^*)\big|\!\big|\!\big|_{1,*} \leqslant \|\widetilde{\psi}_b\|_{\mathrm{Lip,TV}}.$$

It is therefore possible to find $u \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$ and $w \in \mathbb{R}^{\mathcal{D}_K}$ with

$$a(b-b')\iota_K + \nabla\phi(t^*,x^*) = \widetilde{G}_b^{(K)} u + w, \quad |||u|||_1 = a, \quad \text{and} \quad |||w|||_{1,*} \leqslant \frac{1}{2^{K/2}}.$$

Observe that

$$|||\widetilde{G}_b^{(K)} u|||_{1,*} \leqslant |||a(b-b')\iota_K + \nabla\phi(t^*,x^*)|||_{1,*} + |||w|||_{1,*} \leqslant \|\widetilde{\psi}_b\|_{\mathrm{Lip,TV}} + \frac{1}{2^{K/2}},$$

so increasing K if necessary, it is possible to ensure that $\widetilde{G}_b^{(K)} u \in \widetilde{\mathcal{C}}_{b,K} \cap B_R$. It follows by the Lipschitz continuity of $\widetilde{\mathsf{H}}_{b,K}$ established in Proposition 4.8 that

$$\frac{a^2(b-b')}{2} + \partial_t \phi(t^*,x^*) - \widetilde{\mathsf{C}}_{b,K}\big(\widetilde{G}_b^{(K)} u\big) \leqslant \frac{8RM}{2^{K/2}m^2}.$$

Notice that

$$\widetilde{\mathsf{C}}_{b,K}\big(\widetilde{G}_b^{(K)} u\big) = \frac{1}{2}\widetilde{G}_b^{(K)} u\cdot u = \frac{1}{2}G^{(K)}u\cdot u + \frac{1}{2}b|||u|||_1^2 = \frac{1}{2}\widetilde{G}_{b'}^{(K)} u\cdot u + \frac{1}{2}(b-b')a^2 = \widetilde{\mathsf{C}}_{b',K}\big(\widetilde{G}_{b'}^{(K)} u\big) + \frac{a^2(b-b')}{2},$$

so in fact

$$\partial_t \phi(t^*,x^*) - \widetilde{\mathsf{C}}_{b'}^{(K)}\big(\widetilde{G}_{b'}^{(K)} u\big) \leqslant \frac{8RM}{2^{K/2}m}.$$

To replace $\widetilde{G}_{b'}^{(K)} u$ by $\nabla\phi(t^*,x^*)$ observe that

$$\widetilde{G}_{b'}^{(K)} u = \widetilde{G}_b^{(K)} u + (b'-b)\iota_K |||u|||_1 = a(b-b')\iota_K + \nabla\phi(t^*,x^*) - w + a(b'-b)\iota_K = \nabla\phi(t^*,x^*) - w,$$

and leverage the Lipschitz continuity of $\widetilde{\mathsf{H}}_{b',K}$ established in Proposition 4.8 to deduce that

$$\big(\partial_t\phi - \widetilde{\mathsf{H}}_{b',K}(\nabla\phi)\big)(t^*,x^*) \leqslant \mathcal{E}_K \quad \text{for the error term} \quad \mathcal{E}_K := \frac{16RM}{2^{K/2}m}.$$

In particular, the function $(t,x) \mapsto f_{b,b'}^K(t,x) - \mathcal{E}_K t$ is a viscosity subsolution to the Hamilton-Jacobi equation defining $\widetilde{f}_{b'}^K$. A similar argument shows that $(t,x) \mapsto f_{b,b'}^K(t,x) + \mathcal{E}_K t$ is a viscosity supersolution to the Hamilton-Jacobi equation defining $\widetilde{f}_{b'}^K$. It follows by the comparison principle in Theorem 4.1 that for every $\mu \in \mathcal{M}_+$,

$$\big|f_b^{(K)}\big(t,x^{(K)}(\mu)\big) - f_{b'}^{(K)}\big(t,x^{(K)}(\mu)\big)\big| = \big|f_{b,b'}^{(K)}\big(t,x^{(K)}(\mu)\big) - \widetilde{f}_{b'}^{(K)}\big(t,x^{(K)}(\mu)\big)\big| \leqslant \mathcal{E}_K t.$$

Letting K tend to infinity completes the proof. ■

Proof of Theorem 4.5. Recall the definition of the set $\mathcal{M}_{a,+}$ of measures with mass a, and of its image cone of functions $\mathcal{C}_{a,\infty}$. Fix $b > 0$ large enough so the kernel $\widetilde{g}_b$ is positive on $[-1,1]$ and satisfies (**H5**). Lemma 4.24 and the Hopf-Lax representation formula in Theorem 4.3 imply that for any $t > 0$ and $\mu \in \mathcal{M}_+$,

$$\widetilde{f}_b(t,\mu) = \sup_{\nu\in\mathcal{M}_+}\left(\widetilde{\psi}_b(\mu + t\nu) - \frac{t}{2}\int_{-1}^1 \widetilde{G}_{b,\nu}(y)\,\mathrm{d}\nu(y)\right). \tag{4.139}$$

Since the Gateaux derivative density $x \mapsto D_\mu\psi(\mu + t\nu)$ belongs to the set $\mathcal{C}_{a,\infty}$ by assumption, there exists a measure $\eta \in \mathcal{M}_{a,+}$ with $D_\mu \psi(\mu + t\nu, \cdot) = G_\eta$. This means that

$$D_\mu \widetilde{\psi}_b(\mu + t\nu, x) = D_\mu \psi(\mu + t\nu, x) + ab = \int_{-1}^{1} g(xy)\,\mathrm{d}\eta(y) + ab = \int_{-1}^{1} \widetilde{g}_b(xy)\,\mathrm{d}\eta(y),$$

so another application of Theorem 4.3 implies that the supremum in (4.139) is achieved at some $\nu^* \in \mathcal{M}_+$ with

$$\widetilde{G}_{b,\nu^*} = D_\mu \widetilde{\psi}_b(\mu + t\nu^*, \cdot) = \widetilde{G}_{b,\eta}.$$

Evaluating this equality at $x = 0$ reveals that

$$\widetilde{g}_b(0)\int_{-1}^{1} \mathrm{d}\nu^*(y) = \widetilde{g}_b(0)\int_{-1}^{1} \mathrm{d}\eta(y) = \widetilde{g}_b(0)a.$$

Since $\widetilde{g}_b(0) > 0$ by the choice of b, this means that $\nu^* \in \mathcal{M}_{a,+}$ and

$$\widetilde{f}_b(t,\mu) = \sup_{\nu \in \mathcal{M}_{a,+}} \left(\widetilde{\psi}_b(\mu + t\nu) - \frac{t}{2}\int_{-1}^{1} \widetilde{G}_{b,\nu}(y)\,\mathrm{d}\nu(y) \right).$$

It follows by (4.51) that

$$\begin{aligned} f_b(t,\mu) &= \sup_{\nu \in \mathcal{M}_{a,+}} \left(\widetilde{\psi}_b(\mu + t\nu) - \frac{t}{2}\int_{-1}^{1} \widetilde{G}_{b,\nu}(y)\,\mathrm{d}\nu(y) - ab\int_{-1}^{1} \mathrm{d}\mu - \frac{a^2 bt}{2} \right) \\ &= \sup_{\nu \in \mathcal{M}_{a,+}} \left(\psi(\mu + t\nu) + abt\int_{-1}^{1} \mathrm{d}\nu - \frac{t}{2}\int_{-1}^{1} G_\nu(y)\,\mathrm{d}\nu(y) - \frac{bt}{2}\int_{-1}^{1}\int_{-1}^{1} \mathrm{d}\nu\,\mathrm{d}\nu - \frac{a^2 bt}{2} \right) \\ &= \sup_{\nu \in \mathcal{M}_{a,+}} \left(\psi(\mu + t\nu) - \frac{t}{2}\int_{-1}^{1} G_\nu(y)\,\mathrm{d}\nu(y) \right). \end{aligned}$$

This completes the proof. ∎

At the end of Chapter 3, two main obstacles stood in the way of asserting that the limit of the enriched free energy (3.30) should satisfy the infinite-dimensional Hamilton-Jacobi equation (3.91) subject to the initial condition (3.92). The first was that infinite-dimensional Hamilton-Jacobi equations of this form had not been well-studied in the literature, so making rigorous sense of this equation was not possible, and the second was that the concentration (3.75) of the multioverlaps (3.41) was not expected to hold for all choices of parameters. The first of these issues has been resolved in this chapter, where a well-posedness theory for infinite-dimensional Hamilton-Jacobi equations has been discussed. It will be shown in Lemma 6.1 that this theory is general enough to handle the Hamilton-Jacobi (3.91). The second of these issues will be addressed in the next chapter, where a finitary version of the multioverlap concentration result in [15] will be established.

Chapter 5

Multioverlap concentration

In this chapter, a finitary version of the main result in [15] regarding the concentration of the multioverlaps (3.41) is established. To alleviate notation and strive for generality, instead of focusing on the concentration of the multioverlaps associated with the Hamiltonian (3.29) in the stochastic block model, the general setting of optimal Bayesian inference is considered. Extending to this level of generality presents no additional difficulty, and the author suspects that the finitary restatement of the multioverlap concentration results in [15] provided in this chapter will be useful for the analysis of other statistical inference models. The general setting of optimal Bayesian inference is described in Section 5.1, where the main multioverlap concentration results of this chapter are also stated. In Section 5.2, the Franz-de Sanctis identities are introduced, and it is shown how these can be enforced through a small perturbation of the Hamiltonian that does not affect the limit of the free energy. The Franz-de Sanctis identities may be thought of as the Ghirlanda-Guerra identities of optimal Bayesian inference. A random probability measure that satisfies the Ghirlanda-Guerra identities must have an ultrametric support; a deep insight that leads to the appearance of the Poisson-Dirichlet probability cascades in many spin-glass models [50, 95]. The Franz-de Sanctis identities enforce a much simpler and more rigid structure on a random probability measure: all its multioverlaps must concentrate. This is discussed in Section 5.3, where a finitary version of the multioverlap concentration result in [15] is established. The most notable difference between this finitary multioverlap result and that in [15] is that it is uniform over an appropriate class of probability measures, and that multioverlap concentration is shown for any perturbation parameter satisfying a condition that may be verified in practice. The result in [15] holds for a given probability measure, and establishes multioverlap concentration on average over the set of perturbation parameters. This uniformity over random probability measures and additional control on the choice of perturbation parameters is essential in the proof of Theorem 1.6. This chapter is taken from Appendix C in [49].

5.1 Bayesian inference, perturbations, and key concentration results

The general optimal Bayesian inference problem is described following [15]. Consider a ground-truth signal $\sigma^* \in \Sigma_N := \{-1,+1\}^N$ with independent coordinates each generated from a distribution P_i^*. The prior distribution P^* of the model is thus the product law

$$\sigma^* \sim P^* := \prod_{i=1}^N P_i^*. \tag{5.1}$$

Data $\mathcal{D} := \mathcal{D}(\sigma^*)$ is sampled conditionally on the unknown signal σ^* from a probability distribution P_{out},

$$\mathcal{D} \sim P_{\text{out}}\{\cdot|\sigma^*\}, \tag{5.2}$$

and the inference task of the statistician consists of recovering the signal σ^* as accurately as possible given the data $\mathcal{D}$, the likelihood P_{out}, and the prior P^*. In this setting, the posterior of the model can be written explicitly using Bayes' formula as the Gibbs measure, or posterior distribution,

$$G_N(\mathrm{d}\sigma) := \mathbb{P}\{\sigma^* \in \mathrm{d}\sigma|\mathcal{D}\} = \frac{\exp H_N(\sigma)P^*(\mathrm{d}\sigma)}{\int_{\Sigma_N}\exp H_N(\tau)P^*(\mathrm{d}\tau)}, \tag{5.3}$$

associated with the Hamiltonian, or log-likelihood,

$$H_N(\sigma) := \log P_{\text{out}}\big\{\mathcal{D}|\sigma^* = \sigma\big\}. \tag{5.4}$$

Since the posterior distribution and all the hyperparameters of the model are known to the statistician, one speaks of *optimal* Bayesian inference. In addition to working in the context of optimal Bayesian inference, it will be assumed throughout this chapter that the Hamiltonian (5.4) satisfies symmetry between sites. This means that for any permutation ρ of the spin indices,

$$\mathbb{P}\{\sigma^* \in \mathrm{d}\sigma|\mathcal{D}\} \stackrel{d}{=} \mathbb{P}\{\rho(\sigma^*) \in \mathrm{d}\sigma|\mathcal{D}\}, \tag{5.5}$$

where $\stackrel{d}{=}$ denotes equality in distribution. Notice that both the sparse stochastic block model (3.13) and its enriched version (3.29) fall into the setting just described. For instance, in the enriched stochastic block model, one has $P_i^* := \mathrm{Ber}(p)$ and $\mathcal{D} := \widetilde{\mathcal{D}}^{t,\mu[-1,1],\overline{\mu}}$, where the data $\widetilde{\mathcal{D}}^{t,s,\mu}$ was defined in (3.25).

The concentration of the multioverlaps associated with the Hamiltonian (5.4) will be enforced through a small perturbation which will not affect the limit of its associated free energy,

$$\overline{F}_N := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp H_N(\sigma)\,\mathrm{d}P^*. \tag{5.6}$$

Fix an integer $K_+ \geqslant 1$ which will be kept implicit throughout this chapter, and write $\lambda := (\lambda_0, \lambda_1, \ldots, \lambda_{K_+})$ for a perturbation parameter with $\lambda_k \in [2^{-k-1}, 2^{-k}]$ for $0 \leqslant k \leqslant K_+$. Given a sequence $(\varepsilon_N)_{N\geqslant 1}$ with $\varepsilon_N := N^\gamma$ for some $-1/8 < \gamma < 0$, and a standard Gaussian vector $Z_0 := (Z_{0,1}, \ldots, Z_{0,N})$ in $\mathbb{R}^N$, introduce the Gaussian perturbation Hamiltonian

$$H_N^{\text{gauss}}(\sigma, \lambda_0) := \mathcal{H}_0 := \sum_{i\leqslant N}\big(\lambda_0\varepsilon_N\sigma_i^*\sigma_i + \sqrt{\lambda_0\varepsilon_N}Z_{0,i}\sigma_i\big) \tag{5.7}$$

associated with the task of recovering the signal σ^* from the data

$$Y^{\text{gauss}} := \sqrt{\lambda_0\varepsilon_N}\sigma^* + Z_0. \tag{5.8}$$

Notice that

$$1 \geqslant \varepsilon_N \to 0 \quad \text{and} \quad N\varepsilon_N \to +\infty. \tag{5.9}$$

Similarly, consider a sequence $(s_N)_{N\geqslant 1}$ with $s_N := N^\eta$ for $4/5 < \eta < 1$ in such a way that

$$\frac{s_N}{N} \to 0 \quad \text{and} \quad \frac{s_N}{\sqrt{N}} \to +\infty. \tag{5.10}$$

Fix a sequence of i.i.d. random variables $(\pi_k)_{k\geqslant 1}$ with $\mathrm{Poi}(s_N)$ distribution as well as a sequence $e := (e_{jk})_{j,k\geqslant 1}$ of i.i.d. random variables with $\mathrm{Exp}(1)$ distribution. For every $j \leqslant \pi_k$, sample i.i.d. random indices i_{jk} uniformly from the set $\{1,\dots,N\}$, and define the exponential perturbation Hamiltonians by

$$\mathcal{H}_k := \sum_{j\leqslant \pi_k}\left(\log(1+\lambda_k\sigma_{i_{jk}}) - \frac{\lambda_k e_{jk}\sigma_{i_{jk}}}{1+\lambda_k\sigma^*_{i_{jk}}}\right) \quad \text{and} \quad H_N^{\mathrm{exp}}(\sigma,\lambda) := \sum_{1\leqslant k\leqslant K_+}\mathcal{H}_k. \tag{5.11}$$

Observe that H_N^{exp} is the Hamiltonian associated with the task of recovering the signal σ^* from the data

$$Y_{jk}^{\mathrm{exp}} := \frac{e_{jk}}{1+\lambda_k\sigma^*_{i_{jk}}} \tag{5.12}$$

for $j \leqslant \pi_k$ and $k \geqslant 1$. Introduce the perturbed Hamiltonian

$$H_N(\sigma,\lambda) := H_N(\sigma) + H_N^{\mathrm{gauss}}(\sigma,\lambda_0) + H_N^{\mathrm{exp}}(\sigma,\lambda), \tag{5.13}$$

where the randomness of each Hamiltonian is independent of the randomness of the other Hamiltonians. Denote by

$$\overline{F}_N^{\mathrm{pert}}(\lambda) := \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp H_N(\sigma,\lambda)\,\mathrm{d}P^*(\sigma). \tag{5.14}$$

its associated free energy, and by $\langle\cdot\rangle$ its associated Gibbs measure. This means that for any bounded measurable function $f = f(\sigma^1,\dots,\sigma^n)$ of finitely many replicas,

$$\langle f(\sigma^1,\dots,\sigma^n)\rangle := \frac{\int_{\Sigma_N^n} f(\sigma^1,\dots,\sigma^n)\prod_{\ell\leqslant n}\exp H_N(\sigma^\ell,\lambda)\,\mathrm{d}P^*(\sigma^\ell)}{\left(\int_{\Sigma_N}\exp H_N(\sigma,\lambda)\,\mathrm{d}P^*(\sigma)\right)^n}. \tag{5.15}$$

Just like in the settings of the symmetric rank-one matrix estimation problem and of the sparse stochastic block model, this Gibbs measure is a conditional expectation in the sense of (3.32), and therefore satisfies the Nishimori identity (Proposition 2.2). As previously stated, an essential property of the perturbation Hamiltonians (5.7) and (5.11) is that they do not affect the asymptotic behaviour of the free energy (5.6).

Lemma 5.1. *The free energy* (5.6) *and the perturbed free energy* (5.14) *are asymptotically equivalent,*

$$\lim_{N\to+\infty}\left|\overline{F}_N^{\mathrm{pert}}(\lambda) - \overline{F}_N\right| = 0. \tag{5.16}$$

Proof. A direct computation reveals that

$$\left|\overline{F}_N^{\mathrm{pert}}(\lambda) - \overline{F}_N\right| \leqslant \frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N}\left|H_N^{\mathrm{gauss}}(\sigma,\lambda_0)\right| + \frac{1}{N}\mathbb{E}\max_{\sigma\in\Sigma_N}\left|H_N^{\mathrm{exp}}(\sigma,\lambda)\right|.$$

For any spin configuration $\sigma\in\Sigma_N$,

$$\left|H_N^{\mathrm{gauss}}(\sigma,\lambda_0)\right| \leqslant N\varepsilon_N + \sqrt{\varepsilon_N}\sum_{i\leqslant N}|Z_{0,i}| \quad \text{and} \quad \left|H_N^{\mathrm{exp}}(\sigma,\lambda)\right| \leqslant \sum_{1\leqslant k\leqslant K'}\sum_{j\leqslant\pi_k}\left(\log(1+\lambda_k) + \frac{\lambda_k e_{jk}}{1-\lambda_k}\right).$$

Since these bounds are uniform in σ, it follows that

$$\left|\overline{F}_N^{\mathrm{pert}}(\lambda) - \overline{F}_N\right| \leqslant \varepsilon_N + \sqrt{\varepsilon_N}\mathbb{E}Z_{0,1} + \frac{s_N}{N}\sum_{k\geqslant 1}\left(\log(1+\lambda_k) + \frac{\lambda_k}{1-\lambda_k}\right).$$

The third term was obtained by taking the expectation with respect to the randomness of e, and then with respect to the randomness of $(\pi_k)_{k\geqslant 1}$. Leveraging (5.9) and (5.10) to let N tend to infinity completes the proof. ■

In the context of the sparse stochastic block model, this result will imply that studying the free energy associated with the enriched Hamiltonian (3.29) or its perturbed version is equivalent for the purpose of computing the asymptotic mutual information. The advantage of studying the model associated with the perturbed Hamiltonian is that, for almost all perturbation parameters, the multioverlaps

$$R_{\ell_1,\ldots,\ell_n} := \frac{1}{N}\sum_{i\leqslant N}\sigma_i^{\ell_1}\cdots\sigma_i^{\ell_n} \tag{5.17}$$

can be shown to concentrate in the large-volume limit. Here, $(\sigma^\ell)_{\ell\geqslant 1}$ denotes a sequence of i.i.d. replicas sampled from the Gibbs measure (5.15).

To establish the concentration of the multioverlaps (5.17), the arguments in [15] will be followed closely. At this point, the reader may wonder why the results in [15] are not simply used directly. In [15] the concentration of the multioverlaps (5.17) is obtained for some perturbation parameter λ by showing that it holds on average over the set of perturbation parameters. In the proof of Theorem 1.6 it will be important to obtain multioverlap concentration for a specific perturbation parameter, so the existence of a perturbation parameter for which concentration holds does not suffice. However, following the strategy in [15], the main result of this chapter will propose a verifiable condition on a perturbation parameter λ which ensures the concentration of its associated multioverlaps, and this condition will be readily verified for the perturbation parameter of interest in the context of the sparse stochastic block model. The condition proposed is the asymptotic concentration of the quantities

$$\mathcal{L}_0 = \frac{\mathcal{H}_0'}{N\varepsilon_N}, \quad \text{where} \quad \mathcal{H}_0' = \partial_{\lambda_0} H_N^{\mathrm{gauss}}(\sigma,\lambda_0) = \varepsilon_N\left(\sigma\cdot\sigma^* + \frac{\sigma\cdot Z_0}{2\sqrt{\lambda_{0,N}}}\right), \tag{5.18}$$

$$\mathcal{L}_k = \frac{\mathcal{H}_k'}{s_N}, \quad \text{where} \quad \mathcal{H}_k' = \partial_{\lambda_k} H_N^{\exp}(\sigma,\lambda) = \sum_{j\leqslant \pi_k}\sigma_{i_{jk}}\left(\frac{1}{1+\lambda_k\sigma_{i_{jk}}} - \frac{e_{jk}}{(1+\lambda_k\sigma^*_{i_{jk}})^2}\right) \tag{5.19}$$

for $0\leqslant k\leqslant K_+$. More precisely, the concentration of the multioverlaps (5.17) will be established up to a small error for any sequence of perturbation parameters $(\lambda^N)_{N\geqslant 1}$ such that

$$\lim_{N\to+\infty}\mathbb{E}\big\langle(\mathcal{L}_k - \mathbb{E}\langle\mathcal{L}_k\rangle)^2\big\rangle = 0 \tag{5.20}$$

for $0\leqslant k\leqslant K_+$. Through a slight abuse of notation, the Gibbs average (5.15) associated with the perturbation parameters $(\lambda^N)_{N\geqslant 1}$ has been denoted by $\langle\cdot\rangle$. If necessary, this average will be written as $\langle\cdot\rangle_N$ to emphasize its dependence on N.

The concentration of the multioverlaps (5.17) associated with a sequence of perturbation parameters satisfying (5.20) will be obtained in two stages. In Section 5.2, it will be shown that the condition (5.20) implies the Franz-de Sanctis identities, and in Section 5.3 the concentration of the multioverlaps associated with any probability measure that satisfies an approximate version of the Franz-de Sanctis identities will be established.

To state the Franz-de Sanctis identities, it will be convenient to fix a uniform index $i\in\{1,\ldots,N\}$ and an exponential random variable $e\sim\mathrm{Exp}(1)$ independent of all other sources of randomness, and to introduce the random variables

$$y_{ik} := \frac{e}{1+\lambda_k \sigma_i^*}, \quad \theta_{ik}^\ell := \log(1+\lambda_k \sigma_i^\ell) - \lambda_k y_{ik} \sigma_i^\ell, \quad \text{and} \quad d_{ik}^\ell := \frac{y_{ik}\sigma_i^\ell}{1+\lambda_k \sigma_i^*} \tag{5.21}$$

for $1 \leqslant k \leqslant K_+$. The main result of Section 5.2 reads as follows.

Proposition 5.2 (Franz-de Sanctis identities in inference)**.** *For any $1 \leqslant k \leqslant K_+$ and any function f_n of finitely many spins on n replicas and of the signal σ^* with $\|f_n\|_{L^\infty} \leqslant 1$,*

$$\mathbb{E}\big\langle (R_1 - \mathbb{E}\langle R_1\rangle)^2\big\rangle \leqslant \frac{1}{N}, \qquad \mathbb{E}\big\langle (R_{1,2} - \mathbb{E}\langle R_{1,2}\rangle)^2\big\rangle \leqslant 4\mathbb{E}\big\langle (\mathcal{L}_0 - \mathbb{E}\langle \mathcal{L}_0\rangle)^2\big\rangle, \tag{5.22}$$

$$\left|\mathbb{E}\frac{\langle f_n d_{ik}^1 \exp\big(\sum_{\ell\leqslant n}\theta_{ik}^\ell\big)\rangle}{\langle \exp(\theta_{ik})\rangle^n} - \mathbb{E}\langle f_n\rangle \mathbb{E}\frac{\langle d_{ik}\exp(\theta_{ik})\rangle}{\langle \exp(\theta_{ik})\rangle}\right| \leqslant \left(2\mathbb{E}\big\langle (\mathcal{L}_k - \mathbb{E}\langle \mathcal{L}_k\rangle)^2\big\rangle + \frac{16}{s_N}\right)^{1/2}. \tag{5.23}$$

Applying this result along a sequence of perturbation parameters $(\lambda^N)_{N\geqslant 1}$ satisfying (5.20) reveals that

$$\lim_{N\to+\infty}\left|\mathbb{E}\frac{\langle f_n d_{ik}^1 \exp\big(\sum_{\ell\leqslant n}\theta_{ik}^\ell\big)\rangle_N}{\langle \exp(\theta_{ik})\rangle_N^n} - \mathbb{E}\langle f_n\rangle_N \mathbb{E}\frac{\langle d_{ik}\exp(\theta_{ik})\rangle_N}{\langle \exp(\theta_{ik})\rangle_N}\right| = 0 \tag{5.24}$$

for any $1 \leqslant k \leqslant K_+$ and any function f_n of finitely many spins on n replicas and of the signal σ^* with $\|f_n\|_{L^\infty} \leqslant 1$. Observing that the denominators $\langle \exp(\theta_{ik})\rangle_N$ do not depend on the signal σ^*, it is possible to use the Nishimori identity to replace all occurrences of the signal σ^* in (5.24) by another replica. For convenience of notation, this new replica will be denoted by $\sigma^\diamond$ to distinguish it from the signal σ^* and at the same time not occupy any specific index. The equations in (5.24) now read

$$\lim_{N\to+\infty}\left|\mathbb{E}\mathbb{E}_\diamond\frac{\langle f_n d_{ik}^1 \exp\big(\sum_{\ell\leqslant n}\theta_{ik}^\ell\big)\rangle_N}{\langle \exp(\theta_{ik})\rangle_N^n} - \mathbb{E}\mathbb{E}_\diamond\langle f_n\rangle_N \mathbb{E}\mathbb{E}_\diamond\frac{\langle d_{ik}\exp(\theta_{ik})\rangle_N}{\langle \exp(\theta_{ik})\rangle_N}\right| = 0, \tag{5.25}$$

where $\mathbb{E}_\diamond$ denotes the Gibbs average with respect to the replica $\sigma^\diamond$ only, the bracket $\langle\cdot\rangle_N$ denotes the Gibbs average with respect to all other standard replicas, the function f_n depends on finitely many spins on the n standard replicas and $\sigma^\diamond$, and, with some abuse of notation, for $1 \leqslant k \leqslant K_+$,

$$y_{ik} := \frac{e}{1+\lambda_k \sigma_i^\diamond}, \quad \theta_{ik}^\ell := \log(1+\lambda_k \sigma_i^\ell) - \lambda_k y_{ik} \sigma_i^\ell, \quad \text{and} \quad d_{ik}^\ell := \frac{y_{ik}\sigma_i^\ell}{1+\lambda_k \sigma_i^\diamond} \tag{5.26}$$

where $\lambda_k := \lambda_k^N$. The expression (5.25) is now simplified for functions f_n that do not depend on the spin coordinate indexed by 1. Introduce the collection of functions

$$\begin{aligned}\mathsf{F}_n := \big\{&\text{functions } f_n \text{ of finitely many spins } \sigma_i^\ell, \sigma_i^\diamond \text{ with } 2 \leqslant i \leqslant N \\ &\qquad\text{of the } n \text{ standard replicas } (\sigma^\ell)_{\ell\leqslant n} \text{ and the special replica } \sigma^\diamond \text{ with } \|f_n\|_{L^\infty} \leqslant 1\big\},\end{aligned} \tag{5.27}$$

and the quantities

$$y_k := \frac{e}{1+\lambda_k \sigma_1^\diamond}, \quad \theta_k^\ell := \log(1+\lambda_k \sigma_1^\ell) - \lambda_k y_k \sigma_1^\ell, \quad \text{and} \quad d_k^\ell := \frac{y_k\sigma_1^\ell}{1+\lambda_k \sigma_1^\diamond}. \tag{5.28}$$

For functions $f_n \in \mathsf{F}_n$, the symmetry between sites (5.5) and the fact that $i \in \{2,\dots,N\}$ with overwhelming probability in the limit, allow the replacement of the uniform random index $i \in \{1,\dots,N\}$ by the index 1. The

Franz de-Sanctis identities together with assumption (5.20) therefore have the following important implication.

Corollary 5.3 (Asymptotic Franz-de Sanctis identities in inference)**.** *If* (5.20) *holds, then for every* $1 \leqslant k \leqslant K_+$ *and all functions* $f_n \in \mathsf{F}_n$*,*

$$\lim_{N\to+\infty} \mathbb{E}\big\langle (R_1 - \mathbb{E}\langle R_1\rangle_N)^2\big\rangle_N = 0, \qquad \lim_{N\to+\infty} \mathbb{E}\big\langle (R_{1,2} - \mathbb{E}\langle R_{1,2}\rangle_N)^2\big\rangle_N = 0, \tag{5.29}$$

$$\lim_{N\to+\infty} \left| \mathbb{E}\mathbb{E}_\diamond \frac{\langle f_n d_k^1 \exp\big(\sum_{\ell\leqslant n} \theta_k^\ell\big)\rangle_N}{\langle \exp(\theta_k)\rangle_N^n} - \mathbb{E}\mathbb{E}_\diamond \langle f_n\rangle_N \mathbb{E}\mathbb{E}_\diamond \frac{\langle d_k \exp(\theta_k)\rangle_N}{\langle \exp(\theta_k)\rangle_N} \right| = 0. \tag{5.30}$$

To establish multioverlap concentration, this result will essentially be complemented by the observation that the multioverlaps associated with any probability measure that satisfies the Franz-de Sanctis identities must concentrate. As previously mentioned, to study the stochastic block model, it will be important that this result be uniform over an appropriate class of random probability measures now described. For each integer $N \geqslant 1$, consider the set of random probability measures on Σ_N thought of as a subset of $\{-1,0,1\}^{\mathbb{N}}$,

$$\mathcal{G}_N := \Big\{ G \mid G \text{ is a random probability measure on } \Sigma_N \times \{0\}^{\mathbb{N}} \Big\}, \tag{5.31}$$

and introduce its subset

$$\mathcal{G}_N^s := \{ G \in \mathcal{G}_N \mid G \text{ satisfies symmetry between sites } \}. \tag{5.32}$$

A measure $G \in \mathcal{G}_N$ is said to satisfy *symmetry between sites* if, for any sequence of i.i.d. replicas $(\sigma^\ell)_{\ell\geqslant 1}$ sampled from G, any permutation ρ_1 on the finite set $\{1,\ldots,N\}$, and any permutation ρ_2 of finitely many indices

$$\big(\sigma_i^\ell\big)_{i,\ell\geqslant 1} \stackrel{d}{=} \big(\sigma_{\rho_1(i)}^{\rho_2(\ell)}\big)_{i,\ell\geqslant 1}, \tag{5.33}$$

where $\stackrel{d}{=}$ denotes equality in distribution. Notice that each Gibbs measure G_N defined by (5.3) can be thought of as an element of $\mathcal{G}_N$ by setting $\sigma_i = 0$ when $i > N$ for any replica $\sigma \in \Sigma_N$ sampled from G_N. In this way, the symmetry between sites in (5.33) and (5.5) coincide, so in fact $G_N \in \mathcal{G}_N^s$. This identification also suggests that the appropriate notion of the multioverlap (5.17) for a random probability measure $G \in \mathcal{G}_N$ should be

$$R_{\ell_1,\ldots,\ell_n} := \frac{1}{N} \sum_{i\leqslant N} \sigma_i^{\ell_1} \cdots \sigma_i^{\ell_n}, \tag{5.34}$$

where $(\sigma^\ell)_{\ell\geqslant 1}$ denotes a sequence of i.i.d. replicas sampled from the Gibbs measure G. Denoting by $\langle\cdot\rangle_G$ the average with respect to the random probability measure G, the main result of Section 5.3 reads as follows.

Proposition 5.4. *For every* $\varepsilon > 0$ *there exists* $\delta > 0$ *such that the following holds. Let* $N \geqslant \lfloor \delta^{-1} \rfloor$ *and* $G \in \mathcal{G}_N^s$ *be such that for all* $1 \leqslant k \leqslant K_+ := \lfloor \delta^{-1} \rfloor$ *and* $f_n \in \mathsf{F}_n$*,*

$$\mathbb{E}\big\langle (R_1 - \mathbb{E}\langle R_1\rangle_G)^2\big\rangle_G \leqslant \delta, \qquad \mathbb{E}\big\langle (R_{1,2} - \mathbb{E}\langle R_{1,2}\rangle_G)^2\big\rangle_G \leqslant \delta, \tag{5.35}$$

$$\left| \mathbb{E}\mathbb{E}_\diamond \frac{\langle f_n d_k^1 \exp\big(\sum_{\ell\leqslant n} \theta_k^\ell\big)\rangle_G}{\langle \exp(\theta_k)\rangle_G^n} - \mathbb{E}\mathbb{E}_\diamond \langle f_n\rangle_G \mathbb{E}\mathbb{E}_\diamond \frac{\langle d_k \exp(\theta_k)\rangle_G}{\langle \exp(\theta_k)\rangle_G} \right| \leqslant \delta. \tag{5.36}$$

Then for any $1 \leqslant m \leqslant \lfloor \varepsilon^{-1} \rfloor$*, one has* $\mathbb{E}\big\langle (R_{1,\ldots,m} - \mathbb{E}\langle R_{1,\ldots,m}\rangle_G)^2\big\rangle_G \leqslant \varepsilon$*.*

The reader eager to return to the study of the sparse stochastic block model may consider taking Corollary 5.3 and Proposition 5.4 for granted on first reading, and turning directly to Chapter 6 where they are applied.

5.2 Deducing the Franz-de Sanctis identities

To show that the asymptotic concentration (5.20) of the quantities $\mathcal{L}_k$ defined in (5.18) and (5.19) implies the Franz-de Sanctis identities as stated in Proposition 5.2, the concentration of the magnetization R_1 and of the overlap $R_{1,2}$ will be established first. The concentration of the magnetization will be immediate from the Nishimori identity (Proposition 2.2) while the concentration of the overlap will follow from the Gaussian integration by parts formula (see Theorem 4.6 in [50]).

Lemma 5.5. *For any integer $N \geqslant 1$,*

$$\mathbb{E}\big\langle (R_1 - \mathbb{E}\langle R_1\rangle)^2\big\rangle \leqslant \frac{1}{N}. \tag{5.37}$$

Proof. Applying the Nishimori identity reveals that

$$\mathbb{E}\big\langle (R_1 - \mathbb{E}\langle R_1\rangle)^2\big\rangle = \frac{1}{N^2}\sum_{i\leqslant N}\mathbb{E}\big(\sigma_i^* - \mathbb{E}\sigma_i^*\big)^2 \leqslant \frac{1}{N}.$$

This completes the proof. ■

Lemma 5.6. *For any integer $N \geqslant 1$,*

$$\mathbb{E}\big\langle (R_{1,2} - \mathbb{E}\langle R_{1,2}\rangle)^2\big\rangle \leqslant 4\mathbb{E}\big\langle (\mathcal{L}_0 - \mathbb{E}\langle \mathcal{L}_0\rangle)^2\big\rangle. \tag{5.38}$$

Proof. The proof is taken from the Appendix of [15], and it consists in testing the concentration of the overlap $R_{1,*} := \frac{\sigma\cdot\sigma^*}{N}$ against the Hamiltonian $\mathcal{L}_0$ using the Gaussian integration by parts formula. Recalling (5.18) reveals that

$$\mathbb{E}\big\langle (R_{1,*} - \mathbb{E}\langle R_{1,*}\rangle)(\mathcal{L}_0 - \mathbb{E}\langle \mathcal{L}_0\rangle)\big\rangle = \mathbb{E}\big\langle R_{1,*}(R_{1,*} - \mathbb{E}\langle R_{1,*}\rangle)\big\rangle + \frac{1}{2N\sqrt{\lambda_{0,N}}}\mathbb{E}\big\langle R_{1,*}(\sigma\cdot Z_0 - \mathbb{E}\langle \sigma\cdot Z_0\rangle)\big\rangle. \tag{5.39}$$

The Gaussian integration by parts formula and the Nishimori identity imply that

$$\mathbb{E}\langle \sigma\cdot Z\rangle = N\sqrt{\lambda_{0,N}}\big(1 - \mathbb{E}\langle R_{1,*}\rangle\big) \quad \text{and} \quad \mathbb{E}\big\langle R_{1,*}\sigma\cdot Z\big\rangle = N\sqrt{\lambda_{0,N}}\big(\mathbb{E}\langle R_{1,*}\rangle - \mathbb{E}\langle R_{1,*}\rangle^2\big).$$

Substituting these two equalities into (5.39) gives

$$\mathbb{E}\big\langle (R_{1,*} - \mathbb{E}\langle R_{1,*}\rangle)(\mathcal{L}_0 - \mathbb{E}\langle \mathcal{L}_0\rangle)\big\rangle = \frac{1}{2}\mathbb{E}\big\langle (R_{1,*} - \langle R_{1,*}\rangle)^2\big\rangle + \frac{1}{2}\mathbb{E}\big\langle (R_{1,*} - \mathbb{E}\langle R_{1,*}\rangle)^2\big\rangle.$$

(There seems to be a sign error in equation (5.2) of [15].) It follows by the Nishimori identity that

$$\mathbb{E}\big\langle (R_{1,*} - \mathbb{E}\langle R_{1,*}\rangle)(\mathcal{L}_0 - \mathbb{E}\langle \mathcal{L}_0\rangle)\big\rangle \geqslant \frac{1}{2}\mathbb{E}\big\langle (R_{1,2} - \mathbb{E}\langle R_{1,2}\rangle)^2\big\rangle.$$

Invoking the Cauchy-Schwarz inequality and the Nishimori identity completes the proof. ■

The main hurdle in establishing Proposition 5.2 is proving the Franz-de Sanctis identity (5.23). This is done similarly to the Ghirlanda-Guerra identities, by testing the concentration of the quantities

$$\widetilde{\mathcal{L}}_k := \frac{1}{s_N}\sum_{j\leqslant\pi_k}\frac{\sigma_{i_{jk}}e_{jk}}{(1+\lambda_k\sigma^*_{i_{jk}})^2} \tag{5.40}$$

defined for $1 \leqslant k \leqslant K_+$ against an arbitrary function of finitely many spins and the signal σ^*. Notice that $\widetilde{\mathcal{L}}_k$ is the second term in the sum defining each $\mathcal{L}_k$ in (5.19). The reason for focusing only on this second term is that the first term concentrates automatically by the Nishimori identity. This is the content of Proposition 3.4 in [15] which is reproduced here for completeness. A slightly simpler proof than that in [15] is presented. This proof was kindly shared with the author by Dmitry Panchenko.

Lemma 5.7. *For any $1 \leqslant k \leqslant K_+$ and every large enough $N \geqslant 1$,*

$$\mathbb{E}\Big\langle \big(\widetilde{\mathcal{L}}_k - \mathbb{E}\langle \widetilde{\mathcal{L}}_k\rangle\big)^2\Big\rangle \leqslant 2\mathbb{E}\big\langle (\mathcal{L}_k - \mathbb{E}\langle \mathcal{L}_k\rangle)^2\big\rangle + \frac{16}{s_N}. \tag{5.41}$$

Proof. Introduce the quantity

$$g(\sigma, \pi_k) := \sum_{j \leqslant \pi_k} \frac{\sigma_{i_{jk}}}{1 + \lambda_k \sigma_{i_{jk}}}$$

in such a way that $\widetilde{\mathcal{L}}_k = s_N^{-1} g(\sigma, \pi_k) - \mathcal{L}_k$. Write Var for the variance with respect to the measure $\mathbb{E}\langle \cdot \rangle$. Since the variance of a sum of two random variables is bounded by twice the sum of the variance of each of the random variables,

$$\mathrm{Var}\big(\widetilde{\mathcal{L}}_k\big) \leqslant 2\Big(\mathrm{Var}(\mathcal{L}_k) + \frac{1}{s_N^2}\mathrm{Var}(g)\Big). \tag{5.42}$$

By the Nishimori identity and a direct computation,

$$\mathrm{Var}\big(g(\sigma, \pi_k)\big) = \mathbb{E}\bigg(\sum_{j \leqslant \pi_k} \frac{\sigma^*_{i_{jk}}}{1 + \lambda_k \sigma^*_{i_{jk}}}\bigg)^2 - \bigg(\mathbb{E}\sum_{j \leqslant \pi_k} \frac{\sigma^*_{i_{jk}}}{1 + \lambda_k \sigma^*_{i_{jk}}}\bigg)^2. \tag{5.43}$$

Recalling that the coordinates of the signal σ^* are i.i.d., and averaging with respect to the randomness of the indices $(i_{jk})_{j,k \geqslant 1}$ reveals that

$$\begin{aligned}\mathbb{E}\bigg(\sum_{j \leqslant \pi_k} \frac{\sigma^*_{i_{jk}}}{1 + \lambda_k \sigma^*_{i_{jk}}}\bigg)^2 &= \frac{1}{N}\mathbb{E}\sum_{j,j' \leqslant \pi_k} \frac{1}{(1 + \lambda_k \sigma^*_1)^2} + \frac{N^2 - N}{N}\mathbb{E}\sum_{j,j' \leqslant \pi_k} \frac{\sigma^*_1 \sigma^*_2}{(1 + \lambda_k \sigma^*_1)(1 + \lambda_k \sigma^*_2)} \\ &\leqslant \frac{4\mathbb{E}\pi_k^2}{N} + \mathbb{E}\pi_k^2\Big(\mathbb{E}\frac{\sigma^*_1}{1 + \lambda_k \sigma^*_1}\Big)^2,\end{aligned}$$

where it has been used that $\sigma_1^2 = 1$ and $\lambda_k \leqslant 1/2$. Similarly,

$$\bigg(\mathbb{E}\sum_{j \leqslant \pi_k} \frac{\sigma^*_{i_{jk}}}{1 + \lambda_k \sigma^*_{i_{jk}}}\bigg)^2 = \big(\mathbb{E}\pi_k\big)^2\Big(\mathbb{E}\frac{\sigma^*_1}{1 + \lambda_k \sigma^*_1}\Big)^2.$$

Substituting these two bounds into (5.43), recalling (5.10), and choosing N large enough yields

$$\mathrm{Var}\big(g(\sigma, \pi_k)\big) \leqslant \frac{4\mathbb{E}\pi_k^2}{N} + \mathrm{Var}(\pi_k)\Big(\mathbb{E}\frac{\sigma^*_1}{1 + \lambda_k \sigma^*_1}\Big)^2 \leqslant 8 s_N.$$

Together with (5.42), this completes the proof. ∎

Proof of Proposition 5.2. The proof follows that of Theorem 3.3 in [15]; full details are not provided, and instead, the interested reader is encouraged to consult [15]. The Cauchy-Schwarz inequality and the fact that

$\|f_n\|_{L^\infty} \leqslant 1$ imply that

$$\left|\mathbb{E}\left\langle f_n\widetilde{\mathcal{L}}_k(\sigma^1)\right\rangle - \mathbb{E}\langle f_n\rangle\mathbb{E}\left\langle\widetilde{\mathcal{L}}_k(\sigma)\right\rangle\right| \leqslant \mathbb{E}\left\langle(f_n - \langle f_n\rangle)^2\right\rangle^{1/2}\mathbb{E}\left\langle\left(\widetilde{\mathcal{L}}_k - \mathbb{E}\langle\widetilde{\mathcal{L}}_k\rangle\right)^2\right\rangle^{1/2} \leqslant \mathbb{E}\left\langle\left(\widetilde{\mathcal{L}}_k - \mathbb{E}\langle\widetilde{\mathcal{L}}_k\rangle\right)^2\right\rangle^{1/2}.$$

By Lemma 5.7 it therefore suffices to prove that

$$\mathbb{E}\left\langle f_n\widetilde{\mathcal{L}}_k(\sigma^1)\right\rangle = \mathbb{E}\frac{\langle f_n d_{ik}^1\exp\left(\sum_{\ell\leqslant n}\theta_{ik}^\ell\right)\rangle}{\langle\exp(\theta_{ik})\rangle^n} \quad \text{and} \quad \mathbb{E}\left\langle\widetilde{\mathcal{L}}_k(\sigma)\right\rangle = \mathbb{E}\frac{\langle d_{ik}\exp(\theta_{ik})\rangle}{\langle\exp(\theta_{ik})\rangle}. \tag{5.44}$$

Since π_k is independent of all other sources of randomness, taking the expectation with respect to this random variable first shows that

$$\mathbb{E}\left\langle f_n\widetilde{\mathcal{L}}_k(\sigma^1)\right\rangle = \sum_{r\geqslant 1}\frac{s_N^{r-1}}{(r-1)!}\exp(-s_N)\mathbb{E}\left\langle f_n D_{1k}^1\right\rangle_{\pi_k=r}, \tag{5.45}$$

where $D_{1k}^1 := \sigma_{i_{1k}}^1 e_{1k}/(1+\lambda_k\sigma_{i_{1k}}^*)^2$. To simplify this expression, the first replica σ^1 appearing in each of the averages will be isolated. It will be convenient to introduce the quantities

$$\Theta_{jk}^\ell := \log(1+\lambda_k\sigma_{i_{jk}}^\ell) - \frac{\lambda_k e_{jk}\sigma_{i_{jk}}^\ell}{1+\lambda_k\sigma_{i_{jk}}^*} \quad \text{and} \quad \mathcal{H}_k^{r-1}(\sigma^\ell) := \sum_{2\leqslant j\leqslant r}\Theta_{jk}^\ell$$

as well as the partially perturbed Hamiltonian

$$H_N'(\sigma) := H_N(\sigma) + H_N^{\text{gauss}}(\sigma) + \sum_{\substack{1\leqslant k'\leqslant K_+\\ k'\neq k}}\mathcal{H}_{k'}, \tag{5.46}$$

where $\mathcal{H}_k$ is defined in (5.11). Denoting by $\langle\cdot\rangle_{\pi_k=r}'$ the Gibbs measure corresponding to the Hamiltonian $H_N'(\sigma) + \mathcal{H}_k^{r-1}(\sigma^\ell)$ shows that for each $r \geqslant 1$,

$$\mathbb{E}\left\langle f_n D_{1k}^1\right\rangle_{\pi_k=r} = \mathbb{E}\mathbb{E}_{i_{1k}}\mathbb{E}_{e_{1k}}\frac{\langle f_n D_{1k}^1\exp\left(\sum_{\ell\leqslant n}\Theta_{1k}^\ell\right)\rangle_{\pi_k=r}'}{(\langle\exp(\Theta_{1k})\rangle_{\pi_k=r}')^n}. \tag{5.47}$$

Since the uniform random variable i_{1k} and the exponential random variable e_{1k} no longer appear in the Gibbs average $\langle\cdot\rangle_{\pi_k=r}'$, they may be replaced by a uniform random variable $i \in \{1,\ldots,N\}$ and an exponential random variable $e \sim \text{Exp}(1)$ independent of all other sources of randomness as in the statement of the result. To emphasize this change, the random variables D_{1k}^1 and Θ_{1k}^ℓ are also replaced by $d_{ik}^1 := \sigma_i^1 e/(1+\lambda_k\sigma_i^*)^2$ and $\theta_{ik}^\ell := \log(1+\lambda_k\sigma_i^\ell) - \lambda_k\sigma_i^\ell e/(1+\lambda_k\sigma_i^*)$, respectively. Notice that this matches the definitions in (5.21). In this new notation (5.47) reads

$$\mathbb{E}\left\langle f_n D_{1k}^1\right\rangle_{\pi_k=r} = \mathbb{E}\frac{\langle f_n d_{ik}^1\exp\left(\sum_{\ell\leqslant n}\theta_{ik}^\ell\right)\rangle_{\pi_k=r}'}{(\langle\exp(\theta_{ik})\rangle_{\pi_k=r}')^n}.$$

Substituting this into (5.45) and making the change of variables $m = r-1$ reveals that

$$\mathbb{E}\left\langle f_n\widetilde{\mathcal{L}}_k(\sigma^1)\right\rangle = \sum_{m\geqslant 0}\frac{s_N^m}{m!}\exp(-s_N)\mathbb{E}\frac{\langle f_n d_{ik}^1\exp\left(\sum_{\ell\leqslant n}\theta_{ik}^\ell\right)\rangle_{\pi_k=m+1}'}{(\langle\exp(\theta_{ik})\rangle_{\pi_k=m+1}')^n}.$$

Notice that whenever $\pi_k = m+1$, the Hamiltonian defining the Gibbs average $\langle\cdot\rangle_{\pi_k=m+1}'$ is given by $H_N'(\sigma) + \mathcal{H}_k^m(\sigma)$.

This Hamiltonian has the same distribution as the Hamiltonian (5.13) defining the original Gibbs average $\langle\cdot\rangle_{\pi_k=m}$. It follows that

$$\mathbb{E}\big\langle f_n\widetilde{\mathcal{L}}_k(\sigma^1)\big\rangle = \sum_{m\geqslant 0}\frac{s_N^m}{m!}\exp(-s_N)\mathbb{E}\frac{\langle f_n d_{ik}^1\exp\big(\sum_{\ell\leqslant n}\theta_{ik}^\ell\big)\rangle_{\pi_k=m}}{(\langle\exp(\theta_{ik})\rangle_{\pi_k=m})^n} = \mathbb{E}\frac{\langle f_n d_{ik}^1\exp\big(\sum_{\ell\leqslant n}\theta_{ik}^\ell\big)\rangle}{\langle\exp(\theta_{ik})\rangle^n}.$$

This is the first equality in (5.44). The second equality in (5.44) is obtained by taking $n=1$ and $f_1=1$ in the first equality. ■

Arguing in the same way as after the statement of this result, the asymptotic Franz-de Sanctis identities stated in Corollary 5.3 are established. It is now shown that these identities imply the concentration of the multioverlaps (5.17).

5.3 Establishing finitary multioverlap concentration

The finitary version of the multioverlap concentration result stated in Proposition 5.4 is uniform over the class $\mathcal{G}_N^s$ of random probability measures defined in (5.32). Its proof proceeds by contradiction and closely follows Sections 3.5 and 3.7 of [15]. Suppose for the sake of contradiction that there exists some $\varepsilon>0$ such that no matter the choice of $\delta>0$, it is always possible to find some integer $N=N(\delta)\geqslant\lfloor\delta^{-1}\rfloor$ and some random probability measure $G=G(\delta)\in\mathcal{G}_N^s$ such that for all $1\leqslant k\leqslant K_+=\lfloor\delta^{-1}\rfloor$ and $f_n\in\mathsf{F}_n$,

$$\mathbb{E}\big\langle(R_1-\mathbb{E}\langle R_1\rangle_G)^2\big\rangle_G\leqslant\delta,\qquad \mathbb{E}\big\langle(R_{1,2}-\mathbb{E}\langle R_{1,2}\rangle_G)^2\big\rangle_G\leqslant\delta, \tag{5.48}$$

$$\left|\mathbb{E}\mathbb{E}_\diamond\frac{\langle f_n d_k^1\exp\big(\sum_{\ell\leqslant n}\theta_k^\ell\big)\rangle_G}{\langle\exp(\theta_k)\rangle_G^n}-\mathbb{E}\mathbb{E}_\diamond\langle f_n\rangle_G\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_k\exp(\theta_k)\rangle_G}{\langle\exp(\theta_k)\rangle_G}\right|\leqslant\delta, \tag{5.49}$$

and for which there exists some $1\leqslant m=m(\delta)\leqslant\lfloor\varepsilon^{-1}\rfloor$ with

$$\mathbb{E}\big\langle(R_{1,\dots,m}-\mathbb{E}\langle R_{1,\dots,m}\rangle_G)^2\big\rangle_G>\varepsilon. \tag{5.50}$$

Applying the Prokhorov theorem (Theorem A.20 in [50]) on the compact metric space $\{-1,0,+1\}^{\mathbb{N}^2}$, and noticing that there are only finitely many choices for $m=m(\delta)$, it is possible to find a subsequence with $\delta\to 0$ along which the distribution of the array $(\sigma_i^\ell)_{i,\ell\geqslant 1}$ under the averaged Gibbs measure $\mathbb{E}\langle\cdot\rangle_{G(\delta)}$ converges in the sense of finite-dimensional distributions, and along which (5.48) - (5.50) hold for every $k\geqslant 1$ and a fixed $1\leqslant m\leqslant\lfloor\varepsilon^{-1}\rfloor$. Since $N(\delta)\to+\infty$ and $G(\delta)\in\mathcal{G}_N^s$, in the limit, the distribution of spins will be a measure on $\{-1,+1\}^{\mathbb{N}^2}$ which will inherit the symmetry between sites (5.33). By the Aldous-Hoover representation (Theorem 1.4 in [95]), this symmetry implies the existence of some function $\sigma:[0,1]^4\to\{-1,+1\}$ with

$$\big(\sigma_i^\ell\big)_{i,\ell\geqslant 1}\overset{d}{=}\big(\sigma(w,u_\ell,v_i,x_{i,\ell})\big)_{i,\ell\geqslant 1}, \tag{5.51}$$

where w, $(u_\ell)_{\ell\geqslant 1}$, $(v_i)_{i\geqslant 1}$ and $(x_{i,\ell})_{i,\ell\geqslant 1}$ are i.i.d. uniform random variables on $[0,1]$, and where $\overset{d}{=}$ denotes equality in distribution. Since σ takes values in $\{-1,+1\}$, the distribution of the array $(\sigma_i^\ell)_{i,\ell\geqslant 1}$ is encoded by the function

$$\overline{\sigma}(w,u,v):=\mathbb{E}_{x_{i,\ell}}\sigma(w,u,v,x_{i,\ell})=\int_0^1\sigma(w,u,v,x)\,\mathrm{d}x. \tag{5.52}$$

Indeed, the last coordinate $x_{i,\ell}$ is a dummy variable that corresponds to flipping a biased coin to generate a Bernoulli random variable with expected value $\overline{\sigma}(w,u,v)$. To clarify this further, let $\mathrm{d}u$ and $\mathrm{d}v$ denote Lebesgue measure on $[0,1]$, and define the random probability measure

$$G := G_w := \mathrm{d}u \circ \big(u \mapsto \overline{\sigma}(w,u,\cdot)\big)^{-1} \tag{5.53}$$

on the space of functions of $v \in [0,1]$,

$$H := L^2\big([0,1],\mathrm{d}v\big) \cap \big\{\|\overline{\sigma}\|_{L^\infty} \leqslant 1\big\}, \tag{5.54}$$

equipped with the topology of $L^2([0,1],\mathrm{d}v)$. As described in Section 2 of [94], the whole process of generating spins can be broken into the following steps:

(i) generate the asymptotic Gibbs measure $G = G_w$ using the uniform random variable w;

(ii) consider an i.i.d. sequence $\overline{\sigma}^\ell = \overline{\sigma}(w,u_\ell,\cdot)$ of replicas from G, which are functions in H;

(iii) plug in i.i.d. uniform random variables $(v_i)_{i\geqslant 1}$ to obtain the array $\overline{\sigma}^\ell(v_i) = \overline{\sigma}(w,u_\ell,v_i)$;

(iv) use the random variables $(x_{i,\ell})_{i,\ell\geqslant 1}$ to generate $(\sigma_i^\ell)_{i,\ell\geqslant 1}$ by flipping a coin with expected value $\overline{\sigma}^\ell(v_i)$,

$$\sigma_i^\ell := 2\mathbf{1}\bigg\{x_{i,\ell} \leqslant \frac{1+\overline{\sigma}^\ell(v_i)}{2}\bigg\} - 1. \tag{5.55}$$

This suggests that the asymptotic Gibbs average $\langle\cdot\rangle$ should be the average with respect to the random variables $(u_\ell)_{\ell\geqslant 1}$ and $(x_{i,\ell})_{i,\ell\geqslant 1}$ that depend on the replica indices,

$$\langle\cdot\rangle := \mathbb{E}_{(u_\ell),(x_{i,\ell})}. \tag{5.56}$$

It also suggests that the asymptotic multioverlap should be

$$R^\infty_{\ell_1,\ldots,\ell_n}(w,(u_{\ell_j})_{j\leqslant n}) := \mathbb{E}_v \prod_{j\leqslant n} \overline{\sigma}(w,u_{\ell_j},v) = \int_0^1 \prod_{j\leqslant n} \overline{\sigma}(w,u_{\ell_j},v)\,\mathrm{d}v. \tag{5.57}$$

This intuition is confirmed by the two following results adapted from Section 3.5 and the Appendix in [15].

Lemma 5.8. *For any finite set of n replicas and every collection $\{\mathcal{C}_\ell\}_{\ell\leqslant n}$ of finite indices,*

$$\lim_{\delta\to 0} \mathbb{E}\prod_{\ell\leqslant n}\Big\langle \prod_{i\in\mathcal{C}_\ell}\sigma_i^\ell\Big\rangle_{G(\delta)} = \mathbb{E}_{w,(v_i)}\prod_{\ell\leqslant n}\Big\langle\prod_{i\in\mathcal{C}_\ell}\sigma_i^\ell\Big\rangle. \tag{5.58}$$

Proof. Let $\mathcal{C} := \{(i,\ell) \mid \ell\leqslant n \text{ and } i\in\mathcal{C}_\ell\}$. By the weak convergence of finite-dimensional marginal distributions,

$$\lim_{\delta\to 0} \mathbb{E}\prod_{\ell\leqslant n}\Big\langle \prod_{i\in\mathcal{C}_\ell}\sigma_i^\ell\Big\rangle_{G(\delta)} = \lim_{\delta\to 0}\mathbb{E}\Big\langle\prod_{(i,\ell)\in\mathcal{C}}\sigma_i^\ell\Big\rangle_{G(\delta)} = \mathbb{E}\prod_{(i,\ell)\in\mathcal{C}}\sigma(w,u_\ell,v_i,x_{i,\ell}) = \mathbb{E}_{w,(v_i)}\prod_{\ell\leqslant n}\Big\langle\prod_{i\in\mathcal{C}_\ell}\sigma_i^\ell\Big\rangle,$$

where the notation (5.56) has been used. ■

Lemma 5.9. *For any collection of sets $\{\mathcal{L}_i\}_{i\geqslant 1}$ only finitely many of which are not empty,*

$$\lim_{\delta\to 0}\mathbb{E}\Big\langle\prod_{i\geqslant 1}R_{\mathcal{L}_i}\Big\rangle_{G(\delta)} = \mathbb{E}_w\Big\langle\prod_{i\geqslant 1}R^\infty_{\mathcal{L}_i}\Big\rangle. \tag{5.59}$$

Proof. Write $N \geqslant \lfloor \delta^{-1} \rfloor$ for the unique integer with $G(\delta) \in \mathcal{G}_N^s$, and suppose without loss of generality that the sets $\mathcal{L}_i$ for $i \leqslant j$ are not empty while the sets $\mathcal{L}_i$ for $i > j$ are empty. From (5.34),

$$\mathbb{E}\Big\langle \prod_{i\geqslant 1} R_{\mathcal{L}_i}\Big\rangle_{G(\delta)} = \frac{1}{N^j}\sum_{i_1,\ldots,i_j} \mathbb{E}\Big\langle \prod_{\ell_1\in\mathcal{L}_1}\cdots\prod_{\ell_j\in\mathcal{L}_j} \sigma_{i_1}^{\ell_1}\cdots\sigma_{i_j}^{\ell_j}\Big\rangle_{G(\delta)}.$$

The number of terms in this sum for which at least two of the indices $i_1,\ldots,i_j$ are equal is of order N^{j-1}, and is therefore negligible in the limit. Moreover, the symmetry between sites (5.33) implies that whenever $i_1,\ldots,i_j$ are all distinct,

$$\mathbb{E}\Big\langle \prod_{\ell_1\in\mathcal{L}_1}\cdots\prod_{\ell_j\in\mathcal{L}_j} \sigma_{i_1}^{\ell_1}\cdots\sigma_{i_j}^{\ell_j}\Big\rangle_{G(\delta)} = \mathbb{E}\Big\langle \prod_{\ell_1\in\mathcal{L}_1}\cdots\prod_{\ell_j\in\mathcal{L}_j} \sigma_{1}^{\ell_1}\cdots\sigma_{j}^{\ell_j}\Big\rangle_{G(\delta)} = \mathbb{E}\Big\langle \prod_{i\geqslant 1}\prod_{\ell\in\mathcal{L}_i} \sigma_i^{\ell}\Big\rangle_{G(\delta)}.$$

(This seems to fix a small typo in the second-to-last display of the Appendix in [15]). Combining these two observations shows that

$$\lim_{\delta\to 0}\mathbb{E}\Big\langle \prod_{i\geqslant 1} R_{\mathcal{L}_i}\Big\rangle_{G(\delta)} = \mathbb{E}_{w,(u_\ell)}\prod_{i\geqslant 1}\mathbb{E}_{v_i}\prod_{\ell\in\mathcal{L}_i}\mathbb{E}_{x_{i,\ell}}\sigma(w,u_\ell,v_i,x_{i,\ell}) = \mathbb{E}_{w,(u_\ell)}\prod_{i\geqslant 1} R^\infty_{\mathcal{L}_i}.$$

This completes the proof. ∎

In the notation of (5.56) and (5.57), the asymptotic version of (5.48) and (5.49) therefore reads that for any $k \geqslant 1$ and $f_n \in \mathsf{F}_n$,

$$\mathbb{E}\langle (R_1^\infty)^2\rangle = \big(\mathbb{E}\langle R_1^\infty\rangle\big)^2, \qquad \mathbb{E}\langle (R_{1,2}^\infty)^2\rangle = \big(\mathbb{E}\langle R_{1,2}^\infty\rangle\big)^2, \tag{5.60}$$

$$\mathbb{E}\mathbb{E}_\diamond \frac{\langle f_n d_k^1 \exp\big(\sum_{\ell\leqslant n}\theta_k^\ell\big)\rangle}{\langle \exp(\theta_k)\rangle^n} = \mathbb{E}\mathbb{E}_\diamond\langle f_n\rangle \mathbb{E}\mathbb{E}_\diamond \frac{\langle d_k\exp(\theta_k)\rangle}{\langle\exp(\theta_k)\rangle}, \tag{5.61}$$

while the asymptotic version of (5.50) becomes that for some $1 \leqslant m \leqslant \lfloor \varepsilon^{-1}\rfloor$,

$$\mathbb{E}\Big\langle (R^\infty_{1,\ldots,m} - \mathbb{E}\langle R^\infty_{1,\ldots,m}\rangle)^2\Big\rangle > \varepsilon. \tag{5.62}$$

The two most important consequences of the identities (5.60) and (5.61) that will lead to multioverlap concentration are now derived. On the one hand, the concentration of the overlap implies that the system lies in a “thermal pure state”, and that the function $\overline{\sigma}(w,u,v)$ is therefore almost surely independent of u. The proof of this fact is taken from Theorem 3.1 in [91]. On the other hand, the asymptotic Franz-de Sanctis identities in (5.61) imply a decoupling property of the asymptotic Gibbs measure. This is lemma 3.5 in [15].

Lemma 5.10. *If* $\mathbb{E}\langle (R^\infty_{1,2})^2\rangle = (\mathbb{E}\langle R^\infty_{1,2}\rangle)^2$*, then for almost all* $u,v,w \in [0,1]$*,*

$$\overline{\sigma}(w,u,v) = \mathbb{E}_u\overline{\sigma}(w,u,v). \tag{5.63}$$

Proof. Denote by $\cdot$ the inner product on the Hilbert space (5.54),

$$\overline{\sigma}^1\cdot\overline{\sigma}^2 = \mathbb{E}_v\overline{\sigma}^1(w,u_1,v)\overline{\sigma}^2(w,u_2,v) = R^\infty_{1,2},$$

and observe that

$$0=\mathbb{E}_{w,(u_\ell)}\big(R^\infty_{1,2}\big)^2-\mathbb{E}_{w,(u_\ell)}R^\infty_{1,2}R^\infty_{3,4}=\mathbb{E}_w\mathrm{Var}_{(u_\ell)}\overline{\sigma}^1\cdot\overline{\sigma}^2.$$

It follows that for almost all $w\in[0,1]$, the inner product $\overline{\sigma}^1\cdot\overline{\sigma}^2$ of any two replicas sampled from the Gibbs measure G_w is constant almost surely. In other words, the measure G_w is concentrated on a single function which may depend on w. This completes the proof. ■

Lemma 5.11 (A decoupling lemma)**.** *Fix* $\lambda\in\{\lambda_k\mid k\geqslant 1\}$. *If* e_1,e_2 *are independent* $\mathrm{Exp}(1)$ *random variables and, for* $j=1,2$,

$$y_j:=\frac{e_j}{1+\lambda\sigma_j^\diamond},\quad \theta_j:=\log(1+\lambda\sigma_j)-\lambda y_j\sigma_j,\quad \text{and}\quad d_j:=\frac{y_j\sigma_j}{1+\lambda\sigma_j^\diamond},\tag{5.64}$$

then

$$\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_1\exp(\theta_1)d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\exp(\theta_2)\rangle}=\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_1\exp(\theta_1)\rangle}{\langle\exp(\theta_1)\rangle}\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_2)\rangle}.\tag{5.65}$$

Proof. The proof follows Lemma 3.5 in [15]. Fix $M>1$ large, and consider the set $A:=\{e_2\mid 0\leqslant e_2\leqslant M\}$ on which the random variable e_2 is bounded by M. On this set, one has $|\theta_j|\leqslant M_\lambda$ for some constant $M_\lambda<+\infty$, so the denominator on the right side of the expression

$$\frac{\langle d_1\exp(\theta_1)d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\exp(\theta_2)\rangle}=\frac{\langle d_1\exp(\theta_1)d_2\exp(\theta_2)\rangle/\langle\exp(\theta_1)\rangle}{\langle\exp(\theta_1)\exp(\theta_2)\rangle/\langle\exp(\theta_1)\rangle}\tag{5.66}$$

lies in the interval $I:=[\exp(-3M_\lambda),\exp(3M_\lambda)]$. At this point fix $\varepsilon>0$, and use the Weierstrass approximation theorem (Exercise A.6 in [50]) to find a polynomial $P(x):=\sum_{n=0}^r c_nx^n$ that uniformly approximates the function $x\mapsto 1/x$ on the compact interval I, within error ε. Observing that d_1, d_2, θ_1, and θ_2 are bounded on A by some constant that depends on M reveals that

$$\begin{aligned}\frac{\langle d_1\exp(\theta_1)d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\exp(\theta_2)\rangle}\mathbf{1}\{e_2\in A\}&=\frac{\langle d_1\exp(\theta_1)d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\rangle}P\bigg(\frac{\langle\exp(\theta_1)\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\rangle}\bigg)\mathbf{1}\{e_2\in A\}+\mathcal{O}(\varepsilon)\\&=\sum_{n=0}^r c_n\frac{\langle d_1\exp(\theta_1)d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\rangle}\bigg(\frac{\langle\exp(\theta_1)\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\rangle}\bigg)^n\mathbf{1}\{e_2\in A\}+\mathcal{O}(\varepsilon)\\&=\sum_{n=0}^r c_n\frac{\big\langle d_1^1\exp\big(\sum_{\ell\leqslant n+1}\theta_1^\ell\big)d_2^1\exp\big(\sum_{\ell\leqslant n+1}\theta_2^\ell\big)\big\rangle}{\langle\exp(\theta_1)\rangle^{n+1}}\mathbf{1}\{e_2\in A\}+\mathcal{O}(\varepsilon).\end{aligned}$$

Applying (5.61) for each $0\leqslant n\leqslant r$ to the function $f_{n+1}:=d_2^1\exp\big(\sum_{\ell\leqslant n+1}\theta_2^\ell\big)$ for a fixed e_2, and then averaging over the randomness of e_2 gives

$$\begin{aligned}\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_1\exp(\theta_1)d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_1)\exp(\theta_2)\rangle}\mathbf{1}\{e_2\in A\}&=\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_1\exp(\theta_1)\rangle}{\langle\exp(\theta_1)\rangle}\mathbb{E}\mathbb{E}_\diamond\langle d_2\exp(\theta_2)\rangle P\big(\langle\exp(\theta_2)\rangle\big)\mathbf{1}\{e_2\in A\}+\mathcal{O}(\varepsilon)\\&=\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_1\exp(\theta_1)\rangle}{\langle\exp(\theta_1)\rangle}\mathbb{E}\mathbb{E}_\diamond\frac{\langle d_2\exp(\theta_2)\rangle}{\langle\exp(\theta_2)\rangle}\mathbf{1}\{e_2\in A\}+\mathcal{O}(\varepsilon),\end{aligned}$$

where the second equality uses that P uniformly approximates $x\mapsto 1/x$ on the compact interval I, within error ε, and that $d_1,d_2,\theta_1,\theta_2$ are bounded on A by some constant that depends on M. Letting ε tend to zero, and then letting M tend to infinity completes the proof. ■

The second identity in (5.60) therefore implies that, instead of the equality in distribution (5.51), it is actually the case that

$$\left(\sigma_i^\ell\right)_{i,\ell\geqslant 1} \stackrel{d}{=} \left(\sigma(w, v_i, x_{i,\ell})\right)_{i,\ell\geqslant 1} \tag{5.67}$$

for any function $\sigma : [0,1]^3 \to \{-1,+1\}$ such that $\int_0^1 \sigma(w,v,x)\,\mathrm{d}x = \overline{\sigma}(w,v)$. In particular, the Gibbs average (5.56) simplifies to

$$\langle\cdot\rangle = \mathbb{E}_{(x_{i,\ell})} \tag{5.68}$$

while the multioverlap (5.57) becomes

$$R^\infty_{\ell_1,\ldots,\ell_n}(w) = \mathbb{E}_v \prod_{j\leqslant n} \overline{\sigma}(w,v) = \mathbb{E}_v\left(\overline{\sigma}(w,v)^n\right) = \int_0^1 \overline{\sigma}(w,v)^n \,\mathrm{d}v. \tag{5.69}$$

The absurd hypothesis (5.50) may now be contradicted, thus establishing Proposition 5.4. The calculations are very similar in spirit to those in [93, 94], and are taken from Theorem 2.2 in [15].

Proof of Proposition 5.4. The proof follows that of Theorem 2.2 in [15]; full details are not provided, and instead, the reader is encouraged to consult [15]. Recall from (5.67) that $\sigma_j = \sigma(w, v_j, x_j)$ and $\sigma_j^\diamond = \sigma(w, v_j, x_j^\diamond)$. Since all random variables indexed by $j = 1,2$ are independent, denoting by $\mathbb{E}_{|w} = \mathbb{E}_{(e_j),(v_j),x_j,x_j^\diamond}$ the conditional expectation given w, and introducing the random variable

$$Y(w) := \mathbb{E}_{|w} \frac{\langle d_1 \exp(\theta_1)\rangle}{\langle \exp(\theta_1)\rangle} = \mathbb{E}_{|w} \frac{y_1}{1+\lambda\sigma_1^\diamond} \frac{\langle \sigma_1 \exp(\theta_1)\rangle}{\langle \exp(\theta_1)\rangle},$$

which depends implicitly on λ through y_1 and θ_1, transforms (5.65) into

$$\mathbb{E}\Big(\mathbb{E}_{|w}\frac{\langle d_1 \exp(\theta_1)\rangle}{\langle \exp(\theta_1)\rangle}\Big)\Big(\mathbb{E}_{|w}\frac{\langle d_2 \exp(\theta_2)\rangle}{\langle \exp(\theta_2)\rangle}\Big) - \Big(\mathbb{E}\frac{\langle d_1 \exp(\theta_1)\rangle}{\langle \exp(\theta_1)\rangle}\Big)\Big(\mathbb{E}\frac{\langle d_2 \exp(\theta_2)\rangle}{\langle \exp(\theta_2)\rangle}\Big) = \mathbb{E}\mathrm{Var}_{|w} Y(w) = 0.$$

This means that $Y = \mathbb{E}Y$ almost surely. To exploit this fact, through a slight abuse of notation, write σ for σ_1 and observe that conditionally on $\sigma_1^\diamond$,

$$Y(w) = \mathbb{E}_{|w} \int_0^{+\infty} \langle \exp(-\lambda\sigma y)\rangle \frac{\langle \sigma(1+\lambda\sigma)\exp(-\lambda y\sigma)\rangle}{\langle (1+\lambda\sigma)\exp(-\lambda y\sigma)\rangle} y\exp(-y)\,\mathrm{d}y.$$

Using the analyticity of both

$$g_w : \gamma \mapsto g_w(\gamma) := \mathbb{E}_{|w} \int_0^{+\infty} \langle \exp(-\gamma\sigma y)\rangle \frac{\langle \sigma(1+\gamma\sigma)\exp(-\gamma y\sigma)\rangle}{\langle (1+\gamma\sigma)\exp(-\gamma y\sigma)\rangle} y\exp(-y)\,\mathrm{d}y$$

for a fixed w as well as its w-expectation $\mathbb{E}g_w(\gamma)$, it is possible to deduce that $Y(w) = \mathbb{E}Y$ for all λ in a small neighbourhood of the origin. With this in mind, introduce the random variable

$$Z(w) := \mathbb{E}_{|w} \int_0^{+\infty} \langle \sigma(1+\lambda\sigma)\exp(-\lambda\sigma)\rangle y\exp(-y)\,\mathrm{d}y$$

which is deterministic by the first identity in (5.60). This implies that the random variable

$$X(w) := \frac{Z(w)-Y(w)}{\lambda} = \mathbb{E}_{|w} \int_0^{+\infty} \langle \sigma\exp(-\lambda y\sigma)\rangle \frac{\langle \sigma(1+\lambda\sigma)\exp(-\lambda y\sigma)\rangle}{\langle (1+\lambda\sigma)\exp(-\lambda y\sigma)\rangle} y\exp(-y)\,\mathrm{d}y$$

is deterministic for all λ in a small neighbourhood of the origin. In particular, all its λ-derivatives are also independent of w. It will now be deduced from this observation that all multioverlaps concentrate. Given $n \geqslant 1$, applying $\frac{\partial^n}{\partial \lambda^n}$ to the denominator in the expression inside the integral defining X and evaluating at $\lambda = 0$ yields the term

$$n! R_{1,\dots,n+2} \mathbb{E} e(e-1)^n,$$

where e is an $\mathrm{Exp}(1)$ random variable. Since $\mathbb{E}e(e-1)^n = \mathbb{E}(e-1)^{n+1} + \mathbb{E}(e-1)^n > 0$ for all $n \geqslant 1$, the term obtained by applying all derivatives to the denominator in the expression inside the integral defining X produces the multioverlap $R^{\infty}_{1,\dots,n+2}$. If along the way a derivative of λ is applied to any factor other than the denominator, this will not create a new replica, so all those terms will produce a linear combination of multioverlaps on strictly less than $n+2$ replicas which by induction is assumed to be independent of w. This establishes the concentration of all multioverlaps and contradicts (5.50), thus completing the proof. ■

At the end of Chapter 3, two main obstacles stood in the way of asserting that the limit of the enriched free energy (3.30) should satisfy the infinite-dimensional Hamilton-Jacobi equation (3.91) subject to the initial condition (3.92). The first was that infinite-dimensional Hamilton-Jacobi equations of this form had not been well-studied in the literature. This was resolved in Chapter 4. The second was that the concentration (3.75) of the multioverlaps (3.41) was not expected to hold for all choices of parameters. This has been addressed in this chapter. The Hamilton-Jacobi approach can now be applied to the sparse stochastic block model to prove Theorems 1.6 and 1.7.

Chapter 6

A Hamilton-Jacobi approach to the sparse stochastic block model

In this chapter, the Hamilton-Jacobi approach is finally used to analyze the sparse stochastic block model and establish Theorems 1.6 and 1.7. In Section 6.1, the well-posedness result in Theorem 4.4 for infinite-dimensional Hamilton-Jacobi equations of the form (4.6) is leveraged to make sense of the Hamilton-Jacobi equation (3.91) derived in Chapter 3 for the sparse stochastic block model. The solution to this infinite-dimensional Hamilton-Jacobi equation is therefore defined as the limit of the solutions to a family of projected Hamilton-Jacobi equations each posed on a positive half-space as the dimension parameter tends to infinity. The key step in proving Theorem 1.6 is to show that an adequately projected version of the finite-volume free energy in the sparse stochastic block model is an approximate solution to each of these projected Hamilton-Jacobi equations in a sense similar to that in Proposition 2.27. The error term in this approximate Hamilton-Jacobi equation is controlled using similar arguments to those in Lemma 2.29, and, in particular, requires a result on the concentration of the free energy in the sparse stochastic block model. This concentration result is established in Section 6.2. In Section 6.3, this concentration result is combined with the multioverlap concentration results in Chapter 5 to show that an adequately projected and perturbed version of the enriched free energy in the sparse stochastic block model is an approximate solution to each of the projected Hamilton-Jacobi equations. Together with the comparison principle in Theorem 4.1, this leads to an upper bound on the limit of the free energy, and therefore a proof of Theorem 1.6. In Section 6.4, this upper bound on the limit of the free energy is combined with the infinite-dimensional Hopf-Lax formula in Theorem 4.5 and a simple interpolation argument to determine a variational formula for the limit of the free energy in the disassortative sparse stochastic block model, thereby proving Theorem 1.7. Finally, in Section 6.5, a brief discussion on the relation between the results of this thesis and other works in the literature is provided. Some of the author's perspectives on the hurdles that still need to be overcome before tackling the matching upper bound in Theorem 1.6 and closing Conjecture 1.4 are also given. This chapter parallels Section 2.5 of Chapter 2, and its contents rely heavily on Sections 4 and 5 of [49].

6.1 Revisiting the stochastic block model Hamilton-Jacobi equation

In Chapter 3, it was argued that the limit of the enriched free energy (3.30) in the sparse stochastic block model should be the unique solution to the infinite-dimensional Hamilton-Jacobi equation (3.91). At that point, this statement made little rigorous sense since it was still very unclear what it means to be a solution to the infinite-dimensional Hamilton-Jacobi equation (3.91). Using the results of Chapter 4, and specifically Theorem 4.4, this can now be clarified by verifying that the initial condition (3.92) in the sparse stochastic block model satisfies assumptions (**H2**), (**H3'**), and (**H4**). It will be convenient to recall the definition of the kernel g in (3.68), of the closed convex set $\mathcal{K}_{=a,K}$ in (4.21), and of its neighbourhood $\mathcal{K}'_{=a,K}$ in (4.22).

Lemma 6.1. *The initial condition ψ defined in* (3.92) *satisfies* (**H2**), (**H3'**)*, and* (**H4**) *with $a = 1$.*

Proof. Recall that Lemma 3.9 implies the existence of a constant $C < +\infty$ which depends only on c and Δ such that for every integer $N \geqslant 1$, $\mu \in \mathcal{M}_+$ and $t \geqslant 0$,

$$\left|D_\mu \overline{F}_N(t,\mu,x)\right| \leqslant C \quad \text{and} \quad \left|\partial_x D_\mu \overline{F}_N(t,\mu,x)\right| \leqslant C. \tag{6.1}$$

To establish (**H2**), notice that for every integer $N \geqslant 1$ and $\mu, \nu \in \mathcal{M}_+$,

$$\psi_N(\mu) - \psi_N(\nu) = \int_0^1 D_\mu \psi_N\big(\nu + t(\mu - \nu); \mu - \nu\big)\,\mathrm{d}t = \int_0^1 \int_{-1}^1 f_t(x)\,\mathrm{d}\big(\mu - \nu\big)(x)\,\mathrm{d}t$$

for the continuously differentiable function $f_t(x) := D_\mu \psi_N\big(\nu + t(\mu - \nu), x\big)$. To bound this integral by the total variation distance, let $\eta := \mu - \nu \in \mathcal{M}_s$, and use the Hahn-Jordan decomposition (Theorem 3.4 in [57]) to write $\eta = \eta^+ - \eta^-$ for measures $\eta^+, \eta^- \in \mathcal{M}_+$ with the property that for some measurable set $D \subseteq [-1,1]$ and all measurable sets $E \subseteq [-1,1]$, one has $\eta^+(E) = \eta(E \cap D) \geqslant 0$ and $\eta^-(E) = -\eta(E \cap D^c) \geqslant 0$. The triangle inequality and the first bound in (6.1) imply that

$$\begin{aligned} |\psi_N(\mu) - \psi_N(\nu)| &\leqslant \left|\int_0^1 \int_{-1}^1 f_t(x)\,\mathrm{d}\eta^+(x)\right| + \left|\int_0^1 \int_{-1}^1 f_t(x)\,\mathrm{d}\eta^-(x)\right| \leqslant C\big(\eta^+[-1,1] + \eta^-[-1,1]\big) \\ &\leqslant 2C\mathrm{TV}(\mu,\nu). \end{aligned}$$

Using Proposition 3.12 to let N tend to infinity establishes (**H2**). To prove (**H3'**) with $a = 1$, notice that by (4.12) and (3.65), for every $y \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$, there exists some probability measure $\mu^* \in \Pr[-1,1]$ with

$$\partial_{x_k} \psi_N^{(K)}(y) = \frac{1}{|\mathcal{D}_K|} D_\mu \psi_N\big(\mu_y^{(K)}, k\big) = \frac{1}{|\mathcal{D}_K|} G_{\mu^*}(k) + \mathcal{O}\big(N^{-1}\big).$$

If $\mu_K^* := \mu^{(K)}_{x^{(K)}(\mu^*)} \in \mathcal{M}_+^{(K)}$ denotes the projection of μ^* onto $\mathcal{M}_+^{(K)}$, then the mean value theorem implies that

$$|G_{\mu^*}(k) - G_{\mu_K^*}(k)| \leqslant \sum_{k' \in \mathcal{D}_K} \int_{k'}^{k'+2^{-K}} |g(ky) - g(kk')|\,\mathrm{d}\mu^*(y) \leqslant \frac{\|g'\|_{L^\infty}}{2^K},$$

where it has been used that $\mu_K^*(k') = \mu^*[k', k' + 2^{-K})$ for every dyadic $k' \in \mathcal{D}_K$. This means that

$$\partial_{x_k} \psi_N^{(K)}(y) = \frac{1}{|\mathcal{D}_K|} G_{\mu_K^*}(k) + \mathcal{O}\big(2^{-2K}\big) + \mathcal{O}\big(N^{-1}\big) = G^{(K)} x^{(K)}(\mu_K^*)_k + \mathcal{O}\big(2^{-2K}\big) + \mathcal{O}\big(N^{-1}\big)$$

so, for K large enough, one has $\nabla \psi_N^{(K)}(y) = w + \mathcal{O}\big(N^{-1}\big)$ for some $w \in \mathcal{K}'_{=1,K}$. At this point fix $c \in \mathbb{R}$ and $x, x' \in \mathbb{R}^d$ with $(x' - x) \cdot z \geqslant c$ for every $z \in \mathcal{K}'_{=1,K}$. The fundamental theorem of calculus reveals that $\psi_N^{(K)}(x') - \psi_N^{(K)}(x) \geqslant c + \mathcal{O}\big(N^{-1}\big)$. Using Proposition 3.12 to let N tend to infinity and invoking Proposition A.2 gives (**H3'**) with $a = 1$. Finally, (**H4**) is a consequence of Lemma 3.10 and Proposition 3.12. ∎

Together with Theorem 4.4, this result establishes the well-posedness of the infinite-dimensional Hamilton-Jacobi equation (3.91) stated in Theorem 1.5. To establish the lower bound on the mutual information in Theorem 1.6, it will be important to unwrap this well-posedness result and express the unique solution f to the Hamilton-Jacobi equation (3.91) as a limit of solutions to projected equations as in Theorem 4.2. Fix $b > 0$ large enough so that the modified kernel

$$\widetilde{g}_b(z) := g(z) + b \tag{6.2}$$

is strictly positive, and introduce the shifted initial condition $\widetilde{\psi}_b : \mathcal{M}_+ \to \mathbb{R}$ defined by

$$\widetilde{\psi}_b(\mu) := \psi(\mu) + b\int_{-1}^{1} \mathrm{d}\mu \tag{6.3}$$

as in (4.50) with $a = 1$. In the spirit of (4.131) - (4.134), introduce the symmetric matrix

$$\widetilde{G}_b^{(K)} := \frac{1}{|\mathcal{D}_K|^2}\big(\widetilde{g}_b(kk')\big)_{k,k'\in\mathcal{D}_K} \in \mathbb{R}^{\mathcal{D}_K\times\mathcal{D}_K}, \tag{6.4}$$

the projected cone

$$\widetilde{\mathcal{C}}_{b,K} := \Big\{\widetilde{G}_b^{(K)}x \in \mathbb{R}^{\mathcal{D}_K} \mid x \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}\Big\}, \tag{6.5}$$

the projected non-linearity $\widetilde{\mathsf{C}}_{b,K} : \widetilde{\mathcal{C}}_{b,K} \to \mathbb{R}$ defined by

$$\widetilde{\mathsf{C}}_{b,K}\big(\widetilde{G}_b^{(K)}x\big) := \frac{1}{2}\widetilde{G}_b^{(K)}x\cdot x = \frac{1}{2|\mathcal{D}_K|^2}\sum_{k,k'\in\mathcal{D}_K}\widetilde{g}_b(kk')x_k x_{k'}, \tag{6.6}$$

and the closed convex set

$$\widetilde{\mathcal{K}}_{=1,b,K} := \Big\{\widetilde{G}_b^{(K)}x \in \mathbb{R}^{\mathcal{D}_K} \mid x \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0} \text{ and } |||x|||_1 = 1\Big\}. \tag{6.7}$$

Recall the notation $B_{K,R}$ in (4.23) for the ball of radius R centred at the origin in $\mathcal{D}_K$ with respect to the normalized-$\ell^{1,*}$ norm. Invoking Proposition 4.8 gives a uniformly Lipschitz continuous and non-decreasing non-linearity $\widetilde{\mathsf{H}}_{b,K,R}$ which agrees with $\widetilde{\mathsf{C}}_{b,K}$ on $\widetilde{\mathcal{C}}_{b,K} \cap B_{K,R}$. Theorem 4.4 ensures that the Hamilton-Jacobi equation

$$\partial_t \widetilde{f}^{(K)}(t,x) = \widetilde{\mathsf{H}}_{b,K,R}\big(\nabla\widetilde{f}^{(K)}(t,x)\big) \quad \text{on} \quad \mathbb{R}_{>0}\times\mathbb{R}^{\mathcal{D}_K}_{>0} \tag{6.8}$$

subject to the initial condition $\widetilde{f}^{(K)}(0,\cdot) = \widetilde{\psi}_b^{(K)}(\cdot)$ defined according to (4.13), admits a unique solution $\widetilde{f}_{b,R}^{(K)} \in \mathfrak{L}_{\mathrm{unif}}$ which satisfies the Lipschitz bound

$$\sup_{t>0} |||\widetilde{f}_{b,R}^{(K)}(t,\cdot)|||_{\mathrm{Lip},1} = |||\widetilde{\psi}_b^{(K)}|||_{\mathrm{Lip},1} \leqslant \|\widetilde{\psi}_b\|_{\mathrm{Lip,TV}}, \tag{6.9}$$

and has its gradient in the closed convex set $\widetilde{\mathcal{K}}_{=1,b,K}$. It also guarantees that, provided $R > \|\widetilde{\psi}_b\|_{\mathrm{Lip,TV}}$, the solution to the infinite-dimensional Hamilton-Jacobi equation (3.91) is given by

$$f(t,\mu) := \lim_{K\to+\infty}\bigg(\widetilde{f}_{b,R}^{(K)}\big(t,x^{(K)}(\mu)\big) - b|||x^{(K)}(\mu)|||_1 - \frac{bt}{2}\bigg), \tag{6.10}$$

and that this limit does not depend on b or R.

To establish Theorem 1.6, the idea will be to show that an adequately projected, perturbed, and shifted version of the enriched free energy (3.30) is an approximate viscosity subsolution to the Hamilton-Jacobi equation (6.8). This will be achieved by deriving an approximate Hamilton-Jacobi equation for this quantity similar to that in Proposition 2.27. Controlling the error term in this approximate equation will require a concentration result for the perturbed free energy in the sparse stochastic block model.

6.2 Establishing the concentration of the free energy

The concentration of the free energy will be combined with the finitary multioverlap result in Chapter 5 to establish Theorem 1.6. This will mean that the concentration result required is for the perturbed enriched free energy (3.30). Fix an integer $K_+ \geqslant 1$ that will be chosen large enough in the course of this chapter, and write $\lambda := (\lambda_0, \lambda_1, \dots, \lambda_{K_+})$ for a perturbation parameter with $\lambda_k \in [2^{-k-1}, 2^{-k}]$ for $0 \leqslant k \leqslant K_+$. Fix a sequence $(\varepsilon_N)_{N\geqslant 1}$ with $\varepsilon_N := N^\gamma$ for some $-1/8 < \gamma < 0$ as well as a sequence $(s_N)_{N\geqslant 1}$ with $s_N := N^\eta$ for some $4/5 < \eta < 1$, and observe that these sequences satisfy (5.9) and (5.10). Recall the definition of the Gaussian perturbation Hamiltonian $H_N^{\mathrm{gauss}}(\sigma, \lambda_0) = \mathcal{H}_0$ in (5.7) and of the exponential perturbation Hamiltonians $\mathcal{H}_k$ and $H_N^{\exp}(\sigma, \lambda)$ in (5.11), and introduce the perturbed Hamiltonian

$$H_N(\sigma, \lambda) := H_N^{t,\mu}(\sigma) + H_N^{\mathrm{gauss}}(\sigma, \lambda_0) + H_N^{\exp}(\sigma, \lambda), \tag{6.11}$$

where the randomness of each of the Hamiltonians in the sum is independent of the randomness of the other Hamiltonians. Just like in the settings of the symmetric rank-one matrix estimation problem and of the sparse stochastic block model, the Gibbs measure (5.15) associated with this perturbed Hamiltonian is a conditional expectation in the sense of (3.32), and therefore satisfies the Nishimori identity (Proposition 2.2). Arguing as in Lemma 5.1 shows that the perturbed free energy associated with the Hamiltonian (6.11) is asymptotically equivalent to the enriched free energy (3.30) in the sparse stochastic block model. Through a slight abuse of notation, it will therefore be convenient to write

$$F_N(t, \mu, \lambda) := \frac{1}{N} \log \int_{\Sigma_N} \exp H_N(\sigma, \lambda) \, \mathrm{d}P_N^*(\sigma) \quad \text{and} \quad \overline{F}_N(t, \mu, \lambda) := \mathbb{E} F_N(t, \mu, \lambda) \tag{6.12}$$

for this perturbed free energy and its average. The main result of this section is that for each even $p \geqslant 2$, the concentration function

$$v_{N,p} := \sup \Big\{ \mathbb{E} \big| F_N(t, \mu, \lambda) - \overline{F}_N(t, \mu, \lambda) \big|^p \mid \lambda_k \in [2^{-k-1}, 2^{-k}] \text{ for all } k \geqslant 0 \Big\} \tag{6.13}$$

is of order $N^{-p/2}$. For simplicity of notation, the spatial component (3.20) of the enriched Hamiltonian (3.29) will be dropped, and the concentration of the free energy will instead be established for the perturbed Hamiltonian

$$H_N'(\sigma, \lambda) := H_N^t(\sigma) + H_N^{\mathrm{gauss}}(\sigma, \lambda_0) + H_N^{\exp}(\sigma, \lambda). \tag{6.14}$$

This means that the free energy

$$F_N' := \frac{1}{N} \log \int_{\Sigma_N} \exp H_N'(\sigma, \lambda) \, \mathrm{d}P_N^*(\sigma) \tag{6.15}$$

will be shown to concentrate about its average

$$\overline{F}'_N = \frac{1}{N}\mathbb{E}\log\int_{\Sigma_N}\exp H'_N(\sigma,\lambda)\,\mathrm{d}P^*_N(\sigma). \tag{6.16}$$

More precisely, for each even $p \geqslant 2$, the concentration function

$$\nu'_{N,p} := \sup\Big\{\mathbb{E}\big|F'_N - \overline{F}'_N\big|^p \mid \lambda_k \in [2^{-k-1}, 2^{-k}] \text{ for all } k \geqslant 0\Big\} \tag{6.17}$$

will be shown to be of order $N^{-p/2}$. A bound on the more general concentration function $\nu_{N,p}$ can be obtained identically, but the notation becomes too cumbersome for comfort. The main tool to establish the concentration of the perturbed free energy will be the generalized Efron-Stein inequality (Theorem 15.5 in [20]) stated here for convenience.

Lemma 6.2 (Generalized Efron-Stein inequality)**.** *Let $X := (X_1,\dots,X_n)$ and $X' := (X'_1,\dots,X'_n)$ be two independent copies of a vector of independent random variables, and let $f:\mathbb{R}^n\to\mathbb{R}$ be a measurable function. Introduce the random variable $Z := f(X)$, and for each $1\leqslant i\leqslant n$ let $Z'_i := f(X_1,\dots,X_{i-1},X'_i,X_{i+1},\dots,X_n)$. If $q\geqslant 2$, then there exists a constant $C<+\infty$ that depends only on q such that*

$$\mathbb{E}|Z-\mathbb{E}Z|^q \leqslant C\mathbb{E}\Big|\sum_{i\leqslant n}\mathbb{E}_{X'}(Z-Z'_i)^2\Big|^{\frac{q}{2}}. \tag{6.18}$$

It will often be useful to make the dependence of the perturbed Hamiltonian (6.14) on one of its sources of randomness $\sigma^*, \Pi_t, \mathcal{I}_1 := (i_k, j_k)_{k\leqslant\Pi_t}, \mathcal{G} := (G^k_{i_k,j_k})_{k\leqslant\Pi_t}, e := (e_{jk}), \Pi' := (\pi_k)_{k\geqslant 0}, \mathcal{I}_2 := (i_{jk})_{j\leqslant\Pi'}$, and $Z := (Z_{0,i})_{i\leqslant N}$ explicit. Through a slight abuse of notation, it will be convenient to write $H'_N(X)$ when the dependence on the source of randomness X wants to be studied. A key observation that will be used repeatedly without further explanation is the following: given two sources of randomness X and X', a configuration-independent bound on the difference of the Hamiltonians $H'_N(X)$ and $H'_N(X')$,

$$\max_{\sigma\in\Sigma_N}\big|H'_N(X)-H'_N(X')\big| \leqslant Y, \tag{6.19}$$

gives a control by the possibly random Y on the difference of the free energy functionals $F'_N(X)$ and $F'_N(X')$ associated with these Hamiltonians,

$$\big|F'_N(X)-F'_N(X')\big| \leqslant \frac{Y}{N}. \tag{6.20}$$

The key concentration result on the perturbed free energy is the following.

Proposition 6.3. *For every even $p\geqslant 2$, there exists a constant $C<+\infty$ that depends only on p, c and Δ such that*

$$\nu'_{N,p} \leqslant \frac{C(1+t^p)}{N^{p/2}}. \tag{6.21}$$

Proof. To alleviate notation, write $C<+\infty$ for a constant that depends only on p, c, and Δ whose value might change during the argument. Given a source of randomness X, write $\mathbb{E}_X$ for the average with respect to the randomness of X. The proof will rely upon the generalized Efron-Stein inequality in Lemma 6.2 and the fact that

$$\mathbb{E} = \mathbb{E}_{\sigma^*}\mathbb{E}_Z\mathbb{E}_{\Pi'}\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\mathbb{E}_{\Pi_t}\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}.$$

To break down the proof into steps, introduce the averaged free energy functionals

$$\widehat{F}'_N = \mathbb{E}_{\Pi_t}\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}F'_N \quad \text{and} \quad \widetilde{F}'_N = \mathbb{E}_Z\mathbb{E}_{\Pi'}\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}'_N$$

in such a way that

$$\mathbb{E}\big(F'_N - \overline{F}'_N\big)^p \leqslant C\Big(\mathbb{E}\big(F'_N - \widehat{F}'_N\big)^p + \mathbb{E}\big(\widehat{F}'_N - \widetilde{F}'_N\big)^p + \mathbb{E}\big(\widetilde{F}'_N - \overline{F}'_N\big)^p\Big). \tag{6.22}$$

The proof now proceeds in three steps, each showing that one of the terms in this sum is of order $N^{-p/2}$.

Step 1: proving $\mathbb{E}\big(F'_N - \widehat{F}'_N\big)^p = \mathcal{O}\big((t/N)^{p/2}\big)$. The first term in (6.22) is bounded from above by

$$C\Big(\mathbb{E}\big(F'_N - \mathbb{E}_{\mathcal{G}|\sigma^*}F'_N\big)^p + \mathbb{E}\big(\mathbb{E}_{\mathcal{G}|\sigma^*}F'_N - \mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}F'_N\big)^p + \mathbb{E}\big(\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}F'_N - \widehat{F}'_N\big)^p\Big) =: C\big(I + II + III\big), \tag{6.23}$$

and each of the terms I, II and III is now bounded individually. By the generalized Efron-Stein inequality

$$I \leqslant C\mathbb{E}\bigg|\sum_{\ell\leqslant\Pi_t}\mathbb{E}_{\mathcal{G}^{(\ell)}|\sigma^*}\Big(F'_N(\mathcal{G}) - F'_N\big(\mathcal{G}^{(\ell)}\big)\Big)^2\bigg|^{p/2},$$

where $\big(\widetilde{G}^{(\ell)}_{i,j}\big)_{i,j\geqslant 1}$ is an independent copy of $\big(G^{(\ell)}_{i,j}\big)_{i,j\geqslant 1}$. Since $|\Delta| < c$ and all spin configuration coordinates are bounded by one,

$$\begin{aligned}\big|H'_N(\mathcal{G}) - H'_N\big(\mathcal{G}^{(\ell)}\big)\big| &\leqslant |\log(c + \Delta\sigma_{i_\ell}\sigma_{j_\ell})|\big|G^\ell_{i_\ell,j_\ell} - \widetilde{G}^\ell_{i_\ell,j_\ell}\big| + \Big|\log\Big(1 - \frac{c + \Delta\sigma_{i_\ell}\sigma_{j_\ell}}{N}\Big)\Big|\big|G^\ell_{i_\ell,j_\ell} - \widetilde{G}^\ell_{i_\ell,j_\ell}\big| \\ &\leqslant C\big|G^\ell_{i_\ell,j_\ell} - \widetilde{G}^\ell_{i_\ell,j_\ell}\big|.\end{aligned}$$

It follows that

$$\begin{aligned}I &\leqslant \frac{C}{N^p}\mathbb{E}\bigg|\sum_{\ell\leqslant\Pi_t}\mathbb{E}_{\mathcal{G}^{(\ell)}|\sigma^*}\big|G^\ell_{i_\ell,j_\ell} - \widetilde{G}^\ell_{i_\ell,j_\ell}\big|^2\bigg|^{p/2} \leqslant \frac{C}{N^p}\mathbb{E}\bigg|\sum_{\ell\leqslant\Pi_t}\Big(\Big(1 - \frac{2}{N}\Big)G^\ell_{i_\ell,j_\ell} + \frac{1}{N}\Big)\bigg|^{p/2} \\ &\leqslant \frac{C}{N^p}\bigg(\mathbb{E}\bigg|\sum_{\ell\leqslant\Pi_t}G^\ell_{i_\ell,j_\ell}\bigg|^{p/2} + \frac{1}{N^{p/2}}\mathbb{E}\Pi_t^{p/2}\bigg).\end{aligned}$$

By the Poisson colouring theorem (Proposition A.16), conditionally on σ^* and $\mathcal{I}_1$, the random variable $\sum_{\ell\leqslant\Pi_t}G^\ell_{i_\ell,j_\ell}$ is Poisson with mean $\mathbb{E}\Pi_t\frac{c+\Delta\sigma^*_{i_\ell}\sigma^*_{j_\ell}}{N}$. Invoking the Poisson moment bound in Lemma A.14 yields

$$I \leqslant \frac{C}{N^{p+\frac{p}{2}}}\mathbb{E}\Pi_t^{p/2} \leqslant \frac{Ct^{p/2}}{N^{p/2}}. \tag{6.24}$$

Another application of the generalized Efron-Stein inequality gives

$$II \leqslant \mathbb{E}\bigg|\sum_{\ell\leqslant\Pi_t}\mathbb{E}_{\mathcal{I}_1^{(\ell)}}\Big(\mathbb{E}_{\mathcal{G}|\sigma^*}F'_N(\mathcal{I}_1) - \mathbb{E}_{\mathcal{G}|\sigma^*}F'_N\big(\mathcal{I}_1^{(\ell)}\big)\Big)^2\bigg|^{p/2},$$

where $\mathcal{I}_1^{(\ell)}$ has an independent copy (i'_ℓ, j'_ℓ) of (i_ℓ, j_ℓ) at the ℓ'th coordinate but otherwise coincides with $\mathcal{I}_1$. Taylor expanding the logarithm and remembering that $G^\ell_{i_\ell,j_\ell} \in \{0,1\}$, it is readily verified that

$$\big|H'_N(\mathcal{I}_1)\big| \leqslant \big|G^\ell_{i_\ell,j_\ell}\big||\log(c + \Delta\sigma_{i_\ell}\sigma_{j_l})| + \big|1 - G^\ell_{i_\ell,j_\ell}\big|\Big|\log\Big(1 - \frac{c + \Delta\sigma_{i_\ell}\sigma_{j_\ell}}{N}\Big)\Big| \leqslant C\Big(\big|G^\ell_{i_\ell,j_\ell}\big| + \frac{1}{N}\Big).$$

This means that $\left|H_N'(\mathcal{I}_1)-H_N'\big(\mathcal{I}_1^{(\ell)}\big)\right|\leqslant C\big(|G^\ell_{i_\ell,j_\ell}|+|G^\ell_{i'_\ell,j'_\ell}|+\frac{1}{N}\big)$, and therefore

$$\Big(\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\mathcal{I}_1)-\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'\big(\mathcal{I}_1^{(\ell)}\big)\Big)^2\leqslant \frac{C}{N^2}\Big(\mathbb{E}_{\mathcal{G}|\sigma^*}\left|G^\ell_{i_\ell,j_\ell}\right|+\mathbb{E}_{\mathcal{G}|\sigma^*}\left|G^\ell_{i'_\ell,j'_\ell}\right|+\frac{1}{N}\Big)^2\leqslant\frac{C}{N^4}.$$

It follows that

$$II\leqslant\frac{C}{N^{2p}}\mathbb{E}\Pi_t^{p/2}\leqslant\frac{Ct^{p/2}}{N^p}\leqslant\frac{Ct^{p/2}}{N^{p/2}}. \tag{6.25}$$

A final application of the generalized Efron-Stein inequality reveals that

$$III\leqslant\mathbb{E}\Big|\mathbb{E}_{\Pi_t'}\big(\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\Pi_t)-\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\Pi_t')\big)^2\Big|^{p/2},$$

where Π_t' is an independent copy of Π_t. Slightly abusing notation and redefining Π_t' to be the maximum between Π_t and Π_t', it is readily verified that

$$\begin{aligned}\left|H_N'(\Pi_t')-H_N'(\Pi_t)\right|&\leqslant\sum_{\Pi_t\leqslant k\leqslant\Pi_t'}\bigg(\left|G^k_{i_k,j_k}\right|\left|\log(c+\Delta\sigma_{i_k}\sigma_{j_k})\right|+\left|1-G^k_{i_k,j_k}\right|\left|\log\Big(1-\frac{c+\Delta\sigma_{i_k}\sigma_{j_k}}{N}\Big)\right|\bigg)\\&\leqslant C\sum_{\Pi_t\leqslant k\leqslant\Pi_t'}\Big(\left|G^k_{i_k,j_k}\right|+\frac{1}{N}\Big).\end{aligned}$$

It follows that

$$\left|\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\Pi_t)-\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\Pi_t')\right|\leqslant\frac{C}{N}\mathbb{E}_{\mathcal{I}_1}\mathbb{E}_{\mathcal{G}|\sigma^*}\sum_{\Pi_t\leqslant k\leqslant\Pi_t'}\Big(\left|G^k_{i_k,j_k}\right|+\frac{1}{N}\Big)\leqslant\frac{C}{N^2}|\Pi_t'-\Pi_t|,$$

and by Jensen's inequality and the Poisson moment bound in Lemma A.14,

$$III\leqslant\frac{C}{N^{2p}}\mathbb{E}\Big|\mathbb{E}_{\Pi_t'}|\Pi_t'-\Pi_t|^2\Big|^{p/2}\leqslant\frac{C}{N^{2p}}\mathbb{E}|\Pi_t-\mathbb{E}\Pi_t|^p\leqslant\frac{Ct^{p/2}}{N^p}\leqslant\frac{Ct^{p/2}}{N^{p/2}}. \tag{6.26}$$

Combining (6.23)-(6.26) reveals that $\mathbb{E}\big(F_N'-\widehat{F}_N'\big)^p=\mathcal{O}\big((t/N)^{p/2}\big)$.

Step 2: proving $\mathbb{E}\big(\widehat{F}_N'-\widetilde{F}_N'\big)^p=\mathcal{O}\big(N^{-p/2}\big)$. The second term in (6.22) is bounded from above by

$$\begin{aligned}&C\Big(\mathbb{E}\big(\widehat{F}_N'\mathbb{E}_e\widehat{F}_N'\big)^p+\mathbb{E}\big(\mathbb{E}_e\widehat{F}_N'-\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}_N'\big)^p+\mathbb{E}\big(\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}_N'-\mathbb{E}_{\Pi'}\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}_N'\big)^p+\mathbb{E}\big(\mathbb{E}_{\Pi'}\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}_N'-\widetilde{F}_N'\big)^p\Big)\\&=:C\big(I+II+III+IV\big),\end{aligned} \tag{6.27}$$

and each of the terms I, II, III, and IV is now bounded individually. By the generalized Efron-Stein inequality

$$I\leqslant\mathbb{E}\left|\sum_{k\geqslant0}\sum_{j\leqslant\pi_k}\mathbb{E}_{e^{(jk)}}\Big(\widehat{F}_N'(e)-\widehat{F}_N'\big(e^{(jk)}\big)\Big)^2\right|^{p/2},$$

where $e^{(jk)}$ has an independent copy e_{jk}' of e_{jk} at the jk'th coordinate but otherwise coincides with e. Since

$$\left|H_N'(e)-H_N'\big(e^{(jk)}\big)\right|\leqslant\frac{|\lambda_k\sigma_{i_{jk}}|}{|1+\lambda_k\sigma^*_{i_{jk}}|}\left|e_{jk}-e_{jk}'\right|\leqslant\frac{\lambda_k}{1-\lambda_k}\left|e_{jk}-e_{jk}'\right|,$$

and $\lambda_k \in [2^{-k-1}, 2^{-k}]$,

$$I \leqslant \frac{C}{N^p}\mathbb{E}\Big|\sum_{k\geqslant 0}\frac{1}{2^k}\cdot\frac{1}{2^k}\sum_{j\leqslant \pi_k}\mathbb{E}_{e'_{jk}}\big|e_{jk}-e'_{jk}\big|^2\Big|^{p/2}.$$

It follows by two applications of Hölder's inequality and Jensen's inequality that

$$I \leqslant \frac{C}{N^p}\mathbb{E}\sum_{k\geqslant 0}\Big(\frac{1}{2^k}\sum_{j\leqslant \pi_k}\mathbb{E}_{e'_{jk}}\big|e_{jk}-e'_{jk}\big|^2\Big)^{p/2} \leqslant \frac{C}{N^p}\mathbb{E}\sum_{k\geqslant 0}\frac{1}{2^{\frac{kp}{2}}}\pi_k^{\frac{p}{2}-1}\sum_{j\leqslant \pi_k}\mathbb{E}_{e'_{jk}}\big|e_{jk}-e'_{jk}\big|^p.$$

Recalling that $e_{jk} \sim \mathrm{Exp}(1)$ while $\pi_{jk} \sim \mathrm{Poi}(s_N)$, and invoking the Poisson moment bound in Lemma A.14 gives

$$I \leqslant \frac{C}{N^p}\sum_{k\geqslant 0}\frac{1}{2^{\frac{kp}{2}}}\mathbb{E}\pi_k^{\frac{p}{2}} \leqslant C\Big(\frac{s_N}{N^2}\Big)^{p/2} \leqslant \frac{C}{N^{p/2}}. \tag{6.28}$$

Similarly, by the generalized Efron-Stein inequality,

$$II \leqslant C\mathbb{E}\bigg|\sum_{k\geqslant 0}\sum_{j\leqslant \pi_k}\mathbb{E}_{\mathcal{I}_2^{(jk)}}\Big(\mathbb{E}_e\widehat{F}'_N(\mathcal{I}_2)-\mathbb{E}_e\widehat{F}'_N\big(\mathcal{I}_2^{(jk)}\big)\Big)^2\bigg|^{p/2},$$

where $\mathcal{I}_2^{(jk)}$ has an independent copy i'_{jk} of i_{jk} at the jk'th coordinate but otherwise coincides with $\mathcal{I}_2$. By the mean value theorem,

$$\begin{aligned}\big|H'_N(\mathcal{I}_2)-H'_N\big(\mathcal{I}_2^{(jk)}\big)\big| &\leqslant |\log(1+\lambda_k\sigma_{i_{jk}})-\log(1+\lambda_k\sigma_{i'_{jk}})|+\lambda_k e_{jk}\bigg|\frac{\sigma_{i_{jk}}}{1+\lambda_k\sigma^*_{i_{jk}}}-\frac{\sigma_{i'_{jk}}}{1+\lambda_k\sigma^*_{i'_{jk}}}\bigg|\\ &\leqslant C\lambda_k(1+e_{jk}).\end{aligned}$$

It follows once again by two applications of Hölder's inequality and the Poisson moment bound in Lemma A.14 that

$$II \leqslant \frac{C}{N^p}\mathbb{E}\sum_{k\geqslant 0}\frac{1}{2^{\frac{kp}{2}}}\pi_k^{\frac{p}{2}-1}\sum_{j\leqslant \pi_k}(1+e_{jk})^p \leqslant C\Big(\frac{s_N}{N^2}\Big)^{p/2} \leqslant \frac{C}{N^{p/2}}. \tag{6.29}$$

Another application of the generalized Efron-Stein inequality yields

$$III \leqslant C\mathbb{E}\bigg|\sum_{k\geqslant 0}\mathbb{E}_{\Pi'^{(k)}}\Big(\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}'_N(\Pi')-\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}'_N\big(\Pi'^{(k)}\big)\Big)^2\bigg|^{p/2},$$

where $\Pi'^{(k)}$ has an independent copy π'_k of π_k at the k'th coordinate but otherwise coincides with Π'. Slightly abusing notation and redefining $\Pi'^{(k)}$ to be the process with the larger k'th coordinate, it is readily verified that

$$\begin{aligned}\big|H'_N(\Pi')-H'_N(\Pi'^{(k)})\big| &\leqslant \sum_{\pi_k\leqslant j\leqslant \pi'_k}\bigg|\log(1+\lambda_k\sigma_{i_{jk}})-\frac{\lambda_k e_{jk}\sigma_{i_{jk}}}{1+\lambda_k\sigma^*_{i_{jk}}}\bigg|\\ &\leqslant \sum_{\pi_k\leqslant j\leqslant \pi'_k}\bigg(\bigg|\lambda_k\sigma_{i_{jk}}-\frac{\lambda_k e_{jk}\sigma_{i_{jk}}}{1+\lambda_k\sigma^*_{i_{jk}}}\bigg|+C\lambda_k^2\bigg)\\ &\leqslant \sum_{\pi_k\leqslant j\leqslant \pi'_k}\bigg(\lambda_k\bigg|1-\frac{e_{jk}}{1+\lambda_k\sigma^*_{i_{jk}}}\bigg|+C\lambda_k^2\bigg)\\ &\leqslant \lambda_k\sum_{\pi_k\leqslant j\leqslant \pi'_k}\bigg(\frac{|1-e_{jk}|+\lambda_k}{1-\lambda_k}+C\lambda_k\bigg).\end{aligned}$$

It follows by two applications of the Cauchy-Schwarz inequality that

$$
\begin{aligned}
III &\leqslant \frac{C}{N^p}\mathbb{E}\sum_{k\geqslant 0}\frac{1}{2^{\frac{kp}{2}}}\bigg(\sum_{\pi_k\leqslant j\leqslant \pi_k'}\bigg(\frac{|1-e_{jk}|+\lambda_k}{1-\lambda_k}+C\lambda_k\bigg)\bigg)^p\\
&\leqslant \frac{C}{N^p}\mathbb{E}\sum_{k\geqslant 0}\frac{1}{2^{\frac{kp}{2}}}|\pi_k-\pi_k'|^{p-1}\sum_{\pi_k\leqslant j\leqslant \pi_k'}\bigg(\frac{|1-e_{jk}|+\lambda_k}{1-\lambda_k}+C\lambda_k\bigg)^p\\
&\leqslant \frac{C}{N^p}\sum_{k\geqslant 0}\frac{1}{2^{\frac{kp}{2}}}\mathbb{E}|\pi_k-\pi_k'|^p.
\end{aligned}
$$

Since $\mathbb{E}|\pi_k-\pi_k'|^p\leqslant C\mathbb{E}|\pi_k-\mathbb{E}\pi_k|^p\leqslant s_N^{p/2}$ by the Poisson moment bound in Lemma A.14, this implies that

$$
III\leqslant C\Big(\frac{s_N}{N^2}\Big)^{p/2}\leqslant \frac{C}{N^{p/2}}. \tag{6.30}
$$

A final application of the generalized Efron-Stein inequality gives

$$
IV\leqslant C\mathbb{E}\bigg|\sum_{i\leqslant N}\mathbb{E}_{Z^{(i)}}\Big(\mathbb{E}_{\Pi'}\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}_N'(Z)-\mathbb{E}_{\Pi'}\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\widehat{F}_N'\big(Z^{(i)}\big)\Big)^2\bigg|^{p/2},
$$

where $Z^{(i)}$ has an independent copy $Z_{i,0}'$ of $Z_{i,0}$ at the i'th coordinate but otherwise coincides with Z. Combining Hölder's inequality with the bound

$$
\big|H_N'(Z)-H_N'\big(Z^{(i)}\big)\big|\leqslant\sqrt{\lambda_0\varepsilon_N}\big|Z_{i,0}-Z_{i,0}'\big|
$$

reveals that

$$
IV\leqslant C\Big(\frac{\lambda_0\varepsilon_N}{N}\Big)^p\mathbb{E}\Big|\sum_{i\leqslant N}\mathbb{E}_{Z^{(i)}}\big|Z_{i,0}-Z_{i,0}'\big|^2\Big|^{p/2}\leqslant\frac{C}{N^p}N^{\frac{p}{2}-1}\sum_{i\leqslant N}\mathbb{E}\big|Z_{i,0}-Z_{i,0}'\big|^p\leqslant\frac{C}{N^{p/2}}. \tag{6.31}
$$

Together with (6.27)-(6.30), this shows that $\mathbb{E}\big(\widehat{F}_N'-\widetilde{F}_N'\big)^p=\mathcal{O}\big(N^{-p/2}\big)$.

Step 3: proving $\mathbb{E}\big(\widetilde{F}_N'-\overline{F}_N'\big)^p=\mathcal{O}\big((t^2/N)^{p/2}\big)$. Controlling the final term in (6.22) requires more care since $\widetilde{F}_N'$ depends on σ^* both through F_N' and through the conditional expectation $\mathbb{E}_{\mathcal{G}|\sigma^*}$. To simplify notation, write $\mathbb{E}'=\mathbb{E}_Z\mathbb{E}_{\Pi'}\mathbb{E}_{\mathcal{I}_2}\mathbb{E}_e\mathbb{E}_{\Pi_t}\mathbb{E}_{\mathcal{I}_1}$ in such a way that by the generalized Efron-Stein inequality

$$
\begin{aligned}
\mathbb{E}\big(\widetilde{F}_N'-\overline{F}_N'\big)^p&\leqslant C\mathbb{E}\bigg|\sum_{\ell\leqslant N}\mathbb{E}_{\sigma^{*,(\ell)}}\Big(\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\sigma^*)-\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^{*,(\ell)}}F_N'\big(\sigma^{*,(\ell)}\big)\Big)^2\bigg|^{p/2}\\
&\leqslant C\mathbb{E}\bigg|\sum_{\ell\leqslant N}\mathbb{E}_{\sigma^{*,(\ell)}}\Big(\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\sigma^*)-\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'\big(\sigma^{*,(\ell)}\big)\Big)^2\bigg|^{p/2}\\
&\quad+C\mathbb{E}\bigg|\sum_{\ell\leqslant N}\mathbb{E}_{\sigma^{*,(\ell)}}\Big(\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'\big(\sigma^{*,(\ell)}\big)-\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^{*,(\ell)}}F_N'\big(\sigma^{*,(\ell)}\big)\Big)^2\bigg|^{p/2}\\
&=:C(I+II),
\end{aligned}\tag{6.32}
$$

where $\sigma^{*,(\ell)}$ has an independent copy $\widetilde{\sigma}_\ell^*$ of σ_ℓ^* at the ℓ'th coordinate but otherwise coincides with σ^*. Since

$$\left|H_N'(\sigma^*)-H_N'(\sigma^{*,(l)})\right|\leqslant \lambda_0\varepsilon_N|\sigma_i|\left|\sigma_\ell^*-\widetilde{\sigma}_\ell^*\right|+\sum_{k\geqslant 1}\sum_{j:i_{jk}=\ell}\left|\frac{\lambda_k e_{jk}\sigma_\ell}{1+\lambda_k\sigma_\ell^*}-\frac{\lambda_k e_{jk}\sigma_\ell}{1+\lambda_k\widetilde{\sigma}_\ell^*}\right|\leqslant 2\varepsilon_N+\sum_{k\geqslant 1}\sum_{j:i_{jk}=\ell}\frac{2\lambda_k^2 e_{jk}}{(1-\lambda_k)^2},$$

and $\mathbb{E}e_{jk}=1$, the Fubini-Tonelli theorem and the basic properties of the multinomial distribution imply that

$$\begin{aligned}\left|\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'(\sigma^*)-\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'\big(\sigma^{*,(\ell)}\big)\right| &\leqslant \frac{2\varepsilon_N}{N}+\frac{1}{N}\mathbb{E}_{\Pi'}\sum_{k\geqslant 1}\frac{2\lambda_k^2}{(1-\lambda_k)^2}\mathbb{E}_{\mathcal{I}_2}|\{j:i_k=\ell\}|\\ &\leqslant \frac{2}{N}+\frac{1}{N^2}\mathbb{E}_{\Pi'}\sum_{k\geqslant 1}\frac{2\lambda_k^2}{(1-\lambda_k)^2}\pi_k\leqslant \frac{2}{N}+\frac{s_N}{N^2}\leqslant\frac{3}{N}.\end{aligned}$$

It follows that

$$I\leqslant C\Big(\frac{N}{N^2}\Big)^{p/2}=\frac{C}{N^{p/2}}. \tag{6.33}$$

To bound II an interpolation argument will be used. Fix $1\leqslant\ell\leqslant N$, and condition on all sources of randomness other than $\mathcal{G}$. For each $u\in[0,1]$, $G\in\{0,1\}^{\Pi_t}$, and $k\leqslant\Pi_t$, let

$$\begin{aligned}P_u^{1,k}(G)&:=G_k\left(\frac{c+\Delta\sigma_\ell^{*,u}\sigma_{j_k}^*}{N}\right)+(1-G_k)\left(1-\frac{c+\Delta\sigma_\ell^{*,u}\sigma_{j_k}^*}{N}\right),\\ P_u^{2,k}(G)&:=G_k\left(\frac{c+\Delta\sigma_\ell^{*,u}\sigma_{i_k}^*}{N}\right)+(1-G_k)\left(1-\frac{c+\Delta\sigma_\ell^{*,u}\sigma_{i_k}^*}{N}\right),\\ P_u^{3}(G)&:=G_k\left(\frac{c+\Delta(\sigma_\ell^{*,u})^2}{N}\right)+(1-G_k)\left(1-\frac{c+\Delta(\sigma_\ell^{*,u})^2}{N}\right),\end{aligned}$$

where $\sigma_\ell^{*,u}:=(1-u)\sigma_\ell^*+u\widetilde{\sigma}_\ell^*$. Write $\sigma^{*,u}$ for the configuration with ℓ'th coordinate $\sigma_\ell^{*,u}$ which otherwise coincides with σ^*, and introduce the sets

$$\begin{aligned}\mathcal{I}_1^1&:=\{k\mid i_k=l\text{ and }j_k\neq l\}, &\mathcal{I}_1^3&:=\{k\mid i_k=j_k=l\},\\ \mathcal{I}_1^2&:=\{k\mid i_k\neq l\text{ and }j_k=l\}, &\mathcal{I}_1^4&:=\{k\mid i_k\neq l\neq j_k\}.\end{aligned}$$

Let $\widetilde{\mathcal{G}}:=(\mathcal{G}_k)_{k\in\mathcal{I}_1^4}$, $\mathcal{G}^{(k)}:=\mathcal{G}\setminus\mathcal{G}_k$, $\widetilde{G}:=(G_k)_{k\in\mathcal{I}_1^4}$ and $G^{(k)}:=G\setminus G_k$, and define the interpolating free energy

$$\varphi(u):=\sum_{G\in\{0,1\}^{\Pi_t}}F_N'\big(\sigma^{*,(\ell)},G\big)\mathbb{P}\big\{\widetilde{\mathcal{G}}=\widetilde{G}|\sigma^*\big\}\cdot\prod_{k\in\mathcal{I}_1^1}P_u^{1,k}(G)\prod_{k\in\mathcal{I}_1^2}P_u^{2,k}(G)\prod_{k\in\mathcal{I}_1^3}P_u^{3,k}(G)$$

in such a way that $\varphi(1)=\mathbb{E}_{\mathcal{G}|\sigma^*}F_N'\big(\sigma^{*,(\ell)}\big)$ and $\varphi(0)=\mathbb{E}_{\mathcal{G}|\sigma^{*,(\ell)}}F_N'\big(\sigma^{*,(\ell)}\big)$. By the product rule,

$$\begin{aligned}\varphi'(u)&=\sum_{k\in\cup_{i\leqslant 3}\mathcal{I}_1^i}\sum_{G\in\{0,1\}^{\Pi_t}}F_N'\big(\sigma^{*,(\ell)},G\big)\mathbb{P}\big\{\mathcal{G}^{(k)}=G^{(k)}|\sigma^{*,u}\big\}(2G_k-1)\\ &\qquad\left(\frac{\Delta(\widetilde{\sigma}_\ell^*-\sigma_\ell^*)(\sigma_{j_k}^*\mathbf{1}\{k\in\mathcal{I}_1^1\}+\sigma_{i_k}^*\mathbf{1}\{k\in\mathcal{I}_1^2\}+2\sigma_\ell^{*,u}\mathbf{1}\{k\in\mathcal{I}_1^3\})}{N}\right)\\ &=\sum_{k\in\mathcal{I}_1^1}D_k\frac{\Delta(\widetilde{\sigma}_\ell^*-\sigma_\ell^*)\sigma_{j_k}^*}{N}+\sum_{k\in\mathcal{I}_1^2}D_k\frac{\Delta(\widetilde{\sigma}_\ell^*-\sigma_\ell^*)\sigma_{i_k}^*}{N}+\sum_{k\in\mathcal{I}_1^3}D_k\frac{2\Delta\big((1-u)\sigma_\ell^*+u\widetilde{\sigma}_\ell^*\big)(\widetilde{\sigma}_\ell^*-\sigma_\ell^*)}{N}\end{aligned}$$

for $D_k := \mathbb{E}_{\mathcal{G}^{(k)}|\sigma^{*,u}} F'_N\big(\sigma^{*,(\ell)}, \mathcal{G}^{(k)}, \mathcal{G}_k = 1\big) - \mathbb{E}_{\mathcal{G}^{(k)}|\sigma^{*,u}} F'_N\big(\sigma^{*,(\ell)}, \mathcal{G}^{(k)}, \mathcal{G}_k = 0\big)$. Since

$$\left|H'_N\big(\mathcal{G}^{(k)}, \mathcal{G}_k = 1\big) - H'_N\big(\mathcal{G}^{(k)}, \mathcal{G}_k = 0\big)\right| \leqslant \left|\log(c + \Delta\sigma_{i_k}\sigma_{j_k}) - \log\Big(1 - \frac{c + \Delta\sigma_{i_k}\sigma_{j_k}}{N}\Big)\right| \leqslant C,$$

one has $|D_k| \leqslant \frac{C}{N}$, so the fundamental theorem of calculus yields

$$\left|\mathbb{E}_{\mathcal{G}|\sigma^*} F'_N\big(\sigma^{*,(\ell)}\big) - \mathbb{E}_{\mathcal{G}|\sigma^{*,(\ell)}} F'_N\big(\sigma^{*,(l)}\big)\right| \leqslant \sup_{u \in (0,1)} |\varphi'(u)| \leqslant \frac{C}{N^2}\big(|\mathcal{I}_1^1| + |\mathcal{I}_1^2| + |\mathcal{I}_1^3|\big).$$

It follows by the basic properties of the binomial distribution that

$$\Big(\mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^*} F'_N\big(\sigma^{*,(\ell)}\big) - \mathbb{E}'\mathbb{E}_{\mathcal{G}|\sigma^{*,(\ell)}} F'_N\big(\sigma^{*,(\ell)}\big)\Big)^2 \leqslant \frac{C}{N^4}\sum_{1\leqslant i\leqslant 3}\big(\mathbb{E}_{\Pi_t}|\mathcal{I}_1^i|\big)^2 = \frac{C}{N^4}\sum_{1\leqslant i\leqslant 3}\Big(\mathbb{E}_{\Pi_t}\frac{\Pi_t}{N}\Big)^2 \leqslant \frac{Ct^2}{N^2},$$

and thus

$$II \leqslant C\bigg(\frac{Nt^2}{N^2}\bigg)^{p/2} = \frac{Ct^p}{N^{p/2}}. \tag{6.34}$$

Combining (6.32)-(6.34) reveals that $\mathbb{E}\big(\widetilde{F}'_N - \overline{F}'_N\big)^p = \mathcal{O}\big((t^2/N)^{p/2}\big)$. Together with (6.22) and the previous two steps, this completes the proof. ■

This concentration result for the perturbed free energy, or more precisely its extension establishing that the concentration function $\nu_{N,p}$ in (6.13) is of order $N^{-p/2}$, may finally be combined with the multioverlap concentration results in Chapter 5 and the derivative computations in Chapter 3 to prove the lower bound on the mutual information in Theorem 1.6.

6.3 Proving the free energy upper bound

In this section, the limit of the perturbed and enriched free energy (6.12) in the sparse stochastic block model is shown to be bounded from above by the unique solution to the infinite-dimensional Hamilton-Jacobi equation (3.91). Remembering that the perturbation does not affect the limit of the free energy, and recalling the relationship (3.17) between the free energy and the mutual information, this will give the lower bound on the mutual information stated in Theorem 1.6.

The upper bound on the free energy will be established by showing that a suitably projected and shifted version of the perturbed and enriched free energy (6.12) is an approximate subsolution to the Hamilton-Jacobi equation (6.8). For every integer $K \geqslant 1$, $t \geqslant 0$, $x \in \mathbb{R}^{\mathcal{D}_K}_{\geqslant 0}$, and perturbation parameter λ, define the projected version of the perturbed and enriched free energy (6.12) by

$$F_N^{(K)}(t,x,\lambda) := F_N\big(t, \mu_x^{(K)}, \lambda\big) \quad \text{and} \quad \overline{F}_N^{(K)}(t,x,\lambda) := \mathbb{E}F_N^{(K)}(t,x,\lambda). \tag{6.35}$$

In the same spirit as (6.10), given $b \in \mathbb{R}$ such that the kernel $\widetilde{g}_b$ defined in (6.2) is positive on $[-1,1]$, introduce shifted versions of the perturbed and enriched free energy functionals (6.12),

$$F'_N(t,\mu,\lambda) := F_N(t,\mu,\lambda) + b\int_{-1}^1 \mathrm{d}\mu + \frac{bt}{2} \quad \text{and} \quad \overline{F}'_N(t,\mu,\lambda) := \mathbb{E}F'_N(t,\mu,\lambda), \tag{6.36}$$

and for every integer $K \geqslant 1$, $t \geqslant 0$, $x \in \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$, and perturbation parameter λ, denote by

$$F_N'^{(K)}(t,x,\lambda) := F_N'\big(t,\mu_x^{(K)},\lambda\big) \quad \text{and} \quad \overline{F}_N'^{(K)}(t,x,\lambda) := \mathbb{E}F_N'^{(K)}(t,x,\lambda) \tag{6.37}$$

their finite-dimensional projections. Similarly, write

$$\overline{F}_N^{(K)}(t,x) = \overline{F}_N\big(t,\mu_x^{(K)}\big) \quad \text{and} \quad \overline{F}_N'^{(K)}(t,x) = \overline{F}_N^{(K)}(t,x) + b|||x|||_1 + \frac{bt}{2} \tag{6.38}$$

for the finite-dimensional projections of the enriched free energy (6.36) and its translation according to (6.10). Combining Lemmas 3.2, 3.5, and 3.7 with the Arzelà-Ascoli theorem, it is possible to extract a subsequential limit $\widetilde{F}^{(K)}$ from the sequence defined by the second term in (6.38) for varying N. Passing to a further subsequence, and using a diagonalization argument, it is also possible to ensure that for all $(t,x) \in \mathbb{R}_{>0} \times \mathbb{R}_{\geqslant 0}^{\mathcal{D}_K}$,

$$\widetilde{F}^{(K)}(t,x) = \limsup_{N\to+\infty} \overline{F}_N'^{(K)}(t,x). \tag{6.39}$$

The key to establishing Theorem 1.6 will be to show that, in some sense, the subsequential limit $\widetilde{F}^{(K)}$ is an approximate subsolution to the Hamilton-Jacobi equation (6.8) for some $R > \|\widetilde{\psi}_b\|_{\mathrm{Lip,TV}} + \|\widetilde{g}_b\|_{L^\infty} + \|\widetilde{g}_b'\|_{L^\infty} + 1$ which will remain fixed throughout this section.

Consider a smooth function $\phi \in C^\infty\big(\mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K};\mathbb{R}\big)$ with the property that $\widetilde{F}^{(K)} - \phi$ achieves a local maximum at some point $(t_\infty, x_\infty) \in \mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K}$. Recalling that the index K_+ controls the number of terms in the exponential perturbation Hamiltonian (5.11), introduce the parameter

$$\lambda_\infty := \frac{\big(1,2^{-1},2^{-2},\ldots,2^{-K_+}\big) + \big(2^{-1},2^{-2},2^{-3},\ldots,2^{-K_+-1}\big)}{2} \tag{6.40}$$

as well as the smooth function

$$\widetilde{\phi}(t,x,\lambda) := \phi(t,x) + (t-t_\infty)^2 + |x-x_\infty|^2 + |\lambda-\lambda_\infty|^2. \tag{6.41}$$

It is clear that $(t,x,\lambda) \mapsto \widetilde{F}^{(K)}(t,x) - \widetilde{\phi}(t,x,\lambda)$ has a strict local maximum at $(t_\infty, x_\infty, \lambda_\infty)$. Arguing as in the proof of Lemma 5.1 shows that $(t,x,\lambda) \mapsto \overline{F}_N'^{(K)}(t,x,\lambda)$ converges to $(t,x,\lambda) \mapsto \widetilde{F}^{(K)}(t,x)$ locally uniformly. Using Lemma 2.7, it is therefore possible to find a sequence $(t_N, x_N, \lambda_N)_{N\geqslant 1}$ which converges to the point $(t_\infty, x_\infty, \lambda_\infty)$ and has the property that $(t,x,\lambda) \mapsto \overline{F}_N'^{(K)}(t,x,\lambda) - \widetilde{\phi}(t,x,\lambda)$ attains a local maximum at (t_N, x_N, λ_N). More precisely, it is possible to find a constant $C < +\infty$ which is allowed to depend on K, K_+, t_∞, x_∞, and the function ϕ such that

$$\Big(\overline{F}_N'^{(K)} - \widetilde{\phi}\Big)(t_N, x_N, \lambda_N) = \sup\Big\{\Big(\overline{F}_N'^{(K)} - \widetilde{\phi}\Big)(t,x,\lambda) \mid |t-t_N| + |x-x_N| + |\lambda-\lambda_N| \leqslant C^{-1}\Big\}. \tag{6.42}$$

The constant $C < +\infty$ will be used at various places in this section, and it is understood that its value may need to be increased as the argument proceeds; the important point is that it does not depend on N. The choices of λ_∞ in (6.40) and $x_\infty \in \mathbb{R}_{>0}^{\mathcal{D}_K}$ ensure that when N is large enough $(\lambda_N)_k \in (2^{-k-1}, 2^{-k})$ for $0 \leqslant k \leqslant K_+$, and $x_N \in \mathbb{R}_{>0}^{\mathcal{D}_K}$. Increasing $C < +\infty$ if necessary, it is therefore possible to guarantee that for N large enough, the supremum on the right side of (6.42) is taken over triples (t,x,λ) with $t > 0$, $x \in \mathbb{R}_{>0}^{\mathcal{D}_K}$, and $\lambda_k \in [2^{-k-1}, 2^{-k}]$ for $1 \leqslant k \leqslant K_+$. It follows that

$$\partial_t\Big(\overline{F}_N'^{(K)} - \widetilde{\phi}\Big)(t_N, x_N, \lambda_N) = 0, \qquad \nabla_x\Big(\overline{F}_N'^{(K)} - \widetilde{\phi}\Big)(t_N, x_N, \lambda_N) = 0, \tag{6.43}$$

and

$$\nabla_\lambda\Big(\overline{F}_N^{(K)} - \widetilde{\phi}\Big)(t_N, x_N, \lambda_N) = \nabla_\lambda\Big(\overline{F}_N'^{(K)} - \widetilde{\phi}\Big)(t_N, x_N, \lambda_N) = 0. \tag{6.44}$$

The majority of this section will be devoted to leveraging the second equality in (6.44) and the multioverlap concentration results in Chapter 5 to establish the concentration of a finite but very large number of the multi-overlaps (3.41). This finitary multioverlap concentration result will then be combined with the computations in Chapter 3 to deduce the following approximate Hamilton-Jacobi equation for the projected free energy.

Lemma 6.4. *Fix $R > \|\widetilde{\psi}_b\|_{\mathrm{Lip,TV}} + \|\widetilde{g}_b\|_{L^\infty} + \|\widetilde{g}_b'\|_{L^\infty} + 1$. For every $\varepsilon > 0$, there exists a choice of integer $K_+ \geqslant 1$ in the perturbed Hamiltonian (6.11) with the property that for any integer $K \geqslant 1$, it is possible to find a constant $\mathcal{E}_{\varepsilon,K}$ with*

$$\limsup_{N\to+\infty}\Big|\Big(\partial_t \overline{F}_N'^{(K)} - \widetilde{\mathsf{H}}_{b,K,R}\Big(\nabla_x \overline{F}_N'^{(K)}\Big)\Big)(t_N, x_N, \lambda_N)\Big| \leqslant \mathcal{E}_{\varepsilon,K} \tag{6.45}$$

and $\lim_{\varepsilon\to 0}\lim_{K\to+\infty}\mathcal{E}_{\varepsilon,K} = 0$.

For the time being, Theorem 1.6 is proved assuming Lemma 6.4.

Proof of Theorem 1.6 assuming Lemma 6.4. Given $\varepsilon > 0$, invoke Lemma 6.4 to find an integer $K_+ \geqslant 1$ in the perturbed Hamiltonian (6.11) with the property that for any integer $K \geqslant 1$, it is possible to find a constant $\mathcal{E}_{\varepsilon,K}$ with

$$\limsup_{N\to+\infty}\Big|\Big(\partial_t \overline{F}_N'^{(K)} - \widetilde{\mathsf{H}}_{b,K,R}\Big(\nabla_x \overline{F}_N'^{(K)}\Big)\Big)(t_N, x_N, \lambda_N)\Big| \leqslant \mathcal{E}_{\varepsilon,K} \tag{6.46}$$

and $\lim_{\varepsilon\to 0}\lim_{K\to+\infty}\mathcal{E}_{\varepsilon,K} = 0$. Given an integer $K \geqslant 1$, the idea will be to show that the test function $\phi \in C^\infty\big(\mathbb{R}_{>0} \times \mathbb{R}_{>0}^{\mathcal{D}_K}; \mathbb{R}\big)$ introduced above satisfies the subsolution condition for the non-linearity $\widetilde{\mathsf{H}}_{b,K,R}$ at the contact point (t_∞, x_∞) up to the small error $\mathcal{E}_{\varepsilon,K}$. This will mean that the subsequential limit $\widetilde{F}^{(K)}$ of the modified free energy (6.38) is a viscosity subsolution to the Hamilton-Jacobi equation (6.8) up to a small error. The proof proceeds in two steps. First it is shown that the function

$$\widetilde{F}_\varepsilon^{(K)} := \widetilde{F}^{(K)} - t\mathcal{E}_{\varepsilon,K} \tag{6.47}$$

is a viscosity subsolution to the Hamilton-Jacobi equation (6.8), and then the comparison principle in Theorem 4.1 is leveraged to obtain an upper bound on the limit superior of the enriched free energy (3.30) by the solution f to the infinite-dimensional Hamilton-Jacobi equation (3.91) constructed in Section 6.1. The lower bound on the limit inferior of the mutual information then follows from the relationship (3.17) between the free energy and the mutual information.

Step 1: proving $\widetilde{F}_\varepsilon^{(K)}$ is a viscosity subsolution. Since $x_N \to x_\infty$ assume without loss of generality that $(x_N)_{N\geqslant 1} \subseteq \mathbb{R}_{>0}^{\mathcal{D}_K}$. It follows by (6.43) that

$$\big(\partial_t\widetilde{\phi} - \widetilde{\mathsf{H}}_{b,K,R}\big(\nabla_x\widetilde{\phi}\big)\big)(t_N, x_N, \lambda_N) = \Big(\partial_t \overline{F}_N'^{(K)} - \widetilde{\mathsf{H}}_{b,K,R}\Big(\nabla_x \overline{F}_N'^{(K)}\Big)\Big)(t_N, x_N, \lambda_N),$$

so letting N tend to infinity and combining the definition of $\widetilde{\phi}$ with (6.46) yields

$$\big(\partial_t\widetilde{\phi} - \widetilde{\mathsf{H}}_{b,K,R}\big(\nabla_x\widetilde{\phi}\big)\big)(t_\infty, x_\infty) \leqslant \mathcal{E}_{\varepsilon,K}.$$

This shows that the function (6.47) is a viscosity subsolution to the Hamilton-Jacobi equation (6.8).

Step 2: comparison principle. The comparison principle in Theorem 4.1 gives the upper bound

$$\widetilde{F}^{(K)}(t,x) \leqslant \widetilde{f}^{(K)}_{b,R}(t,x) + t\mathcal{E}_{\varepsilon,K}. \tag{6.48}$$

It has implicitly been used that $\widetilde{F}^{(K)}$ and $\widetilde{f}^{(K)}_{b,R}$ are continuous and have the same initial condition by Proposition 3.12. It has also been used that they both belong to the solution space $\mathfrak{L}_{\mathrm{unif}}$ by the discussion in Section 6.1, the derivative computations (3.39) and (3.60), and a simple application of the mean value theorem. With (6.48) in mind, fix a finite measure $\mu \in \mathcal{M}_+$, and introduce the discrete measure

$$\mu^{(K)} := \mu^{(K)}_{x^{(K)}(\mu)}$$

defined in (4.9). It is readily verified that $\overline{\mu}^{(K)} \to \overline{\mu}$ with respect to the Wasserstein distance (3.85). Moreover, an identical argument to that in Lemma 3.10 leveraging the second bound in (3.87) gives a constant $C' < +\infty$ that depends only on c such that

$$\overline{F}_N(t,\mu) \leqslant C'\mu[-1,1]W\big(\overline{\mu},\overline{\mu}^{(K)}\big) + \overline{F}^{(K)}_N\big(t,x^{(K)}(\mu)\big).$$

The letter C' is used instead of C to emphasize that the constant C' does not depend on K. Letting N tend to infinity, recalling (6.38) and (6.39), and leveraging (6.48) yields

$$\limsup_{N\to+\infty} \overline{F}_N(t,\mu) \leqslant C'\mu[-1,1]W\big(\overline{\mu},\overline{\mu}^{(K)}\big) + \widetilde{f}^{(K)}_{b,R}\big(t,x^{(K)}(\mu)\big) - b|||x^{(K)}(\mu)|||_1 - \frac{bt}{2} + t\mathcal{E}_{\varepsilon,K}.$$

Invoking Lemma 6.4 and remembering the definition of the solution f to the infinite-dimensional Hamilton-Jacobi equation (3.91) given in (6.10) to let K tend to infinity and then ε tend to zero establishes the upper bound

$$\limsup_{N\to+\infty} \overline{F}_N(t,\mu) \leqslant f(t,\mu). \tag{6.49}$$

Recalling that the free energy (3.14) in the sparse stochastic block model is given by $\overline{F}_N = \overline{F}_N(1,0)$, where 0 denotes the zero measure, and leveraging the relationship (3.17) between the free energy and the mutual information completes the proof. ∎

The rest of this section is devoted to the proof of Lemma 6.4 which will be established by combining the computations in Chapter 3 with the multioverlap concentration results in Chapter 5. For any perturbation parameter λ, let

$$\lambda_{0,N} := \varepsilon_N \lambda_0, \tag{6.50}$$

and recall the definition of the quantities $\mathcal{L}_0$ in (5.18) and $\mathcal{L}_k$ in (5.19). The importance of these quantities in the present context stems from the fact that for $1 \leqslant k \leqslant K_+$,

$$\partial_{\lambda_0} F^{(K)}_N(t,x,\lambda) = \frac{1}{N}\big\langle \mathcal{H}'_0 \big\rangle, \qquad \partial^2_{\lambda_0} F^{(K)}_N(t,x,\lambda) = \frac{1}{N}\Big(\big\langle (\mathcal{H}'_0 - \langle \mathcal{H}'_0\rangle)^2\big\rangle - \frac{\varepsilon_N^2}{4\lambda_{0,N}^{3/2}}\langle \sigma\rangle \cdot Z_0\Big), \tag{6.51}$$

$$\partial_{\lambda_k} F^{(K)}_N(t,x,\lambda) = \frac{1}{N}\big\langle \mathcal{H}'_k \big\rangle, \qquad \partial^2_{\lambda_k} F^{(K)}_N(t,x,\lambda) = \frac{1}{N}\Big(\big\langle (\mathcal{H}'_k - \langle \mathcal{H}'_k\rangle)^2\big\rangle + \big\langle \mathcal{H}''_k\big\rangle\Big), \tag{6.52}$$

while for $0 \leqslant j \neq k \leqslant K_+$,

$$\partial_{\lambda_k \lambda_j} F^{(K)}_N(t,x,\lambda) = \frac{1}{N}\big(\langle \mathcal{H}'_j \mathcal{H}'_k\rangle - \langle \mathcal{H}'_j\rangle\langle \mathcal{H}'_k\rangle\big). \tag{6.53}$$

Here, and for the remainder of this section, the Gibbs average $\langle\cdot\rangle$ will always be associated with the perturbed Hamiltonian (6.11) evaluated at a triple (t,x,λ) which will be clear from the context. It will also be convenient to record that for $1\leqslant k\leqslant K_+$,

$$\mathcal{H}_k'' := \partial_{\lambda_k}^2 H_N^{\exp}(\sigma,\lambda) = \sum_{j\leqslant\pi_k}\left(-\frac{1}{(1+\lambda_k\sigma_{i_{jk}})^2}+2\frac{\sigma_{i_{jk}}\sigma^*_{i_{jk}}e_{jk}}{(1+\lambda_k\sigma^*_{i_{jk}})^3}\right) \quad \text{and} \quad \left|\mathbb{E}\langle\mathcal{H}_k''\rangle\right|\leqslant Cs_N. \tag{6.54}$$

The key step in the proof of Lemma 6.4 is to obtain the concentration of the multioverlaps (3.41) from Propositions 5.2 and 5.4 by establishing the concentration (5.20) of the quantities $\mathcal{L}_k$ for the Gibbs measure with parameters given by the contact point (t_N,x_N,λ_N). This concentration will be deduced from the fact that the averaged free energy is being touched from above by a smooth function at the contact point, thereby constraining its Hessian at this point. The concentration of the free energy $F_N^{(K)}$ about its average $\overline{F}_N^{(K)}$ established in Proposition 6.3 will also play its part. Due to the constraint on the Hessian at the contact point, it will be possible to extend the concentration result on the free energy into an estimate on the concentration of its gradient. The argument is decomposed into a series of four lemmas. The first two essentially bound the Hessian of the perturbed free energy (6.35) from above and from below as in Steps 1 and 2 of the proof of Lemma 2.29. The third leverages the free energy concentration result in Proposition 6.3 to estimate the uniform L^p-distance between the quenched and averaged free energies (6.35) as in Lemma 2.28. Finally, the fourth extends this to a control on the gradient of the free energy as in Step 3 of the proof of Lemma 2.29.

Lemma 6.5. *For any perturbation parameter λ with $|\lambda|\leqslant C^{-1}$,*

$$\overline{F}_N^{(K)}(t_N,x_N,\lambda_N+\lambda)-\overline{F}_N^{(K)}(t_N,x_N,\lambda_N)-\lambda\cdot\nabla_\lambda\overline{F}_N^{(K)}(t_N,x_N,\lambda_N)\leqslant C|\lambda|^2. \tag{6.55}$$

Proof. Fix a perturbation parameter λ with $|\lambda|\leqslant C^{-1}$, and notice that (6.42) gives

$$\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N+\lambda)-\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N)\leqslant\widetilde{\phi}(t_N,x_N,\lambda_N+\lambda)-\widetilde{\phi}(t_N,x_N,\lambda_N).$$

On the other hand, Taylor's formula with integral remainder implies that

$$\begin{aligned}&\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N+\lambda)-\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N)\\&\qquad=\lambda\cdot\nabla_\lambda\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N)+\int_0^1(1-s)\lambda\cdot\nabla_\lambda^2\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N+s\lambda)\lambda\,\mathrm{d}s,\end{aligned} \tag{6.56}$$

and similarly,

$$\begin{aligned}&\widetilde{\phi}(t_N,x_N,\lambda_N+\lambda)-\widetilde{\phi}(t_N,x_N,\lambda_N)\\&\qquad=\lambda\cdot\nabla_\lambda\widetilde{\phi}(t_N,x_N,\lambda_N)+\int_0^1(1-s)\lambda\cdot\nabla_\lambda^2\widetilde{\phi}(t_N,x_N,\lambda_N+s\lambda)\lambda\,\mathrm{d}s.\end{aligned}$$

Combining (6.43) with the chain rule shows that $\lambda\cdot\nabla_\lambda\widetilde{\phi}(t_N,x_N,\lambda_N)=\lambda\cdot\nabla_\lambda\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N)$, and therefore

$$\int_0^1(1-s)\lambda\cdot\nabla_\lambda^2\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N+s\lambda)\lambda\,\mathrm{d}s\leqslant\int_0^1(1-s)\lambda\cdot\nabla_\lambda^2\widetilde{\phi}(t_N,x_N,\lambda_N+s\lambda)\lambda\,\mathrm{d}s\leqslant C|\lambda|^2.$$

Substituting this into (6.56) gives

$$\begin{aligned}\overline{F}_N^{(K)}(t_N,x_N,\lambda_N+\lambda)-\overline{F}_N^{(K)}(t_N,x_N,\lambda_N) &= \overline{F}_N'^{(K)}(t_N,x_N,\lambda_N+\lambda)-\overline{F}_N'^{(K)}(t_N,x_N,\lambda_N)\\ &\leqslant \lambda\cdot\nabla_\lambda \overline{F}_N'^{(K)}(t_N,x_N,\lambda_N)+C|\lambda|^2\\ &= \lambda\cdot\nabla_\lambda \overline{F}_N^{(K)}(t_N,x_N,\lambda_N)+C|\lambda|^2,\end{aligned}$$

and completes the proof. ∎

Lemma 6.6. *There exists a random variable X with $\mathbb{E}X^2\leqslant C$ such that, for all perturbation parameters λ with $|\lambda|\leqslant C^{-1}$,*

$$F_N^{(K)}(t_N,x_N,\lambda_N+\lambda)-F_N^{(K)}(t_N,x_N,\lambda_N)-\lambda\cdot\nabla_\lambda F_N^{(K)}(t_N,x_N,\lambda_N)\geqslant -X|\lambda|^2. \tag{6.57}$$

Proof. Since t_N and x_N remain fixed throughout, write $F_N^{(K)}(\lambda)$ for $F_N^{(K)}(t_N,x_N,\lambda)$. Introduce the function

$$h(\lambda):=F_N^{(K)}(\lambda)-\frac{\sqrt{\lambda_{0,N}}}{N}\sum_{i\leqslant N}|Z_{0,i}|+\frac{1}{N}\sum_{1\leqslant k\leqslant K_+}\sum_{j\leqslant \pi_k}\big(8\lambda_k^2 e_{jk}-\log(1-\lambda_k)\big). \tag{6.58}$$

Leveraging (6.51) and Hölder's inequality, one can see that

$$\partial_{\lambda_0}^2 h(\lambda)=\frac{1}{N}\big\langle(\mathcal{H}_0'-\langle\mathcal{H}_0'\rangle)^2\big\rangle-\frac{\varepsilon_N^2}{4N\lambda_{0,N}^{3/2}}\langle\sigma\rangle\cdot Z_0+\frac{\varepsilon_N^2}{4N\lambda_{0,N}^{3/2}}\sum_{i\leqslant N}|Z_{0,i}|\geqslant\frac{1}{N}\big\langle(\mathcal{H}_0'-\langle\mathcal{H}_0'\rangle)^2\big\rangle.$$

Using (6.52) and (6.54) reveals that for $1\leqslant k\leqslant K_+$,

$$\partial_{\lambda_k}^2 h(\lambda)=\frac{1}{N}\big\langle(\mathcal{H}_k'-\langle\mathcal{H}_k'\rangle)^2\big\rangle+\frac{1}{N}\Bigg\langle\sum_{j\leqslant\pi_k}\Big(-\frac{1}{(1+\lambda_k\sigma_{i_{jk}})^2}+2\frac{\sigma_{i_{jk}}\sigma_{i_{jk}}^* e_{jk}}{(1+\lambda_k\sigma_{i_{jk}}^*)^3}+16e_{jk}+\frac{1}{(1-\lambda_k)^2}\Big)\Bigg\rangle.$$

Since $\lambda_k\leqslant 1/2$ and all spin configuration coordinates are bounded by one, it is actually the case that

$$\partial_{\lambda_k}^2 h(\lambda)\geqslant\frac{1}{N}\big\langle(\mathcal{H}_k'-\langle\mathcal{H}_k'\rangle)^2\big\rangle.$$

Together with (6.53), this shows that $\nabla_\lambda^2 h$ is positive definite and therefore h is convex. It follows that for any perturbation parameter λ with $|\lambda|\leqslant C^{-1}$,

$$h(\lambda_N+\lambda)\geqslant h(\lambda_N)+\lambda\cdot\nabla_\lambda h(\lambda_N).$$

Remembering the definition of h in (6.58), this shows that the left side of (6.57) is bounded from below by

$$\begin{aligned}&\frac{1}{N}\Bigg(\sqrt{(\lambda_N)_{0,N}+\lambda_{0,N}}-\sqrt{(\lambda_N)_{0,N}}-\frac{\lambda_{0,N}}{2\sqrt{(\lambda_N)_{0,N}}}\Bigg)\sum_{i\leqslant N}|Z_{0,i}|\\ &\quad+\frac{1}{N}\sum_{1\leqslant k\leqslant K_+}\sum_{j\leqslant\pi_k}8e_{jk}\Big((\lambda_N)_k^2+2\lambda_k(\lambda_N)_k-\big((\lambda_N)_k+\lambda_k\big)^2\Big)\\ &\quad\quad+\frac{1}{N}\sum_{1\leqslant k\leqslant K_+}\sum_{j\leqslant\pi_k}\Bigg(\log\bigg(\frac{1-(\lambda_N)_k-\lambda_k}{1-(\lambda_N)_k}\bigg)+\frac{\lambda_k}{1-(\lambda_N)_k}\Bigg).\end{aligned}$$

Increasing C if necessary, Taylor's theorem with differential remainder gives a perturbation parameter $\widetilde{\lambda}$ with $\widetilde{\lambda}_k \in [2^{-k-1}, 2^{-k}]$ for $0 \leqslant k \leqslant K_+$ whose value might not be the same at each occurrence such that

$$\begin{aligned}\sqrt{(\lambda_N)_{0,N} + \lambda_{0,N}} - \sqrt{(\lambda_N)_{0,N}} - \frac{\lambda_{0,N}}{2\sqrt{(\lambda_N)_{0,N}}} &= -\frac{(\lambda_{0,N})^2}{8\widetilde{\lambda}_{0,N}^{3/2}} \geqslant -\sqrt{\varepsilon_N}\lambda_0^2 \geqslant -\lambda_0^2 \\ (\lambda_N)_k^2 + 2\lambda_k(\lambda_N)_k - \big((\lambda_N)_k + \lambda_k\big)^2 &= -\lambda_k^2 \\ \log\left(\frac{1-(\lambda_N)_k - \lambda_k}{1-(\lambda_N)_k}\right) + \frac{\lambda_k}{1-(\lambda_N)_k} &= -\frac{\lambda_k^2}{2(1-\widetilde{\lambda}_k)^2} \geqslant -2\lambda_k^2.\end{aligned}$$

It follows that the left side of (6.57) is bounded from below by

$$-\frac{\lambda_0^2}{N}\sum_{i\leqslant N}|Z_{0,i}| - \frac{1}{N}\sum_{1\leqslant k\leqslant K_+}\lambda_k^2\sum_{j\leqslant \pi_k}(8e_{jk}+2) \geqslant -X|\lambda|^2$$

for the random variable

$$X := \frac{1}{N}\sum_{i\leqslant N}|Z_{0,i}| + \frac{1}{N}\sum_{1\leqslant k\leqslant K_+}\sum_{j\leqslant \pi_k}(8e_{jk}+2).$$

Using the Cauchy-Schwarz inequality, taking the average with respect to the randomness of $(e_{jk})_{j,k\geqslant 1}$ before the average with respect to the randomness of $(\pi_k)_{k\geqslant 1}$, and remembering (5.10) shows that

$$\begin{aligned}\mathbb{E}X^2 &\leqslant \frac{C}{N^2}\left(\mathbb{E}\Big(\sum_{i\leqslant N}|Z_{0,i}|\Big)^2 + \sum_{1\leqslant k\leqslant K_+}\mathbb{E}\Big(\sum_{j\leqslant \pi_k}(8e_{jk}+2)\Big)^2\right) \\ &\leqslant \frac{C}{N^2}\Big(N\mathbb{E}|Z_{0,1}| + (N^2-N)\mathbb{E}|Z_{0,1}Z_{0,2}| + \sum_{1\leqslant k\leqslant K_+}\mathbb{E}\pi_k\sum_{j\leqslant \pi_k}(8e_{jk}+2)^2\Big) \\ &\leqslant \frac{C}{N^2}\big(N^2 + s_N^2 + s_N\big) \leqslant C.\end{aligned}$$

This completes the proof. ∎

Lemma 6.7. *For every $M > 0$ small enough, $p \in [1, +\infty)$, and $\varepsilon > 0$, there exists a constant $C < +\infty$ not depending on N such that*

$$\left(\mathbb{E}\sup_{\|\lambda\|_\infty \leqslant M}\left|\Big(F_N^{(K)} - \overline{F}_N^{(K)}\Big)(t_N, x_N, \lambda_N + \lambda)\right|^p\right)^{\frac{1}{p}} \leqslant CN^{-\frac{1}{2}+\varepsilon}. \tag{6.59}$$

Proof. Let $0 < M < 1/2$ be small enough so that $(\lambda_N + \lambda)_k \in [2^{-k-1}, 2^{-k}]$ for $0 \leqslant k \leqslant K_+$ whenever $\|\lambda\|_\infty \leqslant M$, and for each perturbation parameter λ introduce the random variable

$$Y(\lambda) := \frac{1}{N}\sum_{0\leqslant k\leqslant K_+}|\langle \mathcal{H}_k'\rangle|,$$

where the Gibbs average is associated with the perturbed Hamiltonian (6.11) evaluated at $(t_N, x_N, \lambda_N + \lambda)$. The relevance of these random variables stems from the fact that by the mean value theorem, (6.51), and (6.52), for every λ, λ' in the ℓ^∞-ball of radius M,

$$\left|F_N^{(K)}(t_N, x_N, \lambda_N + \lambda) - F_N^{(K)}(t_N, x_N, \lambda_N + \lambda')\right| \leqslant C\sup_{\|\eta\|_\infty \leqslant M} Y(\eta)\|\lambda - \lambda'\|_1.$$

Averaging this inequality also shows that for every λ, λ' in the ℓ^∞-ball of radius M,

$$\left|\overline{F}_N^{(K)}(t_N, x_N, \lambda_N + \lambda) - \overline{F}_N^{(K)}(t_N, x_N, \lambda_N + \lambda')\right| \leqslant \mathbb{E} \sup_{\|\eta\|_\infty \leqslant M} Y(\eta)\|\lambda - \lambda'\|_1$$

These two bounds imply that for any even integer $q \geqslant 2$,

$$\mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} \left|\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N + \lambda)\right|^q \leqslant \mathbb{E} \sup_{\lambda \in A_\varepsilon} \left|\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N + \lambda)\right|^q + C\varepsilon^q \mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} Y(\lambda)^q,$$

for the set $A_\varepsilon := \varepsilon \mathbb{Z}^{1+K_+} \cap \{\|\lambda\|_\infty \leqslant M\}$. Indeed, every λ is at most at distance $\varepsilon(K_+ + 1)$ from an element in A_ε with respect to the ℓ^1-norm. Bounding the supremum over A_ε by the sum over A_ε and invoking the free energy concentration result in Proposition 6.3 shows that

$$\mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} \left|\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N + \lambda)\right|^q \leqslant C|A_\varepsilon| N^{-\frac{q}{2}} + C\varepsilon^q \mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} |Y(\lambda)|^q. \tag{6.60}$$

To bound the moments of $\sup_{\|\lambda\|_\infty \leqslant M} |Y(\lambda)|$, fix $1 \leqslant k \leqslant K_+$, and observe that Hölder's inequality and (5.19) reveal that

$$\begin{aligned}\mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} |\langle \mathcal{H}_k' \rangle|^q &\leqslant \mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} \left|\sum_{j \leqslant \pi_k} \frac{1}{1 - (\lambda_N + \lambda)_k} + \frac{e_{jk}}{(1 - (\lambda_N + \lambda)_k)^2}\right|^q \\ &\leqslant \mathbb{E}\left|\sum_{j \leqslant \pi_k} \frac{4(1 + e_{jk})}{(1 - 2M)^2}\right|^q \leqslant \mathbb{E}\pi_k^{q-1} \sum_{j \leqslant \pi_k} \sum_{j \leqslant \pi_k} \left|\frac{4(1 + e_{jk})}{(1 - 2M)^2}\right|^q \\ &\leqslant C\mathbb{E}\pi_k^q,\end{aligned}$$

where the last inequality is found by averaging over the randomness of $(e_{jk})_{j,k \geqslant 1}$. Similarly (5.18) and Hölder's inequality give

$$\begin{aligned}\mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} |\langle \mathcal{H}_0' \rangle|^q &\leqslant \varepsilon_N^q \mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} \left||\sigma \cdot \sigma^*| + \frac{|\sigma \cdot Z_0|}{2\varepsilon_N^{q/2}((\lambda_N)_0 + \lambda_0)^{q/2}}\right|^q \leqslant C\varepsilon_N^q \left(N^q + \frac{2^q \mathbb{E}|\sigma \cdot Z_0|^q}{2\varepsilon_N^{q/2}(1 - 2M)^q}\right) \\ &\leqslant C\varepsilon_N^{\frac{q}{2}} N^q.\end{aligned}$$

Combining these two inequalities with the Poisson moment bound in Lemma A.14, and recalling the properties (5.9) and (5.10) of the sequences $(\varepsilon_N)_{N \geqslant 1}$ and $(s_N)_{N \geqslant 1}$ shows that

$$\mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} |Y(\lambda)|^q \leqslant C. \tag{6.61}$$

Substituting this into (6.60) and noticing that $|A_\varepsilon|$ is of order $\varepsilon^{-(K_+ + 1)}$ yields

$$\mathbb{E} \sup_{\|\lambda\|_\infty \leqslant M} \left|\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N + \lambda)\right|^q \leqslant C\left(\varepsilon^{-(K_+ + 1)} N^{-\frac{q}{2}} + \varepsilon^q\right).$$

Taking $1/q$'th powers and choosing $\varepsilon := N^{-\frac{q}{2(q+K_+ +1)}}$ gives

$$\left(\mathbb{E}\sup_{\|\lambda\|_\infty \leqslant M}\left|\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N + \lambda)\right|^q\right)^{\frac{1}{q}} \leqslant CN^{-\frac{q}{2(q+K_+ +1)}}.$$

Notice that the power on the right side can be made arbitrarily close to $-\frac{1}{2}$ by taking q large enough. Invoking Jensen's inequality completes the proof. ∎

Lemma 6.8. *For every $\varepsilon > 0$, there exists a constant $C < +\infty$ not depending on N such that*

$$\mathbb{E}\left|\nabla_\lambda\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)\right|^2 \leqslant CN^{-\frac{1}{2}+\varepsilon}. \tag{6.62}$$

Proof. Given $\mu \in [0, C^{-1}]$, consider the random perturbation parameter

$$\lambda := \mu \cdot \frac{\nabla_\lambda\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)}{\left|\nabla_\lambda\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)\right|}.$$

Combining Lemma 6.5 and 6.6 shows that

$$\begin{aligned}\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)&(t_N, x_N, \lambda_N + \lambda) - \left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)\\ &\geqslant \mu\left|\nabla_\lambda\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)\right| - (C+X)|\lambda|^2.\end{aligned}$$

Rearranging, squaring, and taking expectations yields

$$\mu^2\mathbb{E}\left|\nabla_\lambda\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)\right|^2 \leqslant C\left(\mathbb{E}\sup_{\|\lambda\|_\infty \leqslant C^{-1}}\left|\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N + \lambda)\right|^2 + \mu^4\right),$$

where it has been used that $\mathbb{E}X^2 \leqslant C$ and $|\lambda| = \mu$. Invoking Lemma 6.7 gives

$$\mathbb{E}\left|\nabla_\lambda\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)\right|^2 \leqslant C\left(\frac{1}{N^{1-2\varepsilon}\mu^2} + \mu^2\right).$$

Optimizing over μ leads to the choice $\mu := N^{-\frac{1}{4}+\frac{\varepsilon}{2}}$, and completes the proof. ∎

Lemma 6.9. *For any $1 \leqslant k \leqslant K_+$, there exists a constant $C < +\infty$ not depending on N such that*

$$\mathbb{E}\left\langle(\mathcal{L}_0 - \mathbb{E}\langle\mathcal{L}_0\rangle)^2\right\rangle \leqslant CN^{-\frac{1}{4}} \quad \textit{and} \quad \mathbb{E}\left\langle(\mathcal{L}_k - \mathbb{E}\langle\mathcal{L}_k\rangle)^2\right\rangle \leqslant CN^{-\frac{1}{20}}. \tag{6.63}$$

Here, the Gibbs average $\langle\cdot\rangle$ is associated with the perturbed Hamiltonian (6.11) evaluated at the contact point (t_N, x_N, λ_N).

Proof. A direct computation using (6.51) shows that

$$\begin{aligned}N^2\varepsilon_N^2\mathbb{E}\left\langle(\mathcal{L}_0 - \mathbb{E}\langle\mathcal{L}_0\rangle)^2\right\rangle &= \mathbb{E}\left\langle(\mathcal{H}_0' - \langle\mathcal{H}_0'\rangle)^2\right\rangle + \mathbb{E}\left(\langle\mathcal{H}_0'\rangle - \mathbb{E}\langle\mathcal{H}_0'\rangle\right)^2\\ &= N\partial_{\lambda_0}^2\overline{F}_N^{(K)}(t_N, x_N, \lambda_N) + \frac{\varepsilon_N^2}{4\lambda_{0,N}^{3/2}}\mathbb{E}\langle\sigma\rangle\cdot Z_0 + N^2\mathbb{E}\left(\partial_{\lambda_0}\left(F_N^{(K)} - \overline{F}_N^{(K)}\right)(t_N, x_N, \lambda_N)\right)^2.\end{aligned}$$

It follows by Lemmas 6.5 and 6.8 that for any $\varepsilon > 0$,

$$\mathbb{E}\big\langle (\mathcal{L}_0 - \mathbb{E}\langle \mathcal{L}_0\rangle)^2\big\rangle \leqslant \frac{C}{N^2\varepsilon_N^2}\Big(N + N\varepsilon_N^{2-\frac{3}{2}} + N^{2-\frac{1}{2}+\varepsilon}\Big) = C\Big(N^{2|\gamma|-1} + N^{\frac{3}{2}|\gamma|-1} + N^{2|\gamma|+\varepsilon-\frac{1}{2}}\Big).$$

Remembering that $-1/8 < \gamma < 0$ gives the first bound in (6.63). To establish the second bound, fix $1 \leqslant k \leqslant K_+$. A direct computation using (6.52) yields

$$\begin{aligned} s_N^2\mathbb{E}\big\langle (\mathcal{L}_k - \mathbb{E}\langle \mathcal{L}_k\rangle)^2\big\rangle &= \mathbb{E}\big\langle (\mathcal{H}_k' - \langle \mathcal{H}_k'\rangle)^2\big\rangle + \mathbb{E}\big(\langle \mathcal{H}_k'\rangle - \mathbb{E}\langle \mathcal{H}_k'\rangle\big)^2 \\ &= N\partial_{\lambda_k}^2 \overline{F}_N^{(K)}(t_N, x_N, \lambda_N) - \mathbb{E}\langle \mathcal{H}_k''\rangle + N^2\mathbb{E}\Big(\partial_{\lambda_k}\Big(F_N^{(K)} - \overline{F}_N^{(K)}\Big)(t_N, x_N, \lambda_N)\Big)^2. \end{aligned}$$

Invoking (6.54), and Lemmas 6.5 and 6.8 reveals that for any $\varepsilon > 0$,

$$\mathbb{E}\big\langle (\mathcal{L}_k - \mathbb{E}\langle \mathcal{L}_k\rangle)^2\big\rangle \leqslant \frac{C}{s_N^2}\Big(N + s_N + N^{\frac{3}{2}+\varepsilon}\Big) = C\Big(N^{1-2\eta} + N^{-\eta} + N^{\frac{3}{2}+\varepsilon-2\eta}\Big).$$

Choosing $\varepsilon := 1/20$, and recalling that $-1/8 < \gamma < 0$ and $4/5 < \eta < 1$ completes the proof. ■

This result implies the fundamental assumption (5.20) in Chapter 5. Combining this with Proposition 5.2 and Corollary 5.3, and fixing $\varepsilon > 0$, it is possible to find $\delta > 0$ so the statement of Proposition 5.4 holds. In particular, setting $K_+ := \lfloor \delta^{-1}\rfloor$ in the perturbed Hamiltonian (6.11) ensures that for $1 \leqslant m \leqslant \lfloor \varepsilon^{-1}\rfloor$,

$$\mathbb{E}\big\langle (R_{[m]} - \mathbb{E}\langle R_{[m]}\rangle)^2\big\rangle \leqslant \varepsilon. \tag{6.64}$$

This multioverlap concentration can be combined with the computations in Chapter 3 to finally give a proof of Lemma 6.4.

Proof of Lemma 6.4. To alleviate notation, it will always be implicitly assumed that $\overline{F}_N^{(K)}$ and its derivatives are evaluated at the contact point (t_N, x_N, λ_N). The definition of the modified free energy in (6.36), and Corollary 3.3 reveal that

$$\partial_t \overline{F}_N'^{(K)} = \frac{1}{2}\big(c + \Delta\overline{m}^2\big)\log(c) + \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\big\langle R_{[n]}^2\big\rangle - \frac{c}{2} + \frac{b}{2} + \mathcal{O}(N^{-1}). \tag{6.65}$$

On the other hand, the duality relation (4.12), the definition of the modified free energy in (6.36), and Corollary 3.6 imply that for any $k \in \mathcal{D}_K$,

$$\partial_{x_k}\overline{F}_N'^{(K)} = \frac{1}{|\mathcal{D}_K|}\bigg(\big(c + \Delta\overline{m}k\big)\log(c) + c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\langle R_{[n]}\rangle k^n - c + b\bigg) + \mathcal{O}(N^{-1}).$$

Denoting by $\mu^* := \mathcal{L}(\langle \sigma_i\rangle)$ the law of the Gibbs average of a uniformly sampled spin coordinate, the Nishimori identity (Proposition 2.2) and the definition of $\widetilde{g}_b$ in (6.2) imply that this may be rewritten as

$$\partial_{x_k}\overline{F}_N'^{(K)} = \frac{1}{|\mathcal{D}_K|}\int_{-1}^1 \widetilde{g}_b(ky)\,\mathrm{d}\mu^*(y) + \mathcal{O}\big(N^{-1}\big).$$

The mean value theorem shows that

$$|\mathcal{D}_K|\Big|\frac{1}{|\mathcal{D}_K|}\int_{-1}^{1}\widetilde{g}_b(ky)\,\mathrm{d}\mu^*(y)-\widetilde{G}_b^{(K)}x^{(K)}(\mu^*)_k\Big|\leqslant\sum_{k'\in\mathcal{D}_K}\int_{k'}^{k'+2^{-K}}|\widetilde{g}_b(ky)-\widetilde{g}_b(kk')|\,\mathrm{d}\mu^*(y)\leqslant\frac{\|\widetilde{g}_b'\|_{L^\infty}}{2^K},$$

which means that

$$|||\nabla_x\overline{F}_N'^{(K)}-\widetilde{G}_b^{(K)}x^{(K)}(\mu^*)|||_{1,*}\leqslant\frac{\|\widetilde{g}_b'\|_{L^\infty}}{2^K}+\mathcal{O}\big(N^{-1}\big).$$

In particular, for N large enough,

$$|||\widetilde{G}_b^{(K)}x^{(K)}(\mu^*)|||_{1,*}\leqslant|||\nabla_x\overline{F}_N'^{(K)}|||_{1,*}+\frac{\|\widetilde{g}_b'\|_{L^\infty}}{2^K}+\mathcal{O}\big(N^{-1}\big)\leqslant\|\widetilde{g}_b\|_{L^\infty}+\|\widetilde{g}_b'\|_{L^\infty}+\mathcal{O}\big(N^{-1}\big)\leqslant R$$

Remembering that $\widetilde{\mathsf{H}}_{b,K,R}$ coincides with $\widetilde{\mathsf{C}}_{b,K}$ on $\widetilde{\mathcal{C}}_{b,K}\cap B_{K,R}$, and leveraging the Lipschitz continuity of $\widetilde{\mathsf{H}}_{b,K,R}$ in Proposition 4.8 gives

$$\Big|\widetilde{\mathsf{H}}_{b,K,R}\Big(\nabla_x\overline{F}_N'^{(K)}\Big)-\widetilde{\mathsf{C}}_{b,K}\big(\widetilde{G}_b^{(K)}x^{(K)}(\mu^*)\big)\Big|\leqslant\frac{8RM_b\|\widetilde{g}_b'\|_{L^\infty}}{2^Km_b^2}+\mathcal{O}\big(N^{-1}\big),$$

where $M_b:=\max_{[-1,1]}\widetilde{g}_b$ and $m_b:=\min_{[-1,1]}\widetilde{g}_b>0$. Another application of the mean value theorem shows that

$$\Big|\widetilde{\mathsf{C}}_{b,K}\big(\widetilde{G}_b^{(K)}x^{(K)}(\mu^*)\big)-\frac{1}{2}\int_{-1}^{1}\int_{-1}^{1}\widetilde{g}_b(xy)\,\mathrm{d}\mu^*(y)\,\mathrm{d}\mu^*(x)\Big|\leqslant\frac{\|\widetilde{g}_b'\|_{L^\infty}}{2^K},$$

while a direct computation using the Nishimori identity reveals that

$$\frac{1}{2}\int_{-1}^{1}\widetilde{g}_b(xy)\,\mathrm{d}\mu^*(y)\,\mathrm{d}\mu^*(x)=\frac{1}{2}\big(c+\Delta\overline{m}^2\big)\log(c)+\frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\big(\mathbb{E}\langle R_{[n]}\rangle\big)^2-\frac{c}{2}+\frac{b}{2}.$$

It follows by (6.65) that, up to an error vanishing with N,

$$\Big|\partial_t\overline{F}_N'^{(K)}-\widetilde{\mathsf{H}}_{K,R}\Big(\nabla_x\overline{F}_N'^{(K)}\Big)\Big|\leqslant\Big|\frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\big\langle(R_{[n]}-\mathbb{E}\langle R_{[n]}\rangle)^2\big\rangle\Big|+\frac{8RM_b\|\widetilde{g}_b'\|_{L^\infty}}{2^Km_b^2}+\frac{\|\widetilde{g}'\|_{L^\infty}}{2^K}.$$

Invoking the multioverlap concentration (6.64), noticing that the multioverlaps are bounded by one, and using the formula for the sum of a geometric series implies that, up to an error vanishing with N,

$$\Big|\partial_t\overline{F}_N'^{(K)}-\widetilde{\mathsf{H}}_{K,R}\Big(\nabla_x\overline{F}_N'^{(K)}\Big)\Big|\leqslant\frac{\varepsilon c^2}{2(c-|\Delta|)}+\frac{c}{2}\sum_{n\geqslant\lfloor\varepsilon^{-1}\rfloor}\big(|\Delta|/c\big)^n+\frac{8RM_b\|\widetilde{g}_b'\|_{L^\infty}}{2^Km_b^2}+\frac{\|\widetilde{g}'\|_{L^\infty}}{2^K}.$$

Defining $\mathcal{E}_{\varepsilon,K}$ to be the right side of this expression completes the proof. ■

In the disassortative sparse stochastic block model, it turns out that the non-linearity in the Hamilton-Jacobi equation (3.91) is convex, so the upper bound (6.49) on the limit of the free energy established in the proof of Theorem 1.6 may be combined with the infinite-dimensional Hopf-Lax formula in Theorem 4.5 to obtain the variational formula for the limit mutual information stated in Theorem 1.2.

6.4 Deducing the disassortative free energy variational formula

The limit of the mutual information in the disassortative sparse stochastic block model, $\Delta \leqslant 0$, is well-understood [14, 38], and admits the variational formula stated in Theorem 1.2. The key difference between the disassortative and assortative sparse stochastic block models that makes the former much simpler to study than the latter is essentially the presence of convexity. Indeed, in the disassortative sparse stochastic block model, the non-linearity in the infinite-dimensional Hamilton-Jacobi equation (3.91) is convex. More precisely, for b large enough, the kernel $\widetilde{g}_b$ in (6.2) is convex in the sense of (**H5**). This allows one to invoke the infinite-dimensional Hopf-Lax formula in Theorem 4.5 to transform the upper bound (6.49) on the limit of the free energy into a variational upper bound that matches the one in Theorem 1.2. A simple interpolation argument taken from [14] can then be used to obtain the matching lower bound. Remembering the relationship (3.17) between the free energy and the mutual information leads to a proof of Theorem 1.7.

Lemma 6.10. *If $\Delta \leqslant 0$ and b is large enough, then the function $\widetilde{g}_b : [-1,1] \to \mathbb{R}$ defined in* (6.2) *satisfies* (**H5**).

Proof. By a simple approximation argument, it suffices to establish (4.37) for a discrete signed measure of the form

$$\mu := \frac{1}{|\mathcal{D}_K|} \sum_{k \in \mathcal{D}_K} x_k \delta_k$$

for some $x \in \mathbb{R}^{\mathcal{D}_K}$. For such a measure,

$$\int_{-1}^{1} \int_{-1}^{1} \widetilde{g}_b(xy)\, \mathrm{d}\mu(x)\, \mathrm{d}\mu(y) = \frac{1}{|\mathcal{D}_K|^2} \sum_{k,k' \in \mathcal{D}_K} \widetilde{g}_b(kk') x_k x_{k'} = x^T \widetilde{G}_b^{(K)} x,$$

so (4.37) is equivalent to the non-negative definiteness of each of the matrices $\widetilde{G}_b^{(K)}$. Observe that for any $k, k' \in \mathcal{D}_K$,

$$\big(\widetilde{G}_b^{(K)}\big)_{kk'} = \frac{1}{|\mathcal{D}_K|^2} \widetilde{g}_b(kk') = \frac{1}{|\mathcal{D}_K|^2} \bigg(b + c\log(c) - c + \Delta kk' \log(c) + c \sum_{n \geqslant 2} \frac{(-\Delta/c)^n}{n(n-1)} (kk')^n \bigg).$$

With this in mind, introduce the vectors $\mathbf{k} := (k)_{k \in \mathcal{D}_K}$ and $\imath := (1)_{k \in \mathcal{D}_K}$ as well as the matrix

$$\widetilde{G}_{b,M}^{(K)} := \frac{1}{|\mathcal{D}_K|^2} \bigg((b + c\log(c) - c)\imath\imath^T + \Delta \log(c) \mathbf{k}\mathbf{k}^T + \sum_{2 \leqslant n \leqslant M} \frac{(-\Delta/c)^n}{n(n-1)} \big(\mathbf{k}\mathbf{k}^T\big)^{\odot n} \bigg),$$

where $\odot n$ denotes the n-fold Hadamard product on the space of $\mathcal{D}_K \times \mathcal{D}_K$ matrices. In this notation,

$$\widetilde{G}_b^{(K)} = \frac{1}{|\mathcal{D}_K|^2} \bigg((b + c\log(c) - c)\imath\imath^T + \Delta \log(c) \mathbf{k}\mathbf{k}^T + \sum_{n \geqslant 2} \frac{(-\Delta/c)^n}{n(n-1)} \big(\mathbf{k}\mathbf{k}^T\big)^{\odot n} \bigg) = \lim_{M \to +\infty} \widetilde{G}_{b,M}^{(K)}.$$

Choosing $b > 2c|\log(c)| + c$ ensures that the first two matrices in the sum defining $\widetilde{G}_{b,M}^{(K)}$ are non-negative definite. Using that $\Delta \leqslant 0$ and leveraging the Schur product theorem (Theorem 5.2.1 in [64]) reveals that the matrix $\widetilde{G}_{b,M}^{(K)}$ is a positive linear combination of non-negative definite matrices, and is therefore non-negative definite. Noticing that the limit of non-negative definite matrices is again non-negative definite completes the proof. ∎

Proof of Theorem 1.7. Introduce the functional $\mathcal{P}:\mathbb{R}_{>0}\times\mathcal{M}_+\times\mathcal{M}_+\to\mathbb{R}$ defined by

$$\mathcal{P}(t,\mu,\nu):=\psi(\mu+t\nu)-\frac{t}{2}\int_{-1}^{1}G_\nu(y)\,\mathrm{d}\nu(y). \tag{6.66}$$

Combining Lemma 6.10 with the infinite-dimensional Hopf-Lax formula in Theorem 4.5 reveals that the unique solution to the infinite-dimensional Hamilton-Jacobi equation (3.91) evaluated at the pair $(t,\mu)\in\mathbb{R}_{\geqslant 0}\times\mathcal{M}_+$ admits the variational representation

$$f(t,\mu)=\sup_{\nu\in\Pr[-1,1]}\mathcal{P}(t,\mu,\nu). \tag{6.67}$$

Together with the upper bound (6.49) on the limit of the free energy obtained in the proof of Theorem 1.6, this implies that

$$\limsup_{N\to+\infty}\overline{F}_N\leqslant f(1,0)=\sup_{\nu\in\Pr[-1,1]}\mathcal{P}(1,0,\nu). \tag{6.68}$$

It has been used that the free energy (3.14) in the sparse stochastic block model is given by $\overline{F}_N=\overline{F}_N(1,0)$, where 0 denotes the zero measure. The proof now proceeds in three steps. First, the supremum on the right side of (6.68) is restricted to probability measures $\nu\in\mathcal{M}_p$ with mean $\overline{m}$, then the functional (6.66) is replaced by the functional $\mathcal{P}:\mathcal{M}_p\to\mathbb{R}$ defined in (1.18), and finally the matching lower bound is established.

Step 1: restricting to measures $\nu\in\mathcal{M}_p$. Fix $b>1$ large enough so the kernel $\widetilde{g}_b$ defined in (6.2) is strictly positive on $[-1,1]$. In the same spirit as (4.46) and (4.50), for each integer $N\geqslant 1$, define the function $\widetilde{G}_{b,\nu}:[-1,1]\to\mathbb{R}$ and the initial condition $\widetilde{\psi}_{b,N}:\mathcal{M}_+\to\mathbb{R}$ by

$$\widetilde{G}_{b,\nu}(x):=\int_{-1}^{1}\widetilde{g}_b(xy)\,\mathrm{d}\nu(y)\quad\text{and}\quad\widetilde{\psi}_{b,N}(\mu):=\psi_N(\mu)+b\int_{-1}^{1}\mathrm{d}\mu,$$

and introduce the functional $\widetilde{\mathcal{P}}_{b,N}:\mathbb{R}_{\geqslant 0}\times\mathcal{M}_+\times\mathcal{M}_+\to\mathbb{R}$ defined by

$$\widetilde{\mathcal{P}}_{b,N}(t,\mu,\nu):=\widetilde{\psi}_{b,N}(\mu+t\nu)-\frac{t}{2}\int_{-1}^{1}\widetilde{G}_{b,\nu}(y)\,\mathrm{d}\nu(y).$$

Invoking Theorem 4.5 gives a probability measure $\widetilde{\nu}\in\Pr[-1,1]$ which maximizes the right side of (6.68). It follows by Proposition 3.12 that

$$\sup_{\nu\in\Pr[-1,1]}\mathcal{P}(1,0,\nu)=\lim_{N\to+\infty}\widetilde{\mathcal{P}}_{b,N}(1,0,\widetilde{\nu})-\frac{b}{2}\leqslant\limsup_{N\to+\infty}\sup_{\nu\in\mathcal{M}_+}\widetilde{\mathcal{P}}_{b,N}(1,0,\nu)-\frac{b}{2}. \tag{6.69}$$

An identical argument to that in Lemma 4.22 gives a sequence of maximizing measures $(\nu_N)_{n\geqslant 1}\subseteq\mathcal{M}_+$ with

$$\sup_{\nu\in\mathcal{M}_+}\widetilde{\mathcal{P}}_{b,N}(1,0,\nu)=\widetilde{\mathcal{P}}_{b,N}(1,0,\nu_N). \tag{6.70}$$

By Corollary 3.6 and the approximate equality (3.73), the Gateaux derivative density of the initial condition $\widetilde{\psi}_{b,N}$ at the measure ν_N is given by $D_\mu\widetilde{\psi}_{b,N}(\nu_N,\cdot)=\widetilde{G}_{b,\nu_N^*}+\mathcal{O}\left(N^{-1}\right)$ for some measure $\nu_N^*\in\mathcal{M}_p$. Up to adding errors of $\mathcal{O}(N^{-1})$ throughout, the proof of Theorem 4.3 applies and reveals that each maximizer $\nu_N\in\mathcal{M}_+$ satisfies the approximate first-order condition

$$D_\mu\widetilde{\psi}_{b,N}(\nu_N,\cdot)=\widetilde{G}_{b,\nu_N}(\cdot)+\mathcal{O}\left(N^{-1}\right).$$

This means that

$$\widetilde{G}_{b,\nu_N^*}(\cdot) = D_\mu \widetilde{\psi}_{b,N}(\nu_N,\cdot) + \mathcal{O}\big(N^{-1}\big) = \widetilde{G}_{b,\nu_N}(\cdot) + \mathcal{O}\big(N^{-1}\big).$$

Together with the definition of $\widetilde{g}_b$ in (6.2), this implies that

$$\begin{aligned}(b+c\log(c)-c)\int_{-1}^1 \mathrm{d}\nu_N(y) + \Delta\log(c)\int_{-1}^1 y\,\mathrm{d}\nu_N(y)x + c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\int_{-1}^1 y^n\,\mathrm{d}\nu_N(y)x^n \\ = (b+c\log(c)-c) + \Delta\log(c)\overline{m}x + c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\int_{-1}^1 y^n\,\mathrm{d}\nu_N^*(y)x^n + \mathcal{O}\big(N^{-1}\big).\end{aligned}$$

Since $b>1$ and $c\log(c)-c\geqslant -1$, it must be that for all $n\geqslant 2$,

$$\int_{-1}^1 \mathrm{d}\nu_N(y) = 1+\mathcal{O}\big(N^{-1}\big) \quad \text{and} \quad \int_{-1}^1 y^n\,\mathrm{d}\nu_N(y) = \int_{-1}^1 y^n\,\mathrm{d}\nu_N^*(y) + \mathcal{O}\big(N^{-1}\big).$$

It has been used that without loss of generality $\Delta\neq 0$, since this case is trivial as it corresponds to the situation where the graph $\mathbf{G}_N$ and the assignment vector σ^* are independent. The possibility that c could be equal to one has been accounted for. Applying the Prokhorov theorem (Theorem A.20 in [50]), and passing to a subsequence if necessary, it is therefore possible to ensure that the sequences $(\nu_N)_{n\geqslant 1}$ and $(\nu_N^*)_{n\geqslant 1}$ converge weakly to probability measures $\nu\in\Pr[-1,1]$ and $\nu^*\in\mathcal{M}_p$ such that for all $n\neq 1$,

$$\int_{-1}^1 y^n\,\mathrm{d}\nu(y) = \int_{-1}^1 y^n\,\mathrm{d}\nu^*(y).$$

Since the set of polynomials with degree one coefficient equal to zero form of a sub-algebra of the space of continuous functions on the compact set $[-1,1]$, the Stone-Weierstrass theorem (Theorem A.10 in [50]) implies that $\nu=\nu^*\in\mathcal{M}_p$. Arguing as in the proof of Lemma 6.1, it is readily verified that there exists a constant $C<+\infty$ that depends only on c with

$$\begin{aligned}\big|\widetilde{\psi}_{b,N}(\nu_N)-\widetilde{\psi}_b(\nu^*)\big| &\leqslant \big|\widetilde{\psi}_{b,N}(\nu_N)-\widetilde{\psi}_{b,N}(\overline{\nu}_N)\big| + \big|\widetilde{\psi}_{b,N}(\overline{\nu}_N)-\widetilde{\psi}_{b,N}(\nu^*)\big| + \big|\widetilde{\psi}_{b,N}(\nu^*)-\widetilde{\psi}_b(\nu^*)\big| \\ &\leqslant C\mathrm{TV}(\nu_N,\overline{\nu}_N) + CW(\overline{\nu}_N,\nu^*) + \big|\widetilde{\psi}_{b,N}(\nu^*)-\widetilde{\psi}_b(\nu^*)\big|.\end{aligned}$$

Recalling that the Wasserstein distance (3.85) metrizes the weak convergence of probability measures on $[-1,1]$, observing that $\mathrm{TV}(\nu_N,\overline{\nu}_N)=\big|1-\nu_N[-1,1]\big|$, and using Proposition 3.12 as well as the equality (6.70) to let N tend to infinity in (6.69) shows that

$$\sup_{\nu\in\Pr[-1,1]}\mathcal{P}(1,0,\nu) \leqslant \limsup_{N\to+\infty}\widetilde{\mathcal{P}}_{b,N}(1,0,\nu_N)-\frac{b}{2} = \widetilde{\psi}_b(\nu^*)-\frac{1}{2}\int_{-1}^1 \widetilde{G}_{b,\nu^*}(y)\,\mathrm{d}\nu^*(y)-\frac{b}{2} \leqslant \sup_{\nu\in\mathcal{M}_p}\mathcal{P}(1,0,\nu).$$

Substituting this upper bound into (6.68) gives the upper bound (6.68) with the set of probability measures $\Pr[-1,1]$ replaced by the set of probability measures $\mathcal{M}_p$ with mean $\overline{m}$.

Step 2: replacing the functional (6.66) *by the functional* (1.18). Fix $\nu\in\mathcal{M}_p$, and denote by x_1 and x_2 two independent samples from the probability measure ν. The definition of g in (3.68) implies that

$$\mathcal{P}(1,0,\nu) = \psi(\nu) + \frac{c}{2} - \frac{1}{2}\big(c+\Delta\overline{m}^2\big)\log(c) - \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\big(\mathbb{E}x_1^n\big)^2.$$

A Taylor expansion of the logarithm shows that

$$c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\big(\mathbb{E}x^n\big)^2 = \mathbb{E}(c+\Delta x_1x_2)\log(c+\Delta x_1x_2) - \big(c+\Delta\overline{m}^2\big)\log(c) - \Delta\overline{m}^2. \tag{6.71}$$

It follows that

$$\mathcal{P}(1,0,\nu) = \psi(\nu) + \frac{c}{2} - \frac{1}{2}\mathbb{E}(c+\Delta x_1x_2)\log(c+\Delta x_1x_2) + \frac{\Delta\overline{m}^2}{2} = \mathcal{P}(\nu),$$

where the functional $\mathcal{P}$ on the right side is defined in (1.18). Together with the previous step, this gives the upper bound

$$\limsup_{N\to+\infty}\overline{F}_N \leqslant f(1,0) \leqslant \sup_{\nu\in\mathcal{M}_p}\mathcal{P}(\nu).$$

Step 3: establishing the matching lower bound. Given a measure $\nu\in\mathcal{M}_p$, introduce the interpolating free energy $\varphi(t) := \widetilde{F}_N(t,1-t,\nu)$ for the free energy $\widetilde{F}_N$ defined in (3.24). The derivative computations in Corollary 3.3 and Lemma 3.4 together with a computation identical to that in Corollary 3.6 imply that

$$\begin{aligned}
\varphi'(t) &= \partial_t\widetilde{F}_N(t,1-t,\nu) - \partial_s\widetilde{F}_N(t,1-t,\nu)\\
&= \frac{c}{2} - \frac{1}{2}\big(c+\Delta\overline{m}^2\big)\log(c) + \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\big\langle R_{[n]}^2\big\rangle - c\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\langle R_{[n]}\rangle\mathbb{E}x_1^n.\\
&= \frac{c}{2} - \frac{1}{2}\big(c+\Delta\overline{m}^2\big)\log(c) - \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\big(\mathbb{E}x_1^n\big)^2 + \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\big\langle (R_{[n]}-\mathbb{E}x_1^n)^2\big\rangle.
\end{aligned}$$

It follows by (6.71) that

$$\varphi'(t) = \frac{c}{2} + \frac{\Delta\overline{m}^2}{2} - \frac{1}{2}\mathbb{E}(c+\Delta x_1x_2)\log(c+\Delta x_1x_2) + \frac{c}{2}\sum_{n\geqslant 2}\frac{(-\Delta/c)^n}{n(n-1)}\mathbb{E}\big\langle (R_{[n]}-\mathbb{E}x_1^n)^2\big\rangle.$$

Since the final term in this equality is non-negative, the fundamental theorem of calculus reveals that

$$\overline{F}_N \geqslant \psi_N(\nu) + \frac{c}{2} + \frac{\Delta\overline{m}^2}{2} - \frac{1}{2}\mathbb{E}(c+\Delta x_1x_2)\log(c+\Delta x_1x_2),$$

where it has been used that $\varphi(1) = \overline{F}_N$ and $\varphi(0) = \psi_N(\nu)$. Using Proposition 3.12 to let N tend to infinity gives the lower bound

$$\liminf_{N\to+\infty}\overline{F}_N \geqslant \psi(\nu) + \frac{c}{2} + \frac{\Delta\overline{m}^2}{2} - \frac{1}{2}\mathbb{E}(c+\Delta x_1x_2)\log(c+\Delta x_1x_2) = \mathcal{P}(\nu).$$

Taking the supremum over all measures $\nu\in\mathcal{M}_p$, and remembering the relationship (3.17) between the free energy and the mutual information completes the proof. ■

6.5 Relation to other works and future perspectives

This thesis closes with a brief discussion on the relation between the Hamilton-Jacobi approach and other recent approaches to determining the limit mutual information in the sparse stochastic block model. Future perspectives on Conjecture 1.4, and possible alternatives to this conjecture, are also touched upon. It will facilitate the discussion to point out that the proof of Theorem 1.7 also yields that when $\Delta \le 0$, the limit of the free energy $\overline{F}_N(t,\mu)$ can be identified, for every $t \ge 0$ and $\mu \in \mathcal{M}_+$, as

$$\lim_{N\to+\infty} \overline{F}_N(t,\mu) = f(t,\mu) = \sup_{\nu\in\mathrm{Pr}[-1,1]} \left(\psi(\mu+t\nu) - \frac{t}{2}\int_{-1}^{1} G_\nu(y)\,\mathrm{d}\nu(y) \right), \tag{6.72}$$

and that this identity also remains valid if the supremum is taken over all $\nu \in \mathcal{M}_+$. This alternative representation is at times convenient to operate with variables that can vary freely inside a cone.

The series of recent works [4, 61, 66, 79, 82, 107] described in Section 1.2.2 share strong connections with the Hamilton-Jacobi approach discussed in this thesis. Recall from Remark 3.8 that it is possible to identify a mapping $\Gamma : \mathcal{M}_+ \to \mathrm{Pr}[-1,1]$ such that for every $\mu \in \mathcal{M}_+$, one has $D_\mu \psi(\mu,\cdot) = G_{\Gamma(\mu)}$. This mapping is closely related to the BP operator described in Section 1.2.2. Let σ^* be sampled according to P^*, and conditionally on σ^*, let $\Pi(\mu)$ denote a Poisson point process with intensity measure $(c+\Delta\sigma^* x)\,\mathrm{d}\mu(x)$. Then the probability measure $\Gamma(\mu)$ is defined to be the law of the random variable

$$\frac{\int_{\Sigma_1} \sigma \exp(-\Delta\sigma \int_{-1}^{1} x\,\mathrm{d}\mu) \prod_{x\in\Pi(\mu)} (c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)}{\int_{\Sigma_1} \exp(-\Delta\sigma \int_{-1}^{1} x\,\mathrm{d}\mu) \prod_{x\in\Pi(\mu)} (c+\Delta\sigma x)\,\mathrm{d}P^*(\sigma)}. \tag{6.73}$$

Notice that the condition for the measure ν to be a critical point in the variational problem on the right side of (6.72) can be written as

$$G_\nu = D_\mu \psi(\mu+t\nu,\cdot). \tag{6.74}$$

At least when $\Delta < 0$, the mapping $\nu \mapsto G_\nu$ is injective, so the relation (6.74) can be equivalently written as

$$\nu = \Gamma(\mu+t\nu). \tag{6.75}$$

Restricting to the case of $(t,\mu) = (1,0)$, this boils down to finding fixed points of the mapping Γ. That there is a connection between the variational formula in Theorem 1.7 and some BP fixed point equation has already been observed in [38, 44] and elsewhere. The less classical question is to relate this to the Hamilton-Jacobi equation (3.91) for arbitrary Δ. In finite dimensions, Hamilton-Jacobi equations can be solved for short times using the method of characteristics — see Section 3.5 in [50] for a detailed discussion of this. Moreover, the slope of the characteristic line is computed by evaluating the gradient of the non-linearity at the gradient of the initial condition. In the present context, the characteristic line emanating from a measure $\nu \in \mathcal{M}_+$ is the trajectory

$$t' \mapsto (t', \nu - t'\Gamma(\nu)), \tag{6.76}$$

for t' varying in $\mathbb{R}_{\geqslant 0}$. As long as characteristic lines emanating from different choices of ν do not intersect each other, the value of the solution along each characteristic line can be calculated using the equation and the fact that the gradient of the solution remains constant along each line. The condition (6.75) turns out to be equivalent to asking that the characteristic line emanating from $\mu+t\nu$ passes through the point (t,μ), since the latter condition can be written as $\mu = \mu+t\nu - t\Gamma(\mu+t\nu)$. In other words, for each fixed (t,μ), there is a simple one-to-one correspondence between the fixed points to (6.75) and the characteristic lines that pass through (t,μ). The formula for prescribing the value of the solution along a characteristic line starting from $\mu+t\nu$ is the supremum (6.72). As long as t is sufficiently small that the equation (6.75) has a unique solution for each μ, this gives a clear procedure for computing the solution to (3.91). Once characteristic lines start to intersect, the viscosity solution to (3.91) aggregates these conflicting trajectories in a physically reasonable way, and Conjecture 1.4 corresponds to the idea that the free energy $\overline{F}_N$ is tracking this aggregation in the limit of large N. As such, the Hamilton-Jacobi approach may be a way to circumvent the difficulties that arise in [4, 61, 66, 79, 82, 107] when the BP operator admits multiple fixed points by, in a certain way, selecting the right fixed point.

It is of course not clear that the right fixed point in the BP operator is selected, since only one bound in Conjecture 1.4 has been established. Although the author believes this to be the right conjecture, some alternatives were considered before landing upon it. In particular, the author's original hope was that a variational formula could be found to describe the limit mutual information in the sparse stochastic block model. More precisely, the author hoped that the variational formula (6.72) would remain valid in the case of general Δ. It seems difficult to identify the exact range of validity of this formula; however, the author would be surprised if it holds for arbitrary measures P^*. The author is also fairly convinced that this formula will not generalize to settings with more than two communities.

To see this, return momentarily to the problem of identifying the limit of the free energy in the dense stochastic block model studied in Chapter 2. In this setting, central-limit-theorem effects take place, and one can equivalently study the symmetric rank-one matrix estimation problem, a fully-connected model with Gaussian noise. Such models have been studied extensively [12, 13, 16, 28, 31, 35, 36, 65, 68, 70, 72, 73, 75, 77, 84, 85, 98, 99], and it is known that the limit of their free energy admits a formula analogous to (6.72) provided that their associated non-linearity is convex; however, in general, this formula needs to be modified into a "sup-inf" formulation. Possibly the simplest example in which this happens is for the problem in which a rank-one matrix of the form $XY^\top$ plus noise is observed, where X and Y are two vectors with i.i.d. coordinates. In this setting, the non-linearity C_∞ in the Hamilton-Jacobi equation (3.91) is replaced by the non-convex mapping $(x,y)\mapsto xy$, and the functional to be optimized over as in (6.72) is $\psi(x_0+tx, y_0+ty)-\frac{t}{2}xy$. Finding counter-examples to the formula with only a supremum over y is made relatively easy by considering candidates with, say, $x=0$; in this case, the counter-term xy vanishes, so the parameter y can be chosen as large as desired to maximize the ψ functional and obtain a contradiction. A similar phenomenon also occurs in the context of spin glasses, and a more precise discussion of this point can be found in Section 6.2 of [86].

Coming back to the sparse stochastic block model, this observation can be leveraged to demonstrate that the formula (6.72) would also be invalid in general. To give a concrete example, consider the following scenario, which can be thought of as a problem with four communities, or as a bipartite version of the two-community problem. First colour the N nodes in red or blue, say with groups of sizes about $N/2$, and think of this colouring as fixed, e.g. the red nodes are the first $\lfloor N/2\rfloor$ indices in $\{1,\ldots,N\}$, and this is perfectly known to the statistician. Next, attribute ± 1 labels to each node independently, possibly with different biases according to the colour of the node. Finally, draw links between nodes i and j according to the formula (1.14), with the additional constraint that only links between nodes of different colours are allowed. The task is to study the asymptotic behaviour of the mutual information between the ± 1 labels and the observed graph. This problem is constructed in such a way that, in the limit of diverging average degree, it reduces to the problem of observing a noisy version of $XY^\top$, as discussed in the previous paragraph — the vectors X and Y contain the ± 1 labels of the red and blue nodes respectively. Using the results of [45, 68] to justify the large-degree approximation, or possibly even directly, the author is confident that counter-examples to the formula (6.72) can be produced.

For fully-connected models with possibly non-convex non-linearities such as the XY^T example, the limit of the free energy was identified in the form of a "sup-inf" formula; see [36] for the most general results. Translating this result into the present context would suggest that the limit free energy might be given by

$$\sup_{\rho\in\mathcal{M}_+}\inf_{\nu\in\mathcal{M}_+}\left(\psi(\nu)+\int_{-1}^{1}G_\rho(y)\,\mathrm{d}(\mu-\nu)(y)+\frac{t}{2}\int_{-1}^{1}G_\rho(y)\,\mathrm{d}\rho(y)\right). \tag{6.77}$$

The key ingredient for showing the validity of the corresponding formula in the dense stochastic block model is the convex selection principle discussed in Section 2.6. In particular, this relies on the observation that the enriched free energy is a convex function of its parameters. In the context of the sparse stochastic block model, the question would translate into whether the mapping $(t,\mu)\mapsto\overline{F}_N(t,\mu)$ is convex. However, it was shown in [67] that this mapping is in fact *not* convex in the sparse regime, even in the limit of large N. This non-convexity property not only breaks down the proof strategy in Section 2.7, which the author had originally hoped to carry through to the sparse regime; in fact, it can be leveraged to assert that the quantity (6.77) can *not* be the limit of the free energy in this case. Indeed, the expression in (6.77) is a supremum over ρ of affine functions of (t,μ), so the whole expression is convex in (t,μ). By [67], it is therefore not possible that the expression in (6.77) be the limit of the free energy.

To sum up, if the aim is for a formula that is robust to model changes, then both (6.72) and (6.77) can be ruled out. The author is not aware of alternative candidate variational formulas for the limit of the free energy. This situation seems analogous to that encountered in the context of spin glasses with possibly non-convex interactions, as discussed in Section 6 of [86].

Another non-variational alternative to Conjecture 1.4 would be that the limit of the free energy is the maximal value one gets by plugging every possible solution of the fixed-point equation (6.75) into the functional inside the supremum in (6.72). However, in view of the discussion in the previous paragraph, counter-examples to the variational formula in (6.72) seem to produce counter-examples to this possibility as well.

This apparent lack of a variational formula for the limit of the free energy in the sparse stochastic block model makes the Hamilton-Jacobi approach so appealing. Indeed, at the very least, this approach allows one to phrase a conjecture for the limit of the free energy. Unfortunately, without access to the convex selection principle, the lower bound in Conjecture 1.4 remains open, and little progress has been made on this lower bound since the publication of [49, 67].

Appendix A

Basic results in analysis and probability

In this appendix, five elementary topics in analysis and probability are discussed. In Section A.1, various representations of convex sets are established. The first is the classical result that a closed convex cone coincides with its bi-dual, the second is a non-differential characterization of a Lipschitz function having its gradient in a closed convex set, and the third is a representation of a convex set as the intersection of all the closed and affine half-spaces that contain it. These results are essential in the study of the sparse stochastic block model. In Section A.2, the classical Fenchel-Moreau theorem is extended to the setting of positive half-space. This result plays an important role in establishing the Hopf-Lax variational formula for Hamilton-Jacobi equations on positive half-space in Section 2.4.4. In Section A.3, the subdifferential of a convex function is discussed. More precisely, it is shown that a convex function is differentiable if and only if its subdifferential consists of a single point, and that, unlike the derivative of a general function, the subdifferential of a convex function is amenable to taking limits. These results are fundamental when establishing the convex selection principle in Section 2.6. In Section A.4, the basic properties of semi-continuous envelopes are discussed. These are leveraged in the Perron argument for proving the existence of solutions to Hamilton-Jacobi equations on positive half-space in Section 2.4.2. Finally, in Section A.5, some basic properties of Binomial and Poisson random variables are discussed. More specifically, moment bounds for the Poisson distribution are established, the Binomial-Poisson approximation theorem is proved, and the Poisson colouring theorem is shown. The moment bounds for the Poisson distribution are used in Section 6.2 to obtain the concentration of the free energy, the Binomial-Poisson approximation theorem is used in Section 3.2 to show that the free energy in the sparse stochastic block model can be modified without changing its limiting value, and the Poisson colouring theorem is used in Section 3.4 to compute the limit of the enriched free energy at the initial time. For completeness, a proof of any result used directly in the main body of the thesis is provided.

A.1 Representations of convex sets

A convex set $\mathcal{K} \subseteq \mathbb{R}^d$ is said to be a *cone* if, for all $x \in \mathcal{K}$ and $\lambda > 0$, one has $\lambda x \in \mathcal{K}$. The *dual* of convex set $\mathcal{K}$ is the closed convex cone

$$\mathcal{K}^* = \left\{ x \in \mathbb{R}^d \mid x \cdot y \geqslant 0 \text{ for all } y \in \mathcal{K} \right\}. \tag{A.1}$$

It is clear that any convex set $\mathcal{K}$ is always a subset of its bi-dual $\mathcal{K}^{**}$. Since $\mathcal{K}^{**}$ is a closed cone, a necessary condition for this containment to be equality is that $\mathcal{K}$ be a closed cone; it turns out that this is also a sufficient condition.

Proposition A.1. *If $\mathcal{K} \subseteq \mathbb{R}^d$ is a non-empty closed convex cone, then $\mathcal{K} = \mathcal{K}^{**}$.*

Proof. This is essentially Exercise 2.14 in [50]. It is clear that $\mathcal{K} \subseteq \mathcal{K}^{**}$. Suppose for the sake of contradiction that there exists $x \in \mathcal{K}^{**} \setminus \mathcal{K}$. The supporting hyperplane theorem (Theorem 2.2 in [50]) gives a non-zero vector $v \in \mathbb{R}^d$ with

$$v \cdot x > \sup\{v \cdot y \mid y \in \mathcal{K}\}. \tag{A.2}$$

Given $x_0 \in \mathcal{K}$, which exists as $\mathcal{K}$ is not empty, the assumption that $\mathcal{K}$ is closed implies that $0 = \lim_{\lambda \searrow 0} \lambda x_0 \in \mathcal{K}$. Together with (A.2), this implies that $v \cdot x > 0$. If there were $y_0 \in \mathcal{K}$ with $v \cdot y_0 > 0$, the fact that $\mathcal{K}$ is a cone would imply that $v \cdot x \geqslant \lambda v \cdot y_0$ for all $\lambda > 0$, and letting λ tend to infinity would give a contradiction. It follows from (A.2) that $v \cdot x > 0 = \sup\{v \cdot y \mid y \in \mathcal{K}\}$, where the fact that $0 \in \mathcal{K}$ has been used. The lower bound implies that $-v \in \mathcal{K}^*$ while the upper bound gives $x \cdot (-v) < 0$. This contradicts the assumption that $v \in \mathcal{K}^{**}$ and completes the proof. ■

This result implies that one can verify whether a point $x \in \mathbb{R}^d$ belongs to the convex cone $\mathcal{K} \subseteq \mathbb{R}^d$ by inspecting the sign of $x \cdot v$ for every $v \in \mathcal{K}^*$. In the context of the sparse stochastic block model, it will be important to have a similar criterion to determine whether the gradient of a Lipschitz function that is not necessarily differentiable everywhere belongs to a closed convex set.

Proposition A.2. *If $\mathcal{K} \subseteq \mathbb{R}^d$ is a closed convex set and $\psi : \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz function, then $\nabla \psi \in \mathcal{K}$ if and only if the following holds. For every $c \in \mathbb{R}$ and $x, x' \in \mathbb{R}^d$ with the property that for every $z \in \mathcal{K}$, one has $(x' - x) \cdot z \geqslant c$, then $\psi(x') - \psi(x) \geqslant c$.*

Proof. This is Proposition B.2 in [48]. To begin with, suppose that $\nabla \psi \in \mathcal{K}$, and fix $c \in \mathbb{R}$ and $x, x' \in \mathbb{R}^d$ with the property that for every $z \in \mathcal{K}$, one has $(x' - x) \cdot z \geqslant c$. If ψ were differentiable almost everywhere on the line joining x and x', the fundamental theorem of calculus could be applied to the one-dimensional Lipschitz function $t \mapsto \psi(x + t(x' - x))$ to conclude that

$$\psi(x') - \psi(x) = \int_0^1 \nabla \psi\big(x + t(x' - x)\big) \cdot (x' - x) \, \mathrm{d}t \geqslant c.$$

Although ψ could fail to be differentiable almost everywhere on the line joining x and x', it will now be shown that, given $\varepsilon > 0$, it must be differentiable almost everywhere on some line joining some point $x_\varepsilon \in B_\varepsilon(x)$ and some point $x'_\varepsilon \in B_\varepsilon(x')$. Denote by

$$\mathcal{H} := \left\{ y \in \mathbb{R}^d \mid y \cdot (x' - x) = 0 \right\} \cong \mathbb{R}^{d-1}$$

the hyperplane perpendicular to the line segment joining x and x', and write $\mathcal{A}_{\varepsilon,x} := B_\varepsilon(x) \cap (x + \mathcal{H})$ for the cross-section of $B_\varepsilon(x)$ through x and perpendicular to the line segment joining x and x'. Denote by $\mathcal{L}$ the set of line segments between points in $\mathcal{A}_{\varepsilon,x}$ and points in $\mathcal{A}_{\varepsilon,x'}$ which are parallel to the line segment joining x and x'. For each $y \in \mathcal{A}_{\varepsilon,x}$, write $\ell_y \in \mathcal{L}$ for the unique line segment in $\mathcal{A}_{\varepsilon,x}$ through y, and introduce the set

$$\mathcal{D}_y := \left\{ z \in \ell_y \mid \psi \text{ is not differentiable at } z \right\}$$

of points on ℓ_y at which ψ is not differentiable. If, for every $y \in \mathcal{A}_{\varepsilon,x}$, the set $\mathcal{D}_y$ were of positive one-dimensional Lebesgue measure $m_1(\mathcal{D}_y) > 0$, then the d-dimensional Lebesgue measure of the set of points in

$\cup_{y \in \mathcal{A}_{\varepsilon,x}} \ell_y$ at which ψ is not differentiable would have positive measure,

$$\int_{\mathcal{A}_{\varepsilon,x}} m_1(\mathcal{D}_y)\,\mathrm{d}y > 0.$$

This would contradict Rademacher's theorem (Theorem 2.10 in [50]) on the almost everywhere differentiability of Lipschitz functions. It is therefore possible to find $x_\varepsilon \in \mathcal{A}_{\varepsilon,x}$ with $m_1(\mathcal{D}_{x_\varepsilon}) = 0$. Writing $x'_\varepsilon \in \mathcal{A}_{\varepsilon,x'}$ for the right endpoint of ℓ_{x_ε}, the fundamental theorem of calculus implies that

$$\psi(x'_\varepsilon) - \psi(x_\varepsilon) = \int_0^1 \nabla\psi\big(x_\varepsilon + t(x'_\varepsilon - x_\varepsilon)\big)\cdot(x'-x)\,\mathrm{d}t \geqslant c.$$

Letting ε tend to zero shows that $\psi(x') - \psi(x) \geqslant c$ as required. Conversely, suppose that for every $c \in \mathbb{R}$ and $x, x' \in \mathbb{R}^d$ with the property that $(x'-x)\cdot z \geqslant c$ for every $z \in \mathcal{K}$, one has $\psi(x') - \psi(x) \geqslant c$. Assume for the sake of contradiction that there exists $y \in \mathbb{R}^d$ with $\nabla\psi(y) \notin \mathcal{K}$. The supporting hyperplane theorem (Theorem 2.2 in [50]) gives a non-zero vector $v \in \mathbb{R}^d$ and $\delta > 0$ with

$$v\cdot\nabla\psi(y) + \delta < \inf\{v\cdot z \mid z \in \mathcal{K}\}.$$

It follows that

$$\psi(y+\varepsilon v) - \psi(y) \geqslant \varepsilon\big(v\cdot\nabla\psi(y) + \delta\big).$$

Dividing by ε and letting ε tend to zero reveals that $\nabla\psi(y)\cdot v \geqslant v\cdot\nabla\psi(y) + \delta$. This contradiction completes the proof. ■

In the special case when the convex set $\mathcal{K}$ is a cone, this result implies that the gradient of a Lipschitz function $\psi : \mathbb{R}^d \to \mathbb{R}$ belongs to the closed convex cone $\mathcal{K}$ if and only if ψ is $\mathcal{K}^*$-non-decreasing. Recall from (4.74) that the function ψ is said to be *$\mathcal{K}^*$-non-decreasing* if, for all $x, x' \in \mathbb{R}^d$ with $x' - x \in \mathcal{K}^*$, one has $\psi(x') - \psi(x) \geqslant 0$.

Corollary A.3. *If $\mathcal{K}$ is a non-empty closed convex cone, and $\psi : \mathbb{R}^d \to \mathbb{R}$ is a Lipschitz function, then $\nabla\psi \in \mathcal{K}$ if and only if ψ is $\mathcal{K}^*$-non-decreasing.*

Proof. On the one hand, if $\nabla\psi \in \mathcal{K}$, then Proposition A.2 applied with $c = 0$ shows that ψ is $\mathcal{K}^*$-non-decreasing. On the other hand, if ψ is $\mathcal{K}^*$-non-decreasing but there exists $y \in \mathbb{R}^d$ with $\nabla\psi(y) \notin \mathcal{K}$, then the supporting hyperplane theorem (Theorem 2.2 in [50]) gives a non-zero vector $v \in \mathbb{R}^d$ with

$$v\cdot\nabla\psi(y) < \inf\{v\cdot z \mid z \in \mathcal{K}\}.$$

Observe that $v \in \mathcal{K}^*$. Indeed, if this were not the case, there would exist $z \in \mathcal{K}$ with $v\cdot z < 0$. Since $\mathcal{K}$ is a cone, this would mean that $v\cdot\nabla\psi(y) \leqslant \lambda v\cdot z$ for all $\lambda > 0$, which would lead to a contradiction upon letting λ tend to infinity. Notice also that $0 = \lim_{\lambda\searrow 0}\lambda x_0 \in \mathcal{K}$, where x_0 denotes any point in $\mathcal{K}$. It follows that $\inf\{v\cdot z \mid z \in \mathcal{K}\} = 0$, and therefore that $v\cdot\nabla\psi(y) < 0$. Since $v \in \mathcal{K}^*$, the $\mathcal{K}^*$-non-decreasingness of ψ implies that

$$\psi(y+\varepsilon v) - \psi(y) \geqslant 0.$$

Dividing by ε and letting ε tend to zero leads to a contradiction that completes the proof. ■

Proposition A.4. *If $\mathcal{K}$ is a closed convex set, then*

$$\mathcal{K} = \{x \in \mathbb{R}^d \mid x \cdot v \geqslant c \text{ for all } (v,c) \in \mathcal{A}\}, \tag{A.3}$$

where $\mathcal{A} := \{(v,c) \in \mathbb{R}^{d+1} \mid x \cdot v \geqslant c \text{ for all } x \in \mathcal{K} \text{ and } |v| = 1\}$ is a representative of the set of closed hyper-spaces containing $\mathcal{K}$.

Proof. This is Corollary 4.2.4 of [62]. To alleviate notation, let $C := \{x \in \mathbb{R}^d \mid x \cdot v \geqslant c \text{ for all } (v,c) \in \mathcal{A}\}$. By definition of $\mathcal{A}$, it is clear that $\mathcal{K} \subseteq C$. Suppose for the sake of contradiction that there exists $x \in C \setminus \mathcal{K}$. The hyperplane separation theorem (Theorem 2.2 in [50]) gives $\delta > 0$ and $v \in \mathbb{R}^d$ with $|v| = 1$ such that

$$v \cdot x + \delta < \inf\{v \cdot y \mid y \in \mathcal{K}\}.$$

This means that $(v, v \cdot x + \delta) \in \mathcal{A}$. Since $x \in C$, it follows that $x \cdot v \geqslant x \cdot v + \delta$. This contradiction completes the proof. ∎

A.2 The Fenchel-Moreau theorem on positive half-space

The *convex dual* of a proper function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is the function $f_* : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f_*(\lambda) := \sup_{x \in \mathbb{R}^d} (\lambda \cdot x - f(x)). \tag{A.4}$$

If f_* is proper, then this operation can be iterated to obtain the *convex bi-dual* of f, which is the function $f_{**} : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f_{**}(x) := (f_*)_*(x) = \sup_{\lambda \in \mathbb{R}^d} (x \cdot \lambda - f_*(\lambda)). \tag{A.5}$$

A classical result in convex analysis is that a function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is convex and lower semi-continuous if and only if it is the supremum of its affine minorants (Proposition 2.4 in [50]). The Fenchel-Moreau theorem refines this result by identifying an explicit set of affine minorants of f whose supremum is f.

Proposition A.5 (Fenchel-Moreau on Euclidean space)**.** *If $f : \mathbb{R}^d \to \mathbb{R}$ is convex and lower semi-continuous, then it is equal to its convex bi-dual,*

$$f = f_{**}. \tag{A.6}$$

Proof. See Theorem 2.5 in [50]. ∎

In Section 2.4.4, when establishing the Hopf-Lax variational formula for Hamilton-Jacobi equations on positive half-space, it will be desirable to have a version of this result for functions defined on positive half-space. With the appropriate interpretation of the convex dual, and under the additional assumption of non-decreasingness, the Fenchel-Moreau theorem extends to the setting of positive half-space. The *convex dual* of a proper function $f : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R} \cup \{+\infty\}$ is the function $f^* : \mathbb{R}^d_{\geqslant 0} \to \mathbb{R} \cup \{+\infty\}$ defined by

$$f^*(\lambda) := \sup_{x \in \mathbb{R}^d_{\geqslant 0}} (\lambda \cdot x - f(x)). \tag{A.7}$$

If f^* is proper, the *convex bi-dual* of f is the function $f^{**}:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}\cup\{+\infty\}$ defined by

$$f^{**}(x) := (f^*)^*(x) = \sup_{\lambda\in\mathbb{R}^d_{\geqslant 0}} (x\cdot\lambda - f^*(\lambda)). \tag{A.8}$$

Recall from (2.69) that a function $f:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ is said to be non-decreasing if, for all $x',x\in\mathbb{R}^d_{\geqslant 0}$ with $x'-x\in\mathbb{R}^d_{\geqslant 0}$, one has $f(x)\leqslant f(x')$.

Proposition A.6 (Fenchel-Moreau on positive half-space)**.** *If $f:\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$ is convex, lower semi-continuous, and non-decreasing, then it is equal to its convex bi-dual,*

$$f = f^{**}. \tag{A.9}$$

Proof. For each $x\in\mathbb{R}^d$, denote by $\widetilde{x}\in\mathbb{R}^d_{\geqslant 0}$ its symmetrization, $\widetilde{x} := (\widetilde{x}_1,\ldots,\widetilde{x}_d) := \big(|x|_1,\ldots,|x|_d\big)$, and define the symmetrization $\widetilde{f}:\mathbb{R}^d\to\mathbb{R}$ of the function f by $\widetilde{f}(x) := f(\widetilde{x})$. The proof proceeds in two steps. First, it is shown that $\widetilde{f} = \widetilde{f}_{**}$ on $\mathbb{R}^d$, and then that $\widetilde{f}_{**} = f^{**}$ on $\mathbb{R}^d_{\geqslant 0}$.

Step 1: $\widetilde{f} = \widetilde{f}_{**}$ on $\mathbb{R}^d$. By the Fenchel-Moreau theorem on Euclidean space, it suffices to show that the symmetrization $\widetilde{f}$ of f is lower semi-continuous and convex. The lower semi-continuity of $\widetilde{f}$ is immediate from that of f. To establish convexity, fix $\alpha\in[0,1]$, $t,t'\geq 0$ and $x,x'\in\mathbb{R}^d$. By non-decreasingness of f,

$$\widetilde{f}\big(\alpha t+(1-\alpha)t',\alpha x+(1-\alpha)x'\big)\leqslant f\big(\alpha t+(1-\alpha)t',\alpha\widetilde{x}+(1-\alpha)\widetilde{x}'\big),$$

and by convexity of f,

$$\widetilde{f}\big(\alpha t+(1-\alpha)t',\alpha x+(1-\alpha)x'\big)\leqslant \alpha f(t,\widetilde{x})+(1-\alpha)f(t',\widetilde{x}') = \alpha\widetilde{f}(t,x)+(1-\alpha)\widetilde{f}(t',x').$$

It follows by the Fenchel-Moreau theorem on Euclidean space that the symmetrization $\widetilde{f}$ is equal to its convex bi-dual, $\widetilde{f} = \widetilde{f}_{**}$.

Step 2: $\widetilde{f}_{**} = f^{**}$ on $\mathbb{R}^d_{\geqslant 0}$. Fix $\lambda\in\mathbb{R}^d$, and for every $x\in\mathbb{R}^d$, let $x_\lambda := (x_1\operatorname{sgn}\lambda_1,\ldots,x_d\operatorname{sgn}\lambda_d)$. By symmetry of $\widetilde{f}$,

$$\widetilde{f}_*(\lambda) = \sup_{x\in\mathbb{R}^d}\big(x\cdot\lambda-\widetilde{f}(x)\big) = \sup_{x\in\mathbb{R}^d}\big(x_\lambda\cdot\lambda-\widetilde{f}(x_\lambda)\big) = \sup_{x\in\mathbb{R}^d}\big(x\cdot\widetilde{\lambda}-\widetilde{f}(x)\big) = \widetilde{f}_*(\widetilde{\lambda}).$$

It follows that for every $x\in\mathbb{R}^d_{\geqslant 0}$,

$$\widetilde{f}_{**}(x) = \sup_{\lambda\in\mathbb{R}^d}\big(x\cdot\lambda-\widetilde{f}_*(\lambda)\big) = \sup_{\lambda\in\mathbb{R}^d}\big(x\cdot\lambda-\widetilde{f}_*(\widetilde{\lambda})\big) = \sup_{\lambda\in\mathbb{R}^d_{\geqslant 0}}\big(x\cdot\lambda-\widetilde{f}_*(\lambda)\big), \tag{A.10}$$

where the final equality uses that $x\cdot\lambda\leqslant x\cdot\widetilde{\lambda}$ for every $\lambda\in\mathbb{R}^d$. To simplify this further, observe that for $\lambda\in\mathbb{R}^d_{\geqslant 0}$,

$$\widetilde{f}_*(\lambda) = \sup_{y\in\mathbb{R}^d}\big(y\cdot\lambda-\widetilde{f}(y)\big) = \sup_{y\in\mathbb{R}^d}\big(y\cdot\lambda-f(\widetilde{y})\big) = \sup_{y\in\mathbb{R}^d_{\geqslant 0}}\big(y\cdot\lambda-f(y)\big) = f^*(\lambda),$$

where the penultimate equality uses that $y\cdot\lambda\leqslant\widetilde{y}\cdot\lambda$ for all $y\in\mathbb{R}^d$. Substituting this into (A.10), and invoking the previous step completes the proof. ∎

A.3 The subdifferential of a convex function

The differentiability of a convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ at points of its effective domain,

$$\operatorname{dom} f := \{x \in \mathbb{R}^d \mid f(x) < +\infty\}, \tag{A.11}$$

is best studied through the notion of subdifferential. The *subdifferential* of a convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ at a point $x \in \operatorname{dom} f$ is the set

$$\partial f(x) := \{p \in \mathbb{R}^d \mid f(y) \geqslant f(x) + p \cdot (y - x) \text{ for all } y \in \mathbb{R}^d\}. \tag{A.12}$$

It turns out that a convex function is differentiable at a point x in the interior of its effective domain if and only if its subdifferential at x consists of a singleton. To prove this, it will be convenient to observe that a convex function is differentiable at a point in the interior of its effective domain if and only if its directional derivative is a linear function of the direction. The *directional derivative* of a convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ at a point $x \in \operatorname{int}(\operatorname{dom} f)$ in the direction $\nu \in \mathbb{R}^d$ is the function

$$D_\nu f(x) := \lim_{t \searrow 0} \frac{f(x + t\nu) - f(x)}{t}. \tag{A.13}$$

The convexity of f implies that the difference quotient defining the directional derivative is a decreasing function of t, so the directional derivative is well-defined as a monotone limit. In general, the linearity of the directional derivative does not suffice to characterize the differentiability of a function; however, it does suffice for locally Lipschitz continuous functions, and therefore for points in the interior of the effective domain of a convex function. Indeed, a convex function is locally Lipschitz continuous on the interior of its effective domain (Proposition 2.9 in [50]).

Lemma A.7. *A convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is differentiable at $x \in \operatorname{int}(\operatorname{dom} f)$ if and only if the map $\nu \mapsto D_\nu f(x)$ is linear. In this case, one has $D_\nu f(x) = \nabla f(x) \cdot \nu$.*

Proof. This is Lemma 2.12 in [50]. On the one hand, if f is differentiable, then for any direction vector $\nu \in \mathbb{R}^d$, one has $D_\nu f(x) = \nabla f(x) \cdot \nu$. Conversely, suppose that $\nu \mapsto D_\nu f(x)$ is linear, that is, there exists $a \in \mathbb{R}^d$ such that $D_\nu f(x) = a \cdot \nu$ for every $\nu \in \mathbb{R}^d$. Assume for the sake of contradiction that f is not differentiable at x. Let $(\nu_n)_{n \geqslant 1}$ with $|\nu_n| = 1$ for all $n \geq 1$ and $(t_n)_{n \geqslant 1} \subseteq \mathbb{R}_{>0}$ be a sequence converging to 0 such that the error term

$$R(x, \nu_n, t_n) := \left| \frac{f(x + t_n \nu_n) - f(x)}{t_n} - a \cdot \nu_n \right|$$

does not converge to zero. Up to passing to a subsequence, assume that $(\nu_n)_{n \geqslant 1}$ converges to some ν_0 in the unit sphere. Remembering that a convex function is locally Lipschitz continuous on the interior of its effective domain (Proposition 2.9 in [50]), denote by $L > 0$ the Lipschitz constant of f around x. The triangle inequality implies that

$$R(x, \nu_n, t_n) \leqslant R(x, \nu_0, t_n) + (L + |a|)|\nu_n - \nu_0|.$$

Leveraging the assumption that $D_{\nu_0} f(x) = a \cdot \nu_0$ to let n tend to infinity contradicts the absurd hypothesis that the sequence $(R(x, \nu_n, t_n))_{n \geqslant 1}$ does not converge to zero. This completes the proof. ∎

Proposition A.8. *Let $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ be a convex function. For every point $x \in \operatorname{int}(\operatorname{dom} f)$, the subdifferential $\partial f(x)$ is not empty.*

Proof. This is Proposition 2.11 in [50]. Consider the convex set $C := \{(y,\mu) \in \mathbb{R}^d \times \mathbb{R} \mid f(y) > \mu\}$, and fix $x \in \operatorname{dom} f$. Since $(x, f(x)) \notin C$ the supporting hyperplane theorem (Theorem 2.2 in [50]) gives a non-zero vector $(v,b) \in \mathbb{R}^d \times \mathbb{R}$ with

$$0 \leqslant v \cdot (y-x) + b(\mu - f(x)).$$

for every $(y,\mu) \in C$. Since μ can be arbitrarily large, it must be the case that $b \geqslant 0$. If it were the case that $b = 0$, then $0 \leqslant v \cdot (y-x)$ for all y in a neighbourhood of $x \in \operatorname{int}(\operatorname{dom} f)$, which is not possible since (v,b) is non-zero. This means that $b > 0$ so the vector $p := -v/b$ is well-defined and satisfies

$$\mu \geqslant f(x) + p \cdot (y-x)$$

for all $(y,\mu) \in C$. Letting μ tend to $f(y)$ reveals that $p \in \partial f(x)$, which means that the subdifferential $\partial f(x)$ is not empty. This completes the proof. ∎

Proposition A.9. *A convex function $f : \mathbb{R}^d \to \mathbb{R} \cup \{+\infty\}$ is differentiable at a point $x \in \operatorname{int}(\operatorname{dom} f)$ if and only if $\partial f(x)$ consists of a singleton. In this case, one has $\partial f(x) = \{\nabla f(x)\}$.*

Proof. This is Theorem 2.13 in [50]. The forward direction is established first. Recall from Proposition A.8 that the subdifferential $\partial f(x)$ is not empty. Fix $p \in \partial f(x)$. By definition of the subdifferential, for every $v \in \mathbb{R}^d$ and $\lambda > 0$,

$$f(x+\lambda v) - f(x) \geqslant \lambda v \cdot p.$$

Dividing by λ and letting λ tend to zero shows that $(\nabla f(x) - p) \cdot v \geqslant 0$. Choosing $v := p - \nabla f(x)$ reveals that $p = \nabla f(x)$, so $\partial f(x) = \{\nabla f(x)\}$. Conversely, suppose that the subdifferential is a singleton $\partial f(x) = \{p\}$, and fix a direction vector $\nu \in \mathbb{R}^d$. The convexity of f and the definition of the directional derivative imply that for all $\lambda \in \mathbb{R}$,

$$f(x) + \lambda D_\nu f(x) \leqslant f(x+\lambda \nu).$$

This means that the convex sets

$$C := \{(x+\lambda\nu, f(x) + \lambda D_\nu f(x)) \mid \lambda \in \mathbb{R}\}$$

and $\operatorname{int}(\operatorname{epi} f)$ are disjoint. Recall that the epigraph of f is the convex set

$$\operatorname{epi} f := \{(x,\lambda) \in \mathbb{R}^d \times \mathbb{R} \mid f(x) \leqslant \lambda\}.$$

It follows by the Hahn-Banach separation theorem (Theorem 2.3 in [50]) that there exists a non-zero vector $(a,b) \in \mathbb{R} \times \mathbb{R}^d$ with

$$a(f(x) + \lambda D_\nu f(x)) + b \cdot (x + \lambda \nu) \leqslant a\mu + b \cdot y \tag{A.14}$$

for all $(y,\mu) \in \operatorname{int}(\operatorname{epi} f)$ and $\lambda \in \mathbb{R}$. Taking $\lambda = 0$ shows that

$$af(x) + b \cdot x \leqslant a\mu + b \cdot y \tag{A.15}$$

for all $(y,\mu) \in \operatorname{int}(\operatorname{epi} f)$. Since μ can be arbitrarily large, it must be the case that $a \geqslant 0$. If it were the case that $a = 0$, then $0 \leqslant b \cdot (y-x)$ for all $y \in \mathbb{R}^d$, which is not possible since (a,b) is non-zero. Dividing through by a and letting μ tend to $f(y)$ in (A.15) shows that $-b/a \in \partial f(x)$, and therefore $b/a = -p$. Combining this with

(A.14) and letting μ tend to $f(y)$ in the resulting bound gives

$$f(x)+\lambda D_v f(x)-p\cdot(x+\lambda v)\leqslant f(y)-p\cdot y$$

for all $y\in\mathbb{R}^d$ and $\lambda\in\mathbb{R}$. Taking $y=x$ reveals that $\lambda(D_v f(x)-p\cdot v)\leqslant 0$ for all $\lambda\in\mathbb{R}$, which implies that $D_v f(x)=p\cdot v$. In particular, the map $v\mapsto D_v f(x)$ is linear. Invoking Lemma A.7 completes the proof. ■

In the context of the Hamilton-Jacobi approach, one has a sequence $(\overline{F}_N)_{N\geqslant 1}$ of differentiable and convex free energy functionals defined on $\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}$, and one is interested in the differential properties of the possibly non-differentiable convex limit $f:\mathbb{R}_{\geqslant 0}\times\mathbb{R}^d_{\geqslant 0}\to\mathbb{R}$. This requires understanding how the subdifferential interacts with limits.

Proposition A.10. *Let $f:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ be a convex function, and let $(x_n,p_n)_{n\geqslant 1}$ be a sequence of points in $\operatorname{dom} f\times\mathbb{R}^d$ with $p_n\in\partial f(x_n)$ for each $n\geqslant 1$ that converges to some point $(x,p)\in\operatorname{dom} f\times\mathbb{R}^d$. If f is lower semi-continuous at $x\in\operatorname{dom} f$, then $p\in\partial f(x)$.*

Proof. This is Proposition 2.14 in [50]. Fix $y\in\mathbb{R}^d$ as well as $n\geqslant 1$. Since $p_n\in\partial f(x_n)$,

$$f(y)\geqslant f(x_n)+p_n\cdot(y-x_n)$$

Letting n tend to infinity and using the lower semi-continuity of f at x completes the proof. ■

Remark A.11. If the point x in Proposition A.10 belongs to $\operatorname{int}(\operatorname{dom} f)$, the continuity of f at x is automatically satisfied since a convex function is locally Lipschitz continuous on the interior of its effective domain (Proposition 2.9 in [50]).

Proposition A.12. *For each integer $n\geqslant 1$, let $f_n:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ be a differentiable convex function. If $(f_n)_{n\geqslant 1}$ converges pointwise to some function $f:\mathbb{R}^d\to\mathbb{R}\cup\{+\infty\}$ and $x\in\operatorname{int}(\operatorname{dom} f)$ is such that the sequence of derivatives $(\nabla f_n(x))_{n\geqslant 1}$ converges to some vector $p\in\mathbb{R}^d$, then $p\in\partial f(x)$.*

Proof. This is Proposition 2.15 in [50]. Fix $y\in\mathbb{R}^d$ as well as $n\geqslant 1$. Since f_n is differentiable at the interior point $x\in\operatorname{int}(\operatorname{dom} f)$, one has $\nabla f_n(x)\in\partial f_n(x)$ by Proposition A.8. It follows by definition of the subdifferential that

$$f_n(y)\geqslant f_n(y)+\nabla f_n(x)(y-x).$$

Letting n tend to infinity and using the pointwise convergence of f_n to f completes the proof. ■

A.4 semi-continuous envelopes

Fix a set $X\subseteq\mathbb{R}^d$ endowed with a norm $\|\cdot\|$. A function $u:X\to\mathbb{R}$ is said to be *upper semi-continuous* at a point $x\in X$ if

$$u(x)\geqslant\limsup_{y\to x}u(y):=\lim_{r\searrow 0}\sup\big\{u(y)\mid y\in X\text{ with }\|y-x\|\leqslant r\big\},\tag{A.16}$$

and it is said to be *lower semi-continuous* at a point $x\in X$ if

$$u(x)\leqslant\liminf_{y\to x}u(y):=\lim_{r\searrow 0}\inf\big\{u(y)\mid y\in X\text{ with }\|y-x\|\leqslant r\big\}.\tag{A.17}$$

Moreover, the *upper semi-continuous envelope* of u is the function $\overline{u} : X \to \mathbb{R}$ defined by

$$\overline{u}(x) := \limsup_{y \to x} u(y) = \lim_{r \searrow 0} \sup \{u(y) \mid y \in X \text{ with } \|y - x\| \leqslant r\}, \tag{A.18}$$

while its *lower semi-continuous envelope* is the function $\underline{u} : X \to \mathbb{R}$ defined by

$$\underline{u}(x) := \liminf_{y \to x} u(y) = \lim_{r \searrow 0} \inf \{u(y) \mid y \in X \text{ with } \|y - x\| \leqslant r\}. \tag{A.19}$$

The following proposition collects the basic properties of semi-continuous envelopes. This result is used in Section 2.4.2 with $X := \mathbb{R}_{\geqslant 0} \times \mathbb{R}^d_{\geqslant 0}$ and $\|(t,x)\| := |t| + |x|$.

Proposition A.13. *The semi-continuous envelopes of a locally bounded function $u : X \to \mathbb{R}$ satisfy the following basic properties.*

(i) $\underline{u}(x) \leqslant u(x) \leqslant \overline{u}(x)$ *for all $x \in X$.*

(ii) $\overline{u}(x) = \min\{v(x) \mid u \leqslant v$ *and v is upper semi-continuous*$\}$ *for all $x \in X$. In particular, $\overline{u}$ is upper semi-continuous.*

(iii) $\underline{u}(x) = \max\{v(x) \mid v \leqslant u$ *and v is lower semi-continuous*$\}$ *for all $x \in X$. In particular, $\underline{u}$ is lower semi-continuous.*

(iv) u is upper semi-continuous at $x \in X$ if and only if $u(x) = \overline{u}(x)$.

(v) u is lower semi-continuous at $x \in X$ if and only if $u(x) = \underline{u}(x)$.

Proof. This is Proposition B.3 in [48]. To deduce the properties of the lower semi-continuous envelope from the corresponding properties of the upper semi-continuous envelope the observation that

$$\underline{u}(x) = \liminf_{y \to x} u(y) = -\limsup_{y \to x} (-u(y)) = -(\overline{-u})(x) \tag{A.20}$$

will be leveraged.

(i) This is immediate from the definition of the semi-continuous envelopes in (A.18) and (A.19).

(ii) If v is an upper semi-continuous function with $u \leqslant v$, taking the limsup as y tends to x on both sides of the inequality $u(y) \leqslant v(y)$, and leveraging the upper semi-continuity of v reveals that

$$\overline{u}(x) = \limsup_{y \to x} u(y) \leqslant \limsup_{y \to x} v(y) \leqslant v(x).$$

This implies that

$$\overline{u}(x) \leqslant \inf\{v(x) \mid u \leqslant v \text{ and } v \text{ is upper semi-continuous}\}.$$

To show that this infimum is achieved and that this inequality is, in fact, an equality, it suffices to prove that $\overline{u}$ is itself upper semi-continuous. Fix $x \in X$ as well as $\varepsilon > 0$, and find $r > 0$ with

$$\overline{u}(x) + \varepsilon > \sup\{u(y) \mid y \in X \text{ with } \|y - x\| \leqslant r\}.$$

The triangle inequality reveals that for any $z \in X$ with $\|z - x\| < r$,

$$\overline{u}(x) + \varepsilon \geqslant \sup\{u(y) \mid y \in X \text{ with } \|y - z\| \leqslant r - \|x - z\|\} \geqslant \overline{u}(z).$$

It follows that $\limsup_{z\to x}\overline{u}(z) \leqslant \overline{u}(x)$, so $\overline{u}$ is upper semi-continuous at x. Since x is arbitrary, this establishes the claim.

(iii) Combining the previous part with (A.20) shows that

$$\underline{u}(x) = -(\overline{-u})(x) = \max\big\{-v(x) \mid -u \leqslant v \text{ and } v \text{ is upper semi-continuous}\big\}.$$

Observing that v is upper semi-continuous if and only $-v$ is lower semi-continuous establishes the claim.

(iv) If u is upper semi-continuous at x, then

$$\overline{u}(x) = \limsup_{y\to x} u(y) \leqslant u(x).$$

Together with the inequality $u(x) \leqslant \overline{u}(x)$, this shows that $u(x) = \overline{u}(x)$. On the other hand, if $\overline{u}(x) = u(x)$, then

$$\limsup_{y\to x} u(y) = \overline{u}(x) = u(x) \leqslant u(x)$$

so u is upper semi-continuous at x.

(v) Observe that u is lower semi-continuous at $x \in X$ if and only if $-u$ is upper semi-continuous at $x \in X$. The previous part implies that this is the case if and only if $-u(x) = (\overline{-u})(x)$. Invoking (A.20) completes the proof. ■

A.5 Binomial and Poisson random variables

A random variable X is *Bernoulli* with probability of success $p \in (0,1)$, denoted by $X \sim \mathrm{Ber}(p)$, if it takes values in $\{0,1\}$ and is equal to one with probability p,

$$\mathbb{P}\{X = 1\} = p. \tag{A.21}$$

A random variable Y is *Binomial* with probability of success $p \in (0,1)$ and number of trials $n \in \mathbb{N}$, denoted by $Y \sim \mathrm{Bin}(n,p)$, if it is the sum of n independent and identically distributed Bernoulli random variables $(X_i)_{i\leqslant n}$ with probability of success p, that is, $Y = \sum_{i\leqslant n} X_i$. This means that Y takes values in $\{0,1,\ldots,n\}$, and for any $k \in \{0,1,\ldots,n\}$,

$$\mathbb{P}\{Y = k\} = \binom{n}{k} p^k (1-p)^{n-k}. \tag{A.22}$$

A random variable Π is *Poisson* with mean $\lambda > 0$, denoted by $\Pi \sim \mathrm{Poi}(\lambda)$, if it takes values in the natural numbers including zero, and for any $k \in \{0,1,2,\ldots\}$,

$$\mathbb{P}\{\Pi = k\} = \frac{\lambda^k}{k!}\exp(-\lambda). \tag{A.23}$$

Jensen's inequality ensures that for any integer $k \geqslant 1$, one has the lower moment bound $\lambda^k = (\mathbb{E}\Pi)^k \leqslant \mathbb{E}\Pi^k$. It turns out that the converse bound holds up to a multiplicative constant provided that $\mathbb{E}\Pi \geqslant 1$.

Proposition A.14. *If $\Pi \sim \mathrm{Poi}(\lambda)$ is a Poisson random variable with a mean larger than one, $\lambda \geqslant 1$, and $k \geqslant 2$*

is an integer, then there exists a constant $C_k < +\infty$ such that

$$\lambda^k \leqslant \mathbb{E}\Pi^k \leqslant C\lambda^k \quad \text{and} \quad \mathbb{E}\big(\Pi - \mathbb{E}\Pi\big)^k \leqslant C\lambda^{\lfloor k/2 \rfloor}. \tag{A.24}$$

Proof. This is Lemma B.2 in [49]. Denote by $\left\{ {k \atop j} \right\}$ the number of ways to partition a k element set into j non-empty subsets. In combinatorics, such numbers are known as Stirling numbers of the second kind, and they have the property that for any integer $m \geqslant 0$,

$$m^k = \sum_{j=0}^{k} \left\{ {k \atop j} \right\} (m)_j, \tag{A.25}$$

where $(m)_j = m(m-1)\cdots(m-j+1)$ is the falling factorial. Combining (A.25) with the basic properties of the Poisson distribution reveals that

$$\mathbb{E}\Pi^k = \sum_{m\geqslant 0}\sum_{j\leqslant k} \left\{ {k \atop j} \right\} (m)_j \frac{\lambda^m}{m!}\exp(-\lambda) = \sum_{j\leqslant k} \left\{ {k \atop j} \right\} \lambda^j \sum_{m\geqslant j} \frac{\lambda^{m-j}}{(m-j)!}\exp(-\lambda) = \sum_{j\leqslant k} \left\{ {k \atop j} \right\} \lambda^j \leqslant \max(1,\lambda^k) B_k,$$

where $B_k := \sum_{j=0}^{k} \left\{ {k \atop j} \right\}$ denotes the k'th Bell number. This establishes the first bound in (A.24). It is now shown by induction that for each $k \geqslant 2$, the function $M_k(\lambda) := \mathbb{E}(\Pi - \mathbb{E}\Pi)^k$ is a polynomial of degree $\lfloor k/2 \rfloor$. The base case holds since $M_2(\lambda) = \mathrm{Var}(\Pi) = \lambda$, so assume the result holds for all $2 \leqslant i \leqslant k$. By the product rule

$$\begin{aligned}
M_k'(\lambda) &= -\sum_{m\geqslant 0} k(m-\lambda)^{k-1}\frac{\lambda^m}{m!}\exp(-\lambda) + \sum_{m\geqslant 0}(m-\lambda)^k m \frac{\lambda^{m-1}}{m!}\exp(-\lambda) - M_k(\lambda) \\
&= -kM_{k-1}(\lambda) + \sum_{m\geqslant 0}(m-\lambda)^k (m-\lambda+\lambda)\frac{\lambda^{m-1}}{m!}\exp(-\lambda) - M_k(\lambda) \\
&= -kM_{k-1}(\lambda) + \frac{1}{\lambda}\big(M_{k+1}(\lambda) + \lambda M_k(\lambda)\big) - M_k(\lambda) \\
&= -kM_{k-1}(\lambda) + \frac{1}{\lambda}M_{k+1}(\lambda).
\end{aligned}$$

Invoking the induction hypothesis shows that $M_{k+1}(\lambda)$ has degree $\max(\lfloor k/2 \rfloor, 1 + \lfloor (k-1)/2 \rfloor)$. This completes the proof. ∎

The Binomial and Poisson distributions are intimately related. Indeed, a classical approximation theorem in probability theory is that, in the appropriate regime, the Poisson distribution can be approximated by the Binomial distribution. This result is known as the Binomial-Poisson approximation, and it can be obtained by controlling the total variation distance between a Poisson distribution and an approximating Binomial distribution. Given a separable metric space S, the *total variation distance* between two probability measures $\mathbb{P},\mathbb{Q} \in \Pr(S)$ is defined by

$$\mathrm{TV}(\mathbb{P},\mathbb{Q}) := \sup\big\{ |\mathbb{P}(A) - \mathbb{Q}(A)| \mid A \text{ is a measurable subset of } S \big\}. \tag{A.26}$$

Approximating any measurable function with values in S by a sequence of simple functions, one can verify that the total variation distance admits the dual representation

$$\mathrm{TV}(\mathbb{P},\mathbb{Q}) = \sup\left\{ \left| \int_{-1}^{1} f(x)\,\mathrm{d}\mathbb{P}(x) - \int_{-1}^{1} f(x)\,\mathrm{d}\mathbb{Q}(x) \right| \mid f : S \to [0,1] \text{ measurable} \right\}. \tag{A.27}$$

Using the Hahn-Jordan decomposition (Theorem 3.4 in [57]), it is also possible to show that

$$\mathrm{TV}(\mathbb{P},\mathbb{Q}) = \inf\big\{\mathbb{P}\{X \neq Y\} \mid X \sim \mathbb{P} \text{ and } Y \sim \mathbb{Q}\big\}. \tag{A.28}$$

This is the representation of the total variation distance that will be used to establish the Binomial-Poisson approximation theorem. More precisely, it will be applied to discrete probability measures supported on the set of natural numbers. In this case, it can be established using the Kantorovich-Rubinstein theorem (Theorem 4.15 in [90]).

Proposition A.15 (Bernoulli-Poisson approximation)**.** *Consider independent Bernoulli random variables $X_i \sim \mathrm{Ber}(p_i)$ for $i \leqslant n$, and let $\lambda_n := \sum_{i=1}^n p_i$. If $Y_n := \sum_{i\leqslant n} X_i$ and $\Pi_n \sim \mathrm{Poi}(\lambda_n)$, then*

$$\mathrm{TV}(Y_n,\Pi_n) \leqslant \sum_{i\leqslant n} p_i^2. \tag{A.29}$$

Proof. This is Theorem 2.4 in [90]. Temporarily fix $i \leqslant n$, and for each $k \geqslant 0$, introduce the constant

$$c_k := \sum_{\ell=0}^{k} \frac{p_i^\ell}{\ell!} e^{-p_i}.$$

Define the random variables X_i and X_i^* on $[0,1]$ endowed with the Borel σ-algebra and the Lebesgue measure $\mathbb{P}$ by

$$X_i := \mathbf{1}_{(1-p_i,1]} \quad \text{and} \quad X_i^* := \sum_{k=1}^{+\infty} k\mathbf{1}_{(c_{k-1},c_k]}$$

in such a way that $X_i \sim \mathrm{Ber}(p_i)$ and $X_i^* \sim \mathrm{Poi}(p_i)$. Since $1-p_i \leqslant e^{-p_i} = c_0$, the random variables X_i and X_i^* can only fail to be equal on the intervals $(1-p_i, c_0]$ and $(c_1,1]$. This means that

$$\mathbb{P}\{X_i \neq X_i^*\} \leqslant c_0 - (1-p_i) + 1 - c_1 = p_i(1-e^{-p_i}) \leqslant p_i^2.$$

At this point, construct pairs $(X_i, X_i^*)_{i\leqslant n}$ on separate coordinates of the product space $[0,1]^n$ endowed with the product Borel σ-algebra and the product Lebesgue measure, thus making them independent. Since the sum of independent Poisson random variables is again Poisson with parameter given by the sum of the individual parameters, the random variable $S_n^* := \sum_{i\leqslant n} X_i^*$ is equal in distribution to Π_n. It follows by the union bound that

$$\mathrm{TV}(Y_n,\Pi_n) \leqslant \mathbb{P}\{Y_n \neq S_n^*\} \leqslant \sum_{i=1}^n \mathbb{P}\{X_i \neq X_i^*\} \leqslant \sum_{i=1}^n p_i^2.$$

This completes the proof. ■

Another result that highlights the strong connections between the Binomial and Poisson distributions is the Poisson colouring theorem. To motivate this result, consider N balls each of which is painted red with probability $p \in (0,1)$ and blue with probability $1-p$. The number of red balls N_r is Binomial with probability of success p and number of trials n, while the number of blue balls N_b is Binomial with probability of success $1-p$ and number of trials n. The Poisson colouring theorem considers this same problem when, instead of having a deterministic number of balls N, the number of balls N is Poisson with mean λ. In this case it turns out that N_r is again Poisson with mean λp while N_b is Poisson with mean $\lambda(1-p)$.

Proposition A.16 (Poisson colouring)**.** *Let $N \sim \mathrm{Poi}(\lambda)$ be a Poisson random variable with mean $\lambda > 0$, and* *let* $(X_i)_{i\geqslant 1}$ *be independent and identically distributed random variables independent of* N *taking values in* $\{1,\dots,k\}$ *with* $\mathbb{P}\{X_1 = j\} = p_j$ *for* $1 \leqslant j \leqslant k$*. If* N_j *denotes the number of random variables among* $(X_i)_{i\leqslant N}$ *taking value* $j \in \{1,\dots,k\}$*, then* $(N_j)_{j\leqslant k}$ *are independent Poisson random variables with* $N_j \sim \mathrm{Poi}(\lambda p_j)$ *for* $1 \leqslant j \leqslant k$*.*

Proof. This is Exercise 1.2.9 in [90]. Fix non-negative integers $n_1,\dots,n_k \geqslant 0$, and let $n := n_1,\dots,n_k$ be their sum. Bayes' formula implies that

$$\mathbb{P}\{N_1 = n_1,\dots,N_k = n_k\} = \mathbb{P}\{N_1 = n_1,\dots,N_k = n_k \mid N = n\}\mathbb{P}\{N = n\}$$

Since the random variables $(X_i)_{i\geqslant 1}$ are independent of N, conditionally on the event $\{N = n\}$, the random vector $(N_1,\dots,N_k)$ is multinomial with probabilities of success $p_1,\dots,p_k$ and number of trials n. It follows that

$$\mathbb{P}\{N_1 = n_1,\dots,N_k = n_k\} = \frac{n!}{n_1!\dots n_k!} p_1^{n_1}\cdots p_k^{n_k} \frac{\lambda^n e^{-\lambda}}{n!} = \prod_{j=1}^{k} \frac{\lambda^{n_j} e^{-\lambda p_j}}{n_j!} p_j^{n_j}.$$

This shows that $N_1,\dots,N_k$ are independent Poisson random variables with means $\lambda p_1,\dots,\lambda p_k$, and completes the proof. ∎

Bibliography

[1] Emmanuel Abbe. “Community detection and stochastic block models: recent developments”. In: *Journal of Machine Learning Research* 18.1 (2017), pp. 6446–6531.

[2] Emmanuel Abbe, Afonso S. Bandeira, and Georgina Hall. “Exact Recovery in the Stochastic Block Model”. In: *IEEE Transactions on Information Theory* 62.1 (2016), pp. 471–487.

[3] Emmanuel Abbe and Andrea Montanari. “Conditional random fields, planted constraint satisfaction, and entropy concentration”. In: *Theory of Computing* 11 (Dec. 2015), pp. 413–443.

[4] Emmanuel Abbe et al. “Stochastic block model entropy and broadcasting on trees with survey”. In: *Proceedings of Thirty Fourth Conference on Learning Theory*. Vol. 134. PMLR, 2021, pp. 1–25.

[5] Dimitris Achlioptas, Assaf Naor, and Yuval Peres. “Rigorous location of phase transitions in hard optimization problems”. In: *Nature* 435 (7043 2005), pp. 759–764.

[6] Lada A. Adamic and Natalie Glance. “The Political Blogosphere and the 2004 U.S. Election: Divided They Blog”. In: *Association for Computing Machinery* (2005), pp. 36–43.

[7] Edoardo M. Airoldi et al. “Mixed Membership Stochastic Blockmodels”. In: *Journal of Machine Learning Research* 9.65 (2008), pp. 1981–2014.

[8] Luis Alvarez et al. “Axioms and fundamental equations of image processing”. In: *Arch. Rational Mech. Anal.* 123.3 (1993), pp. 199–257. ISSN: 0003-9527.

[9] Luigi Ambrosio and Jin Feng. “On a class of first order Hamilton-Jacobi equations in metric spaces”. In: *Journal of Differential Equations* 256.7 (2014), pp. 2194–2245.

[10] Brian Ball, Brian Karrer, and M. E. J. Newman. “Efficient and principled method for detecting communities in networks”. In: *Physical Review E* 84.3 (2011).

[11] Jean Barbier, Chun Lam Chan, and Nicolas Macris. “Mutual Information for the Stochastic Block Model by the Adaptive Interpolation Method”. In: *2019 IEEE International Symposium on Information Theory*. IEEE Press, 2019, pp. 405–409.

[12] Jean Barbier and Nicolas Macris. “The adaptive interpolation method: a simple scheme to prove replica formulas in Bayesian inference”. In: *Probability Theory and Related Fields* 174.3-4 (2019), pp. 1133–1185.

[13] Jean Barbier, Nicolas Macris, and Léo Miolane. “The layered structure of tensor estimation and its mutual information”. In: *2017 55th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2017, pp. 1056–1063.

[14] Jean Barbier and Dmitry Panchenko. “Free Energy in the Stochastic Block Model In Sparse Regime: Disassortative case”. In: *Personal communication* ().

[15] Jean Barbier and Dmitry Panchenko. "Strong replica symmetry in high-dimensional optimal Bayesian inference". In: *Communications in Mathematical Physics* 393 (2022), pp. 1199–1239.

[16] Jean Barbier et al. "Mutual information for symmetric rank-one matrix estimation: A proof of the replica formula". In: *Advances in Neural Information Processing Systems (NIPS)*. Vol. 29. 2016, pp. 424–432.

[17] M. Bardi and I. Capuzzo-Dolcetta. *Optimal Control and Viscosity Solutions of Hamilton-Jacobi-Bellman Equations*. Modern Birkhäuser Classics. Birkhäuser Boston, 2008.

[18] Béla Bollobás, Svante Janson, and Oliver Riordan. "The Phase Transition in Inhomogeneous Random Graphs". In: *Random Struct. Algorithms* 31.1 (2007), pp. 3–122.

[19] Ravi B. Boppana. "Eigenvalues and graph bisection: An average-case analysis". In: *28th Annual Symposium on Foundations of Computer Science*. 1987, pp. 280–285.

[20] Stéphane Boucheron, Gábor Lugosi, and Pascal Massart. *Concentration inequalities*. Oxford University Press, Oxford, 2013, pp. x+481.

[21] T. N. Bui et al. "Graph bisection algorithms with good average case behavior". In: *Combinatorica* 7.2 (1987), pp. 171–191.

[22] Irineo Cabreros, Emmanuel Abbe, and Aristotelis Tsirigos. "Detecting community structures in Hi-C genomic data". In: *Conference on Information Science and Systems*. 2016, pp. 584–589.

[23] Francesco Caltagirone, Marc Lelarge, and Léo Miolane. "Recovering Asymmetric Communities in the Stochastic Block Model". In: *IEEE Transactions on Network Science and Engineering* 5.3 (2018), pp. 237–246.

[24] P. Cardaliaguet and M. Quincampoix. "Deterministic differential games under probability knowledge of initial condition". In: *International Game Theory Review* 10.1 (2008), pp. 1–16.

[25] Pierre Cardaliaguet. *Notes on mean field games*. Tech. rep. 2010.

[26] Pierre Cardaliaguet and Anne Souquière. "A differential game with a blind player". In: *SIAM Journal on Control and Optimization* 50.4 (2012), pp. 2090–2116.

[27] Hong-Bin Chen. "Hamilton-Jacobi equations for nonsymmetric matrix inference". In: *arXiv e-prints*, arXiv:2006.05328 (2020). arXiv: 2006.05328.

[28] Hong-Bin Chen. "Hamilton-Jacobi equations for nonsymmetric matrix inference". In: *Ann. Appl. Probab.* 32.4 (2022), pp. 2540–2567. ISSN: 1050-5164.

[29] Hong-Bin Chen and Jiaming Xia. "Fenchel-Moreau identities on convex cones". In: *arXiv e-prints* (2020). eprint: arXiv:2011.06979.

[30] Hong-Bin Chen and Jiaming Xia. "Hamilton-Jacobi equations for inference of matrix tensor products". In: *arXiv e-prints*, arXiv:2009.01678 (2020). arXiv: 2009.01678.

[31] Hong-Bin Chen and Jiaming Xia. "Hamilton-Jacobi equations for inference of matrix tensor products". In: *Ann. Inst. Henri Poincaré Probab. Stat.* 58.2 (2022), pp. 755–793. ISSN: 0246-0203.

[32] Hong-Bin Chen and Jiaming Xia. "Hamilton-Jacobi equations from mean-field spin glasses". In: *arXiv e-prints*, arXiv:2201.12732 (2022). arXiv: 2201.12732.

[33] Hong-Bin Chen and Jiaming Xia. "Hamilton-Jacobi equations with monotone nonlinearities on convex cones". In: *arXiv e-prints*, arXiv:2206.12537 (2022). arXiv: 2206.12537.

[34] Hong-Bin Chen and Jiaming Xia. "Limiting free energy of multi-layer generalized linear models". In: *arXiv e-prints*, arXiv:2108.12615 (2021). arXiv: 2108.12615.

[35] Hong-Bin Chen and Jiaming Xia. "Limiting free energy of multi-layer generalized linear models". In: *Preprint arXiv:2108.12615* (2021).

[36] Hongbin Chen, Jean-Christophe Mourrat, and Jiaming Xia. "Statistical inference of finite-rank tensors". en. In: *Annales Henri Lebesgue* 5 (2022), pp. 1161–1189.

[37] Jingchun Chen and Bo Yuan. "Detecting functional modules in the yeast protein-protein interaction network". In: *Bioinformatics* 22.18 (2006), pp. 2283–2290.

[38] Amin Coja-Oghlan et al. "Information-theoretic thresholds from the cavity method". In: *Advances in Mathematics* 333 (2018), pp. 694–795.

[39] Michael G. Crandall, Hitoshi Ishii, and Pierre-Louis Lions. "User's guide to viscosity solutions of second order partial differential equations". In: *Bulletin of the American Mathematical Society* 27 (1992), pp. 1–67.

[40] Michael G. Crandall and Pierre-Louis Lions. "Hamilton-Jacobi equations in infinite dimensions. I. Uniqueness of viscosity solutions". In: *Journal of Functional Analysis* 62.3 (1985), pp. 379–396.

[41] Michael G. Crandall and Pierre-Louis Lions. "Hamilton-Jacobi equations in infinite dimensions. II. Existence of viscosity solutions". In: *Journal of Functional Analysis* 65.3 (1986), pp. 368–405.

[42] Michael G. Crandall and Pierre-Louis Lions. "Hamilton-Jacobi equations in infinite dimensions. III". In: *Journal of Functional Analysis* 68.2 (1986), pp. 214–247.

[43] Michael G. Crandall and Richard Newcomb. "Viscosity Solutions of Hamilton-Jacobi Equations at the Boundary". In: *Proceedings of the American Mathematical Society* 94.2 (1985), pp. 283–290.

[44] Aurelien Decelle et al. "Asymptotic analysis of the stochastic block model for modular networks and its algorithmic applications". In: *Physical Review E* 84 (2011), p. 066106.

[45] Yash Deshpande, Emmanuel Abbe, and Andrea Montanari. "Asymptotic mutual information for the balanced binary stochastic block model". In: *Information and Inference: A Journal of the IMA* 6.2 (2016), pp. 125–170.

[46] Yash Deshpande and Andrea Montanari. "Information-theoretically optimal sparse PCA". In: *IEEE International Symposium on Information Theory*. 2014, pp. 2197–2201.

[47] Tomas Dominguez. "The ℓ^p-Gaussian-Grothendieck problem with vector spins". In: *Electronic Journal of Probability* 27 (2022), pp. 1–46.

[48] Tomas Dominguez and Jean-Christophe Mourrat. "Infinite-dimensional Hamilton-Jacobi Equations for Statistical Inference on Sparse Graphs". In: *arXiv e-prints*, 2209.04516 (2022). arXiv: arXiv:2209.04516.

[49] Tomas Dominguez and Jean-Christophe Mourrat. "Mutual information for the sparse stochastic block model". In: *arXiv e-prints*, 2209.04513 (2022). arXiv: arXiv:2209.04513.

[50] Tomas Dominguez and Jean-Christophe Mourrat. "Statistical mechanics of mean-field disordered systems — A Hamilton-Jacobi approach". In: *arXiv e-prints*, 2311.08976 (2023). arXiv: arXiv:2311.08976.

[51] M.E Dyer and A.M Frieze. "The solution of some random NP-hard problems in polynomial expected time". In: *Journal of Algorithms* 10.4 (1989), pp. 451–489.

[52] Ahmed El Alaoui and Florent Krzakala. "Estimation in the spiked Wigner model: a short proof of the replica formula". In: *2018 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2018, pp. 1874–1878.

[53] L.C. Evans. *Partial Differential Equations*. Graduate studies in mathematics. American Mathematical Society, 2010. ISBN: 9780821849743.

[54] Jin Feng and Markos Katsoulakis. "A comparison principle for Hamilton-Jacobi equations related to controlled gradient flows in infinite dimensions". In: *Archive for Rational Mechanics and Analysis* 192.2 (2009), pp. 275–310.

[55] Jin Feng and Thomas G. Kurtz. *Large deviations for stochastic processes*. Vol. 131. Mathematical Surveys and Monographs. American Mathematical Society, Providence, RI, 2006, pp. xii+410.

[56] Stephen E. Fienberg, Michael M. Meyer, and Stanley S. Wasserman. "Statistical Analysis of Multiple Sociometric Relations". In: *Journal of the American Statistical Association* 80.389 (1985), pp. 51–67.

[57] G.B. Folland. *Real Analysis: Modern Techniques and Their Applications*. Pure and Applied Mathematics: A Wiley Series of Texts, Monographs and Tracts. Wiley, 1999.

[58] Wilfrid Gangbo, Truyen Nguyen, and Adrian Tudorascu. "Hamilton-Jacobi equations in the Wasserstein space". In: *Methods and Applications of Analysis* 15.2 (2008), pp. 155–183.

[59] Wilfrid Gangbo and Andrzej Swicech. "Optimal transport and large number of particles". In: *Discrete and Continuous Dynamical Systems. Series A* 34.4 (2014), pp. 1397–1441.

[60] Prem K Gopalan and David M Blei. "Efficient discovery of overlapping communities in massive networks". In: *Proc. Natl. Acad. Sci. USA* 110.36 (2013), pp. 14534–14539.

[61] Yuzhou Gu and Yury Polyanskiy. "Uniqueness of BP fixed point for the Potts model and applications to community detection". In: (Preprint, arXiv:2303.14688).

[62] J.B. Hiriart-Urruty and C. Lemaréchal. *Fundamentals of Convex Analysis*. Grundlehren Text Editions. Springer Berlin Heidelberg, 2004. ISBN: 9783540422051.

[63] Paul W. Holland, Kathryn Blackmond Laskey, and Samuel Leinhardt. "Stochastic blockmodels: First steps". In: *Social Networks* 5.2 (1983), pp. 109–137.

[64] Roger A. Horn and Charles R. Johnson. *Topics in Matrix Analysis*. Cambridge University Press, 1991.

[65] Jonathan Kadmon and Surya Ganguli. "Statistical mechanics of low-rank tensor decomposition". In: *Advances in Neural Information Processing Systems*. 2018, pp. 8201–8212.

[66] Varun Kanade, Elchanan Mossel, and Tselil Schramm. "Global and local information in clustering labeled block models". In: *IEEE Transactions on Information Theory* 62.10 (2016), pp. 5906–5917.

[67] Anastasia Kireeva and Jean-Christophe Mourrat. "Breakdown of a concavity property of mutual information for non-Gaussian channels". In: (Preprint, arXiv:2304.05129).

[68] Marc Lelarge and Léo Miolane. "Fundamental limits of symmetric low-rank matrix estimation". In: *Probability Theory and Related Fields* 173.3 (2019), pp. 859–929.

[69] Thibault Lesieur, Florent Krzakala, and Lenka Zdeborová. "Phase transitions in sparse PCA". In: *IEEE International Symposium on Information Theory*. 2015, pp. 1635–1639.

[70] Thibault Lesieur et al. "Statistical and computational phase transitions in spiked tensor estimation". In: *2017 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2017, pp. 511–515.

[71] Greg Linden, Brent Smith, and Jeremy York. "Amazon.com Recommendations: Item-to-Item Collaborative Filtering". In: *IEEE Internet Comput.* 7 (2003), pp. 76–80.

[72] Clément Luneau, Jean Barbier, and Nicolas Macris. "Mutual information for low-rank even-order symmetric tensor estimation". In: *Information and Inference: A Journal of the IMA* 10.4 (2021), pp. 1167–1207.

[73] Clément Luneau, Nicolas Macris, and Jean Barbier. "High-dimensional rank-one nonsymmetric matrix decomposition: the spherical case". In: *2020 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2020, pp. 2646–2651.

[74] Laurent Massoulié. "Community Detection Thresholds and the Weak Ramanujan Property". In: *Proceedings of the Forty-Sixth Annual ACM Symposium on Theory of Computing*. Association for Computing Machinery, 2014, pp. 694–703.

[75] Vaishakhi Mayya and Galen Reeves. "Mutual information in community detection with covariate information and correlated networks". In: *2019 57th Annual Allerton Conference on Communication, Control, and Computing (Allerton)*. IEEE. 2019, pp. 602–607.

[76] M Mézard, G Parisi, and R Zecchina. "Analytic and algorithmic solution of random satisfiability problems". In: *Science (New York, N.Y.)* 297.5582 (Aug. 2002), pp. 812–815. ISSN: 0036-8075.

[77] Léo Miolane. "Fundamental limits of low-rank matrix estimation: the non-symmetric case". In: *Preprint, arXiv:1702.00473* (2017).

[78] Elchanan Mossel, Joe Neeman, and Allan Sly. "A Proof of the Block Model Threshold Conjecture". In: *Combinatorica* 38.3 (2018), pp. 665–708.

[79] Elchanan Mossel, Joe Neeman, and Allan Sly. "Belief propagation, robust reconstruction and optimal recovery of block models". In: *The Annals of Applied Probability* 26.4 (2016), pp. 2211–2256.

[80] Elchanan Mossel, Joe Neeman, and Allan Sly. "Consistency thresholds for the planted bisection model". In: *Electronic Journal of Probability* 21 (2016), pp. 1–24.

[81] Elchanan Mossel, Joe Neeman, and Allan Sly. "Reconstruction and estimation in the planted partition model". In: *Probab. Theory Related Fields* 162.3-4 (2015), pp. 431–461. ISSN: 0178-8051.

[82] Elchanan Mossel and Jiaming Xu. "Local algorithms for block models with side information". In: *Proceedings of the 2016 ACM Conference on Innovations in Theoretical Computer Science*. 2016, pp. 71–80.

[83] Jean-Christophe Mourrat. "Free energy upper bound for mean-field vector spin glasses". In: *Annales de l'Institut Henri Poincaré, Probabilités et Statistiques* 59.3 (2023), pp. 1143–1182.

[84] Jean-Christophe Mourrat. "HamiltonJacobi equations for finite-rank matrix inference". In: *The Annals of Applied Probability* 30.5 (2020), pp. 2234–2260.

[85] Jean-Christophe Mourrat. "HamiltonJacobi equations for mean-field disordered systems". In: *Annales Henri Lebesgue* 4 (2021), pp. 453–484.

[86] Jean-Christophe Mourrat. "Nonconvex interactions in mean-field spin glasses". In: *Probability and Mathematical Physics* 2.2 (2021), pp. 61–119.

[87] Jean-Christophe Mourrat. “The Parisi formula is a Hamilton-Jacobi equation in Wasserstein space”. In: *Canadian Journal of Mathematics* 74 (Jan. 2021), pp. 1–23.

[88] Jean-Christophe Mourrat and Dmitry Panchenko. “Extending the Parisi formula along a Hamilton-Jacobi equation”. In: *Electronic Journal of Probability* 25.none (2020), pp. 1–17.

[89] M E J Newman, D J Watts, and S H Strogatz. “Random graph models of social networks”. In: *Proc. Natl. Acad. Sci. USA* 99 (2002), pp. 2566–2572.

[90] Dmitry Panchenko. *Lecture notes on probability theory*. 2019.

[91] Dmitry Panchenko. “On the replica symmetric solution of the K-sat model”. In: *Electronic Journal of Probability* 19.67 (2014), pp. 1–17.

[92] Dmitry Panchenko. “Spin glass models from the point of view of spin distributions”. In: *The Annals of Probability* 41.3A (2013), pp. 1315–1361.

[93] Dmitry Panchenko. “Structure of 1-RSB Asymptotic Gibbs Measures in the Diluted p-Spin Models”. In: *Journal of Statistical Physics* 155.1 (2014), pp. 1–22.

[94] Dmitry Panchenko. “Structure of Finite-RSB Asymptotic Gibbs Measures in the Diluted Spin Glass Models”. In: *Journal of Statistical Physics* 162.1 (2015), pp. 1–42.

[95] Dmitry Panchenko. *The Sherrington-Kirkpatrick model*. Springer Monographs in Mathematics. Springer, New York, 2013, pp. xii+156. ISBN: 978-1-4614-6288-0; 978-1-4614-6289-7.

[96] Raj G. Patel and Tomas Dominguez et al. “Application of Tensor Neural Networks to Pricing Bermudan Swaptions”. In: (2023). arXiv: 2304.09750 [q-fin.CP].

[97] Raj G. Patel and Tomas Dominguez et al. “Quantum-Inspired Tensor Neural Networks for Option Pricing”. In: (2022). arXiv: 2212.14076 [q-fin.PR].

[98] Galen Reeves. “Information-theoretic limits for the matrix tensor product”. In: *IEEE Journal on Selected Areas in Information Theory* 1.3 (2020), pp. 777–798.

[99] Galen Reeves, Vaishakhi Mayya, and Alexander Volfovsky. “The geometry of community detection via the mmse matrix”. In: *2019 IEEE International Symposium on Information Theory (ISIT)*. IEEE. 2019, pp. 400–404.

[100] Tom Richardson and Rüdiger Urbanke. “An Introduction to the Analysis of Iterative Coding Systems”. In: *Codes, Systems, and Graphical Models* (2001), pp. 1–37.

[101] C. E. Shannon. “A mathematical theory of communication”. In: *The Bell System Technical Journal* 27.3 (1948), pp. 379–423.

[102] Jianbo Shi and Jitendra Malik. “Normalized Cuts and Image Segmentation”. In: *IEEE Trans. Pattern Anal. Mach. Intell.* 22.8 (2000), pp. 888–905.

[103] T Sørlie, C M Perou, et al. “Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications”. In: *Proc. Natl. Acad. Sci. USA* 98.19 (2001), pp. 10869–10874.

[104] P. E. Souganidis. “A Remark About Viscosity Solutions of Hamilton-Jacobi Equations at the Boundary”. In: *Proceedings of the American Mathematical Society* 96.2 (1986), pp. 323–329.

[105] Yuchung J. Wang and George Y. Wong. “Stochastic Blockmodels for Directed Graphs”. In: *Journal of the American Statistical Association* 82.397 (1987), pp. 8–19.

[106] Harrison C. White, Scott A. Boorman, and Ronald L. Breiger. “Social Structure from Multiple Networks. I. Blockmodels of Roles and Positions”. In: *American Journal of Sociology* 81.4 (1976), pp. 730–780.

[107] Qian Yu and Yury Polyanskiy. “Ising Model on Locally Tree-like Graphs: Uniqueness of Solutions to Cavity Equations”. In: (Preprint, arXiv:2211.15242).