UNIVERSITA' COMMERCIALE "LUIGI BOCCONI"

PhD SCHOOL

PhD program in Statistics

Cycle: XXXII

Disciplinary Field (code): INF/01

Online Learning, Physics and Algorithms

Advisor: Riccardo ZECCHINA

Co-Advisor: Nicolò CESA-BIANCHI

PhD Thesis by

Riccardo DELLA VECCHIA

ID number: 3031356

Year 2021

Contents

List of Figures

Acknowledgements

First of all, I would like to thank my Ph.D. advisors Riccardo Zecchina and Nicolò Cesa-Bianchi for introducing me to machine learning and giving me the possibility to contribute to scientific research in this vast field. I am also grateful to Carlo Lucibello and Carlo Baldassi for their help in the part of my research that was more closely related to physics, for their insightful comments and suggestions. I also have deep gratitude towards Nicolò Cesa-Bianchi, for the kind hospitality in the Computer Science Department of Milano University, where I learned so much of online learning from him and Tom Cesari. I thank Nicolò also for all the fantastic collaborators he introduced me to. Thanks to my colleagues in BNP and the ARTLab at Bocconi for your time, support and laughs of these years.

Finally, this thesis wouldn't have been possible without my beloved ones' support in the ups and downs of these years. A special thank you to my family.

Abstract

In recent years, we have witnessed an increasing cross-fertilization between the fields of computer science, statistics, optimization and the statistical physics of learning. The area of machine learning is at the interface of these subjects. We start with an analysis in the statistical physics of learning, where we analyze some properties of the loss landscape of simple models of neural networks using the computer science formalism of Constraint Satisfaction Problems. Some of the techniques we employ are probabilistic, but others have their root in the studies of *disorder systems* in the statistical physics literature. After that, we focus mainly on *online prediction* problems, which were initially investigated in statistics but are now very active areas of research also in computer science and optimization, where they are studied in the adversarial case through the lens of *(online) convex optimization*. We are particularly interested in the *cooperative* setting, where we show that cooperation improves learning. More specifically, we give efficient algorithms and unify previous works under a simplified and more general framework.

Chapter 1

Introduction

1.1 Statistical Physics, Algorithms and Neural Networks

Problems with many degrees of freedom (*variables*) but also many constraints are ubiquitous in science. Most of the times, the problem is to find a value of the variables which satisfies all constraints, or the most probable configuration of variables given the constraints and some a priori measure. Such problems occur in various branches of scientific research and are crucial in several domains. The satisfiability problem is also at the core of the theory of computational complexity in computer science. Error correcting codes is one of the main topics of information theory. Learning from examples is an essential process in cognitive neuroscience. Reconstruction of neuron interactions from multi-electrode recording is a problem which is becoming more and more critical. All these problems can be formulated in a common language [Mezard and Montanari, 2009], and have a strong relationship to fundamental issues in statistical physics like the existence of phase transition, and the possibility of glassy phases. They can also be cast into a somewhat generic formalism, based a graphical representation of the topology of constraints [Kschischang et al., 2001], which allows applying a general *message passing* strategy to all of them. Some of these message passing algorithms have shown strikingly good performance, solving some problems in satisfiability or perceptron learning that are unreachable by any other algorithms.

Formally, a generic *Constraint Satisfaction Problem* (CSP) can be defined in terms of configurations of N variables $x_i \in X_i$, subject to M constraints $\psi_\mu : D_\mu \to \{0,1\}$. Each constraint μ involves a subset $\partial\mu$ of the variables, which we collectively represent as $x_{\partial\mu} = \{x_i : i \in \partial\mu\} \in D_\mu$, and we define $\psi_\mu\left(x_{\partial\mu}\right) = 1$ if the constraint is satisfied, 0 otherwise. For the case of binary spin variables, one has $X_i = X = \{-1,+1\}$. It is possible to define an energy function of the system simply as the number of violated constraints, namely:

$$E(x) = \sum_\mu E_\mu\left(x_{\partial\mu}\right) = \sum_\mu \left(1 - \psi_\mu\left(x_{\partial\mu}\right)\right) \ ,$$

where a solution of a CSP is then a zero-energy configuration. Statistical Physics provides (among others) tools to state whether one can find solutions to the problem at a given constraint density, $\alpha = M/N$, where the thermodynamic limit $M, N \to \infty$ is taken. Interestingly, in random CSPs, the system undergoes a sharp transition at a critical

density α_C, going from a phase where exponentially many zero-energy configurations are present with very high probability, the so-called SAT phase, to a phase where the problem is no longer satisfiable, and at best a small fraction of the constraints will be necessarily violated, the UNSAT phase. This will be the topic of Chapter 2 of this thesis, where the model that we will study is a simple example of neural networks.

Artificial neural networks (NNs) are among the most widely used tools in data science and have exceptional performances in complex recognition tasks [LeCun et al., 2015]. The remarkable output of these systems has paved the way for important opportunities for machine learning in a vast number of applications. In NNs with simple as well as large and complex architectures, learning from data is a computationally demanding task in which a large number of connection weights are iteratively tuned through heuristic improvements over the basic stochastic gradient descent (SGD) algorithm. In practical applications, NNs are trained on big datasets, and the aforementioned heuristic algorithms often find solutions with good generalization properties. How learning, despite huge numbers of parameters and strong nonlinearities, occurs in these systems, without getting trapped in configurations corresponding to local minima with poor prediction performance, is not well understood. However, theoretical progress could be useful to shape NNs architecture and new learning algorithms.

In the past decades, various methods borrowed from statistical physics have been successful in studying the fundamental properties of neural-like systems [Advani et al., 2013]. From this viewpoint, data under the form of training examples plays the role of quenched disorder, and the synaptic weights of the network (the learning parameters of the machine learning algorithm), play the role of statistical mechanical degrees of freedom. In the zero-temperature limit, these degrees of freedom are optimized, or learned, by minimizing an energy function, just as the ground state is usually uncovered in the context of physics. Indeed, many machine learning algorithms can be formulated as the minimization of a (possibly very rough) data dependent energy function on a high-dimensional space of parameters - a topic that has been very much studied in the statistical physics of disordered systems and in particular of spin glasses [Mézard et al., 1987]. The techniques developed in the context of statistical physics have been successfully applied also to study the geometry of the solution space and corresponding phase transitions in random constraint satisfaction problems (CSPs) [Mezard and Montanari, 2009, Krzaka…a et al., 2007], inspiring new powerful heuristic algorithms that can find solutions to these problems in the hard phase close to the SAT/UNSAT threshold [Mézard et al., 2002, Braunstein et al., 2005].

In a recent series of papers Baldassi et al. [2015, 2016b,a, 2020b], Pittorino et al. [2020] propose to study artificial NNs performance in terms of a non-equilibrium statistical physics framework. They perform a large deviation analysis to shed light on the inner algorithmic working of models of artificial NNs and deep architectures. Theoretical and numerical evidence shows that the large deviation measure - named *local entropy* - reveals the existence of a class of exponentially rare solutions in the optimization landscape (the network weight space) which have surprisingly good generalization properties. These solutions are clustered in dense regions, in the case of the discrete weights, and have radically different properties from the ones dominating an equilibrium measure [Baldassi et al., 2015]. These exponentially rare minima are hard to find for most algorithms, but they become very attractive for others that are properly

designed. The same picture holds for complex NNs architectures with continuous weights trained on real-world benchmarks [Chaudhari et al., 2019, Pittorino et al., 2020], where flat minimizers are the equivalent for the continuous weight space of the regions of dense solutions. In fact, already Sagun et al. [2016] found that the spectrum of the Hessian of the loss function is composed of two parts: the bulk centred near zero, and outliers away from the bulk, confirming the hypothesis that good minimizer of the loss landscape for NNs are minima that are surrounded by an exponential number of other good ones.

The large deviation analysis of the local entropy measure and the out-of-equilibrium measure named *Robust Ensemble* (RE) were introduced by Baldassi et al. [2016a], and provide a framework in which it is possible to interpret and understand the shortcomings of the standard equilibrium analysis and, by a better understanding of the geometrical structure of the solutions space of NNs, to design new effective algorithms for learning in these systems. A general theory for these robust minima can be constructed moving from the *Gibbs Measure* used in the standard statistical mechanics formulation, i.e.

$$p\left(x\right)=\frac{1}{Z(\beta)}\,e^{-\beta E(x)}\quad\text{with}\quad Z(\beta)=\sum_{\{x\}}e^{-\beta E(x)}\,, \tag{1.1}$$

to a measure that suppresses the role of narrow local minima and favors these rare but very dense regions, i.e.

$$P\left(x;\beta,y,\lambda\right)=\frac{1}{Z\left(\beta,y,\lambda\right)}\,e^{y\,\Phi(x,\beta,\lambda)} \tag{1.2}$$

where $\Phi\left(x,\beta,\lambda\right)$ is called the *local free entropy* and y has the role of inverse temperature. Φ involves both the energy term and a distance constraint between configurations enforced by a Lagrange multiplier λ:

$$\Phi\left(x,\beta,\lambda\right)=\ln\sum_{\{x'\}}e^{-\beta E(x')-\lambda\,d(x,x')}\,, \tag{1.3}$$

where $d(\cdot,\cdot)$ is a distance measure between two configurations. In this setting, not only low-energy but also dense regions of configurations with high local entropy are favoured. The parameter λ controls how narrow/wide a minimum needs to be to have a significant statistical weight. At this stage, a direct estimation of the local free entropy Φ is not an easy task. However, by choosing y to be a non-negative integer, one can rewrite the partition function as:

$$Z\left(\beta,y,\lambda\right)=\sum_{\{x^*\}}e^{y\,\Phi(x^*,\beta,\lambda)} \tag{1.4}$$

$$=\sum_{\{x^*\}}\sum_{\{x^a\}}e^{-\beta\sum_{a=1}^{y}E(x^a)-\lambda\,\sum_{a=1}^{y}d(x^*,x^a)}. \tag{1.5}$$

This form of Z can be interpreted as describing a system of $y+1$ interacting replicas of the original system, one of which acts as reference x^*, and the other $\{x^a\}_{a=1}^{y}$ are subject to the energy $E\left(x\right)$ and to the interaction with the reference. This formulation is also particularly apt to the construction of algorithms which are tuned to explore robust regions of the energy landscape. By simply replicating the model and adding an elastic

interaction term between replicas, one can induce the dynamics to converge on local entropy minima.

To demonstrate the utility of the local entropy measure for the implementation of new efficient algorithms, Baldassi et al. [2016a] also introduce a fast Entropy-driven Monte Carlo (EdMC) strategy relying exclusively on a local entropy estimate to sample solutions of general random CSPs efficiently. Furthermore, the RE is applied to Markov Chain Monte Carlo (MCMC), message passing and gradient descent algorithms, targeting the robust dense states and resulting in improvements in their performance. Replicated gradient descent is related to elastic averaged stochastic gradient descent, used in complex deep artificial NNs [Zhang et al., 2015], implying that the geometrical structure of the RE may provide an explanation for its effectiveness and a framework for further research on learning in NNs as it has been recently confirmed in [Pittorino et al., 2020].

Despite all the progress in algorithmic performances of NNs, many questions are still open about their capabilities of generalizing well and yet little has been rigorously proved in this regard. What has been observed, is that minimizers' flatness consistently correlates with good generalization, but there has been little rigorous work in exploring the condition of existence of such minimizers, even in toy models. In Chapter 2 of this thesis, we investigate with rigorous probabilistic methods a simple neural network model, the symmetric perceptron, with binary weights. We perform the first steps toward the rigorous proof of the existence of a dense cluster in certain regimes of the parameters, by computing the first and second moment upper bounds for the existence of pairs of arbitrarily close solutions. Moreover, we present a non rigorous derivation of the same bounds for sets of y solutions at fixed pairwise distances using some quite sophisticated techniques from the theory of disordered systems. In particular, we make use of the replica method for which we refer the interested reader to the monographs by Mézard et al. [1987] and Mezard and Montanari [2009], since a full exposition of the method is out of the scope of this thesis.

1.2 Online Learning

Online learning is the process of answering a sequence of questions given (maybe partial) knowledge of the correct answers to previous questions and possibly additional available information. This setting differs from *batch learning* techniques in which the entire training data set is processed all at the same time [Shalev-Shwartz and Ben-David, 2014]. The study of online learning algorithms constitutes an essential domain in machine learning, and it has interesting theoretical properties and practical applications. Many reviews are available on the topic. The interested reader can look for example at [Orabona, 2019, Hazan, 2019, Bubeck et al., 2012], to name a few. In particular, online learning refers to the framework of regret minimization under worst-case assumptions, and since many problems in it are particularly well fitted to be studied through the lens of mathematical optimization, the field is also referred to as *Online Convex Optimization* (OCO).

Formally, OCO can be seen as a game between a player and an environment, which happens through a sequence of consecutive rounds. At each round t the learner/player

has to choose a prediction in the convex set $\mathcal{X} \subseteq \mathbb{R}^k$, that is called *decision set*. For all $t = 1, 2, \ldots$ a *loss function* $\ell_t(\cdot)\colon \mathcal{X} \to [0, 1]$ is chosen by the environment (possibly in an adversary way). The learner makes a prediction $x_t \in \mathcal{X}$, suffers a loss $\ell_t(x_t)$ and receives some feedback. The learner's goal is to minimize the *regret*, defined for any *time horizon* T by

$$R_T = \sup_{x \in \mathcal{X}} R_T(x) \qquad \text{where} \qquad R_T(x) = \sum_{t=1}^{T} \ell_t(x_t) - \sum_{t=1}^{T} \ell_t(x).$$

Within the regret framework, one can analyze situations in which the data are not independent and identically distributed from a probability distribution but is possible to guarantee that the algorithm is "learning" something. For example, online learning is used to analyze click prediction problems, routing on a network, convergence to equilibrium of repeated games. It can also be used to analyze stochastic algorithms, e.g., Stochastic Gradient Descent, but the adversarial nature of the analysis might give suboptimal results.

A particularly important case of OCO is the problem of learning with *expert advice*. In this setting, the decision set is the probability simplex over a finite set of elements typically referred to as experts. By defining a loss for each expert, we can define the loss of all these distributions x as the expectation of the loss of a random expert (drawn according to x). This setting is very important since, in many real-life applications (weather forecast, stock-price prediction, etc.), the options are indeed limited to a finite set, and after following the advice of an expert, it is possible to measure how good the advice of the other experts really was respect to the best expert in hindsight. This corresponds to the full-feedback setting.

In the case of *multi-armed bandits* (MAB) instead, the feedback does not comprise the outcome of all possible actions. The two most important types of partial feedback are bandit feedback and semi-bandit feedback. In the former case, the agent just receives a feedback f_t which corresponds to the loss $\ell_t(x_t)$ that he pays when playing x_t. In the second case instead, the loss is a linear function, and the agent receives as feedback the single components of the vector $f_t = \big(x_t(1)\ell_t(1), \ldots, x_t(k)\ell_t(k)\big)$ for $x_t \in \{0, 1\}^k$ while paying a loss $\langle \ell_t, x_t \rangle$. This is a fundamental paradigm in online learning and has seen exponential growth in publications over the last decades. The name comes from imagining a gambler at a row of slot machines (sometimes known as "one-armed bandits"), who has to decide which machines to play, how many times to play each machine and in which order to play them, and whether to continue with the current machine or try a different machine. Despite the name, the original motivation for this model was indeed different, and it comes from clinical trials. In clinical trials, each arm corresponds to one treatment, and rewards measure the efficacy of this treatment on a patient [Thompson, 1933]. Even though medical studies motivated the initial research in multi-armed bandits, it is a field where researchers have not consistently employed bandits for their analyses.

Multi-armed bandit problem is interesting for machine learning because it presents the so-called exploration versus exploitation dilemma. Indeed, there is a clear tradeoff between discovering which treatment is the most effective (exploration) and administering the best treatment to as many patients as possible (exploitation). On many aspects,

this problem represents the "hydrogen atom" of reinforcement learning, in the sense that it is the basic model from which one can get insights before tackling more complicated problems. For example, many algorithms for reinforcement learning and partial monitoring have their roots in the bandit setting [Auer et al., 2009, Bartók, 2013]. For further reading on different bandit models, we refer to the recent book by Lattimore and Szepesvári [2018], where the authors also point out to the many applications in which bandits already play a fundamental role today. For example, they have a prominent role in online advertising, where customers arrive sequentially at a fast rate. Big tech companies also use bandit algorithms to optimize user interfaces, provide personalized news or content, and much more Li et al. [2010], Chapelle et al. [2014], Kveton et al. [2015]. Furthermore, Monte-Carlo Tree Search, which also uses bandit theory, is a crucial component of the algorithm that has defeated the world champion of Go in 2016 [Silver et al., 2016], reaching a goal that many believed to be at least a decade away.

In a world where distributed systems are ubiquitous, is worth investigating how these online convex optimization techniques behave in a *cooperative framework*. Cooperation is useful for many problems in large-scale learning systems in finance, online advertising, wireless sensor networks, climate informatics, where such an approach has shown empirical performance advantages compared to the global (i.e., non-spatially distributed) online learning counterparts. The goal in a cooperative setting is to minimize the regret in a communication network, which is modelled as a graph $\mathcal{G}$. After T time steps the regret is

$$R_T = \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \ell_t(x_t(v)) - \inf_{x \in \mathcal{X}} \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \ell_t(x)\,. \tag{1.6}$$

where $\mathcal{S}_t$ is a stochastic set of "active" agents v that made a prediction at time t. Similarly to the single agent case, this is the difference between the cumulative loss of the active agents and the loss that they would have incurred had they consistently made the best prediction in hindsight. In this framework, the protocols might differ. In Chapter 3, we focus on the efficiency of the algorithms that are used, while, in Chapter 4, the focus is on cooperation when the feedback is broadcast through the network. In this last setting, it is natural the appearance of delays since agents have a certain spatial distance, which is given by shortest-path distance on the communication graph.

1.3 Outline of the thesis

The thesis is structured in the following way.

Chapter 2 is based on [Baldassi et al., 2020a]. Motivated by the good generalization properties of flat minimizers in deep NNs, we consider the symmetric perceptron with binary weights and study the existence of this type of solutions. We phrase the learning problem as a constraint satisfaction problem, where the analogous of a flat minimizer becomes a large and dense cluster of solutions, while the narrowest minimizers are isolated solutions. We perform the first steps toward the rigorous proof of the existence of a dense cluster in certain regimes of the parameters, by computing the first and second moment upper bounds for the existence of pairs of arbitrarily close solutions. Moreover, we present a non rigorous derivation of the same bounds for sets of y solutions at fixed pairwise distances.

Chapter 3 is based on [Della Vecchia and Cesari, 2020]. In this chapter we investigate how online convex optimization techniques behave in a cooperative framework for a combinatorial action set and under semi-bandit feedback. Furthermore, at each time step, just some of the agents are stochastically activated and requested to make a prediction. These are the agents that participate in the total loss of the system. Then, neighbors of active agents receive semi-bandit feedback and exchange some succinct local information. As usual, the goal is to minimize the network regret. Interestingly, the main challenge in such a context is to control the computational complexity of the resulting algorithm while retaining minimax optimal regret guarantees. We introduce Coop-FTPL, a cooperative version of the well-known Follow The Perturbed Leader algorithm, that implements a new loss estimation procedure that we call Coop-GR.

Chapter 4 is based on [Cesa-Bianchi et al.]. In this chapter we introduce and analyze an online learning setting in which a network of agents solves a common online convex optimization problem, in the full and partial feedback setting (bandit and semi-bandit), by sharing feedback with their network neighbours. Such shared feedback is broadcast through the network and we study its impact on the global performance of the agents. We study the problem under two types of feedback, under the full-information feedback we study the family of algorithm of Online Mirror Descent (OMD), but, since this doesn't directly give the interesting case of Hedge we resort to a specific analysis for it that makes use of the update of Follow The Regularized Leader (FTRL). The other important case is partial information feedback. We study the case of a network of agents that cooperate to solve the same nonstochastic bandit problem, and we extend the analysis also to the case of semi-bandits on m-sets. In Section 4.4 we present the main novelty of our paper, which is an algorithm and an analysis that lets one transform a general algorithm that plays with delays into an algorithm on the communication network and retains a neat study for the total regret.

Chapter 2

Clustering of solutions in the symmetric binary perceptron

2.1 Abstract

The geometrical features of the (non-convex) loss landscape of neural network models are crucial in ensuring successful optimization and, most importantly, the capability to generalize well. While minimizers' flatness consistently correlates with good generalization, there has been little rigorous work in exploring the condition of existence of such minimizers, even in toy models. Here we consider a simple neural network model, the symmetric perceptron, with binary weights. Phrasing the learning problem as a constraint satisfaction problem, the analogous of a flat minimizer becomes a large and dense cluster of solutions, while the narrowest minimizers are isolated solutions. We perform the first steps toward the rigorous proof of the existence of a dense cluster in certain regimes of the parameters, by computing the first and second moment upper bounds for the existence of pairs of arbitrarily close solutions. Moreover, we present a non rigorous derivation of the same bounds for sets of y solutions at fixed pairwise distances.

2.2 Dense clusters in the symmetric binary perceptron

The problem of learning to classify a set random patterns with a *binary perceptron* has been a recurrent topic since the very beginning of the statistical physics studies of neural networks models [Gardner and Derrida, 1988]. The learning problem consists in finding the optimal binary assignments of the connection weights which minimize the number of misclassifications of the patterns. We shall refer to such set of optimal assignments as the space of solutions of the perceptron. In spite of the extremely simple architecture of the model, the learning task is highly non convex and its geometrical features are believed to play a role also in more complex neural architectures [Watkin et al., 1993, Seung et al., 1992, Engel and Van den Broeck, 2001].

For the case of random i.i.d. patterns, the space of solutions of the binary perceptron is known to be dominated by an exponential number of isolated solutions [Krauth and Mézard, 1989] which lie at a large mutual Hamming distances [Huang et al., 2013,

Huang and Kabashima, 2014] (golf course landscape). An even larger number of local minima have been shown to exist [Horner, 1992].

The study of how the number of these isolated solutions decreases as more patterns are learned provides the correct prediction for the so-called capacity of the binary perceptron, i.e. the maximum number of random patterns that can be correctly classified. However, the same analysis does not provide the insight necessary for understanding the behavior of learning algorithms: one would expect that finding solutions in a golf course landscape should be difficult for search algorithms, and indeed Monte Carlo based algorithms satisfying detailed balance get stuck in local minima; yet, empirical results have shown that many learning algorithms, even simple ones, are able to find solutions efficiently [Braunstein and Zecchina, 2006, Baldassi et al., 2007, Baldassi, 2009, Baldassi and Braunstein, 2015].

These empirical results suggested that the solutions which were not the dominant ones in the Gibbs measure, and were as such neglected in the analysis of the capacity, could in fact play an important algorithmic role. As discussed in refs. [Baldassi et al., 2015, 2016b] this turned out to be the actual case: the study of the dominant solutions in the Gibbs measure theory does not take into account the existence of rare (sub-dominant) regions in the solution space which are those found by algorithms. Revealing those rare, accessible regions required a large deviation analysis based on the notion of *local entropy*, which is a measure of the density of solutions in an extensive region of the configuration space (see the precise definition in the next section). The regions of maximal local entropy are extremely dense in solutions, such that (for finite N) nearly every configuration in the region is a solution. More recently, the existence of high local entropy / flat regions has been found also in multi-layer networks with continuous weights, and their role has been connected to the structural characteristics of deep neural networks [Baldassi et al., 2019, 2020b].

All the above results rely on methods of statistical mechanics of disordered systems which are extremely powerful and yet not fully rigorous. It is therefore important to corroborate them with rigorous bounds [Ding and Sun, 2019]. In a recent paper [Aubin et al., 2019], Aubin et al. have studied a simple variant of the binary perceptron model for which the rigorous bounds provided by first and second moment methods can be shown to be tight. The authors have been able to confirm the predictions of the statistical physics methods concerning the capacity of the model, and the golf course nature of the space of solutions. The model that the authors have studied has a modified activation criterion compared to the traditional perceptron, replacing the Heaviside step function by a function with an even symmetry.

The goal of the present paper is to study the existence of dense regions in the the symmetrized binary perceptron model. In sec. 2.2 we define the model and, as a preliminary step, we present the results of the replica-method large deviation analysis, which predicts that the phenomenology for the symmetrized model is the same as for the traditional one, and thus that high local entropy regions exist. If these predictions are correct, then it should be possible, at least for some range of the parameters, to choose any integer number $y \geq 2$ and find a threshold $x_c(y)$ such that for any $x < x_c(y)$ there is an exponential number of groups of y solutions all at mutual Hamming distance $\lfloor Nx \rfloor$. In the remainder of the paper we try to verify this statement, by employing the first and second moment methods where possible. In sec. 2.3 we address the $y = 2$

case: we extend the analysis of ref. [Aubin et al., 2019] and show rigorously (except for a numerical optimization step) that, for small enough constraint density α, there exist an exponential number of pairs of solutions at arbitrary $O(N)$ Hamming distance. In sec. 2.4 we study the general y case. For $y = 3$ or 4, we can derive a rigorous upper bound that coincides with the non-rigorous results for general y. As for the lower bound, only the $y = 2$ case can be derived rigorously (and again it coincides with the non-rigorous results that we also derive). All the results are thus consistent with the existence of high local entropy regions, as predicted by the large deviation study.

2.2.1 Model definition

We investigate the rectangular-binary-perceptron (RBP) problem introduced in ref. [Aubin et al., 2019]. The RBP has the key property of having a symmetric activation function, characterized by a parameter $K > 0$. Given a vector of binary weights $\mathbf{w} \in \{\pm 1\}^N$ and an input $\boldsymbol{\xi} \in \mathbb{R}^N$ (an example), we say that $\mathbf{w}$ satisfies the example if $|\boldsymbol{\xi} \cdot \mathbf{w}| < K$.[1] This symmetry simplifies the theoretical analysis and allows to obtain tighter bounds for the storage capacity through the first and second moment methods.

For a given set of inputs $\boldsymbol{\xi}^\mu \in \mathbb{R}^N$, with $\mu = 1, \dots, M$, the RBP problem can be expressed as a constraint satisfaction problem (CSP) over the binary weights. Throughout the paper we will assume the entries ξ_i^μ to be *i.i.d.* Gaussian variables with zero mean and variance $1/N$. A binary vector $\mathbf{w} \in \{\pm 1\}^N$ is called a *solution* of the problem if it satisfies

$$\sum_{i=1}^{N} w_i \xi_i^\mu \in I_K \qquad \forall \mu \in [M]\,, \tag{2.1}$$

where $I_K = [-K, K]$. Equivalently, a vector $\mathbf{w}$ is a solution of the RBP problem iff the function $\mathbb{X}_{\boldsymbol{\xi},K} : \{-1, 1\}^N \to \{0, 1\}$, defined as

$$\mathbb{X}_{\boldsymbol{\xi},K}(\mathbf{w}) = \prod_{\mu=1}^{M} \mathbb{1}\left(\sum_{i=1}^{N} w_i \xi_i^\mu \in I_K \right), \tag{2.2}$$

is equal to one, where we have denoted with $\mathbb{1}(p)$ an indicator function that is 1 if the statement p is true and 0 otherwise.

The *storage capacity* is then defined similarly to the satisfiability threshold in random constraint satisfaction problems: we denote the constraint density as $\alpha \equiv M/N$ and define the storage capacity $\alpha_c(K)$, also known as SAT-UNSAT transition point, as the infimum of densities α such that, in the limit $N \to \infty$, with high probability (over the choice of the matrix ξ_i^μ) there are no solutions. It is natural to conjecture that the converse also holds, i.e. that the storage capacity $\alpha_c(K)$ equals the supremum of α such that in the limit $N \to \infty$ solutions exist with high probability. In this case we would say the storage capacity is a *sharp threshold*.

[1] This setting corresponds to a binary classification problem with training examples from a single class. This simplifies the analysis.

2.2.2 Replicated Systems and Dense Clusters

In order to obtain a geometric characterization of the solution space, we consider the Hamming distance of any two configurations $\boldsymbol{w}^1$ and $\boldsymbol{w}^2$, defined by

$$d_{\mathrm{H}}\left(\mathbf{w}^1, \mathbf{w}^2\right) \equiv \sum_{i=1}^{N} \left(1 - w_i^1 w_i^2\right) / 2.$$

Even if an exponential number of solutions exist for $\alpha < \alpha_c(K)$, the overwhelming majority are *isolated*: for each such solution, there exists a radius $r_{\min}$ such that the number of other solutions within a distance $\lfloor N r_{\min} \rfloor$ is sub-exponential. We are interested instead in the presence of *dense regions*, which are characterized by the fact that there is a configuration around which the number of solutions within a given radius $\lfloor Nr \rfloor$ is exponential for all r in some neighborhood of 0. We speak of *ultra-dense* regions when the logarithm of the density of solutions tends exponentially fast to 0 as $r \to 0$.

Suppose now that a dense region around some reference configuration exists, choose a sufficiently small value $r > 0$, and call x the typical distance between any two solutions at distance r from the reference. In general, $0 < x \leq 2r$, and for an ultra-dense region $x = 2r(1-r)$ in the limit of large N. Therefore for any x below some threshold there should exist an exponential number of solutions at mutual normalized distance x.

We thus investigate the problem of finding a set of y solutions of the RBP problem, where y is an arbitrary natural number, with all pairwise distances constrained to some value $\lfloor Nx \rfloor$. The existence (for some range of α) of such set of solutions, w.h.p. in the large N limit, for arbitrarily large values of y and all x in some neighborhood of 0, is a necessary condition for the presence of dense regions. These sets of y solutions would coexist with an exponentially larger number of isolated solutions, and therefore the usual tools of statistical physics are not sufficient to reveal their presence, and a large deviation analysis is necessary [Baldassi et al., 2015].

As a starting point for the analysis we introduce the partition function of the model with y real replicas, $\mathcal{Z}_y$, accounting for the number of such sets (up to a $y!$ symmetry factor). For any fixed (normalized) distance $x \in [0, 1]$, this is given by

$$\mathcal{Z}_y(x, K, \boldsymbol{\xi}) \equiv \sum_{\{\mathbf{w}^a\}_{a=1}^{y}} \prod_{a=1}^{y} \mathbb{X}_{\boldsymbol{\xi},K}(\mathbf{w}^a) \prod_{a<b}^{y} \mathbb{1}\left(d_H\left(\mathbf{w}^a, \mathbf{w}^b\right) = \lfloor Nx \rfloor\right). \tag{2.3}$$

The summation here is over the 2^{yN} spin configurations. We denote with $\alpha_c^y(x, K)$ the SAT/UNSAT threshold (if it exists) in the $N \uparrow \infty$ limit and under the probability distribution for $\boldsymbol{\xi}$ described in the previous Section. The asymptotic behavior is captured by the (normalized) local entropy ϕ_y defined by[2]

[2]We use a simpler definition compared to ref. [Baldassi et al., 2015] here, avoiding the explicit use of a reference configuration. The technical justification for this can be found in ref. [Baldassi et al., 2020b]; intuitively, the reference is defined implicitly as the barycenter, and the results are basically equivalent for large y.

$$\phi_y(x,K,\alpha) = \lim_{N\to\infty} \frac{1}{yN} \mathbb{E}_\xi \ln \mathcal{Z}_y(x,K,\xi)\,. \tag{2.4}$$

The interpretation of this quantity is as follows. If ϕ_y is positive, the number of groups of y solutions is exponential. For any group of y solutions that contributes to the sum in $\mathcal{Z}_y$ we can use their barycenter (which will be at distance $r = \frac{1-\sqrt{1-2x}}{2}$ from each of them) as a reference configuration, and in the limit of large y the sum is dominated by the regions with the highest density of solutions at distance r from their center, provided they are evenly distributed. Also in this limit the logarithm of the density of solutions is computed as $\phi_y(x,K,\alpha) - \phi_y(x,K,0) = \phi_y(x,K,\alpha) - H_2\left(\frac{1-\sqrt{1-2x}}{2}\right)$ where $H_2(r) = -r\ln r - (1-r)\ln(1-r)$ is the two-state entropy function. If a dense region exists around a configuration, we should observe a positive ϕ_y for all y and for all x in some neighborhood of 0, and for ultra-dense regions we should have $\lim_{y\to\infty} \phi_y(x,K,\alpha) = H_2\left(\frac{1-\sqrt{1-2x}}{2}\right) - O\left(e^{-\frac{1}{x}}\right)$ for sufficiently small x.[3]

The computation of ϕ_y can be approached by rigorous techniques only for small y, as discussed in the next sections. In the general case, for any finite y and in the $y \to \infty$ limit, it can be carried out at present only using the non-rigorous replica method of statistical physics of disordered systems. The computations for this model follow entirely those of ref. [Baldassi et al., 2015] and are reported in Appendix 2.6.1.

The replica analysis in the $y \to \infty$ limit strongly suggests the existence of ultra-dense regions of solutions: as shown in fig. 2.1, for $K = 1$ and for sufficiently small x the curves for α below the SAT-UNSAT transition, i.e. $\alpha < \alpha_c \simeq 1.815\ldots$, tend to collapse onto the curve for $\alpha = 0$, implying that these regions are maximally dense in their immediate surroundings (nearly all configurations are solutions in an extensive region centered around their barycenter). Furthermore, there is a transition at around $\alpha_U \simeq 1.58$ after which the curves are no longer monotonic. Overall, this is the same phenomenology that was observed (and confirmed by numerical simulations) for the standard binary perceptron model in ref. [Baldassi et al., 2015], and we interpret it in the same way, i.e. we speculate that ultra-dense sub-dominant regions of solutions exist, and that the break of monotonicity at $\alpha_U \simeq 1.58$ signals a transition[4] between two regimes: one for low α in which the ultra-dense regions are immersed in a vast connected structure, and one at high α in which the structure of the dense solutions fragments into separate regions that are no longer easily accessible.[5]

These results were obtained with the so-called replica-symmetric ansatz, and they

[3] Although these are in principle necessary conditions, and not sufficient, the latter scenario of a log-density going to 0 in particular seems very unlikely in the absence of ultra-dense regions, and indeed when the matter was investigated numerically for the standard perceptron model these rare regions were found and their properties were in good agreement with the theory in a wide range of parameters [Baldassi et al., 2015, 2016b].

[4] In ref. [Baldassi et al., 2015] it was shown that some geometric constraints are violated in a region of x for $\alpha \geq \alpha_U$ implying the onset of strong symmetry-breaking effects, with numerical evidence supporting the switch to a different regime.

[5] It should be noted that in the standard binary perceptron case (i.e. with sign activation) there is empirical evidence only for the first scenario of a vast connected structure with ultra-dense regions in it, while the second scenario of fragmented regions has never been directly observed at large N, arguably due to the intrinsic algorithmic hardness of finding such regions.

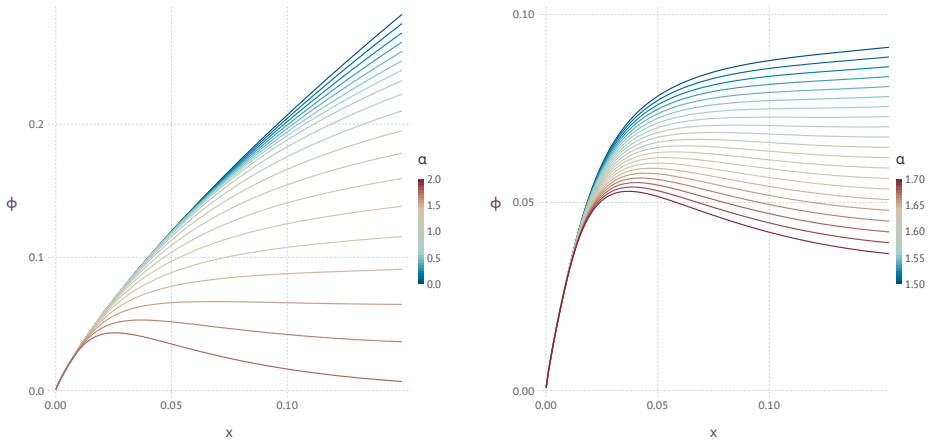

Figure 2.1: Plot of free local entropy of eq. (2.3) as a function of the normalized Hamming distance between solutions x, obtained with the replica method using the replica-symmetric ansatz (see Appendix 2.6.1 for the details). In both figures the value of the half-width of the channel is $K = 1$. (Left) Curves for $\alpha = 0$ up to $\alpha = 1.8$ in steps of 0.1. When the distance x approaches zero we see that all curves tend to coincide with the curve for $\alpha = 0$, meaning that there exist regions of solutions that are maximally dense (nearly all configurations are solutions) in their immediate surroundings. (Right) Zoom on the interval of values of α where there is a change in monotonicity, which we interpret as signaling a fragmentation of the dense clusters into separate pieces. We refine the step of α to 0.01, and we find that the change happens for $\alpha_U \simeq 1.58$.

are certainly not exact. However, as in previous studies [Baldassi et al., 2015], the corrections (which would require the use of a replica-symmetry-broken ansatz) only become numerically relevant at relatively large α (e.g. we may expect small corrections to the value of α_U, and larger effects close to α_c), and they don't affect the qualitative picture, the emerging phenomenology and its physical interpretation.

2.3 Pairs of solutions ($y = 2$): rigorous bounds

We are able to derive rigorous lower and upper bounds for the existence of pairs of solutions, i.e. for the $y = 2$ case, without resorting to the replica method.

The idea of the derivation follows very closely the strategy used in refs. [Mézard et al., 2005, Daudé et al., 2008] for the random K-SAT problem.

We define a SAT-x-pair as a pair of binary weights $\mathbf{w}^1, \mathbf{w}^2 \in \{-1, 1\}^N$, which are both solutions of the CSP, and whose Hamming distance is $d_H\left(\mathbf{w}^1, \mathbf{w}^2\right) = \lfloor Nx \rfloor$. The number of such pairs is $\mathcal{Z}_{y=2}\left(x, K, \boldsymbol{\xi}\right)$, see eq. (2.3).

2.3.1 Upper bound: the first moment method

In this section we are interested in finding an upper-bound (which depends on x) to the critical capacity of pairs of solutions. To do that we use the following lemma.

Lemma 1 (First moment method). *If the random variable X is non-negative and integer-valued then we have*

$$\mathbb{P}\left[X > 0\right] \leq \mathbb{E}\left[X\right]. \tag{2.5}$$

Theorem 1. *For each K and $0 < x < 1$, and for all α such that*

$$\alpha > \alpha_{UB}\left(x, K\right) \equiv -\frac{\ln 2 + H_2\left(x\right)}{\ln f_1\left(x, K\right)}, \tag{2.6}$$

there are no SAT-x-pairs w.h.p.

Proof. Let us apply eq. (2.5) it to the random variable $\mathcal{Z}_{y=2}$. We get:

$$\mathbb{P}\left[\mathcal{Z}_{y=2}\left(x, K, \boldsymbol{\xi}\right) > 0\right] \leq \mathbb{E}\left[\mathcal{Z}_{y=2}\left(x, K, \boldsymbol{\xi}\right)\right] = 2^N \binom{N}{\lfloor Nx \rfloor} \mathbb{P}\left[v_1 \in I_K, v_2 \in I_K\right]^M \tag{2.7}$$

where we have introduced the two Gaussian random variables v_1 and v_2, with $\mathbb{E}\left[v_1\right] = \mathbb{E}\left[v_2\right] = 0$, $\mathbb{E}\left[v_1^2\right] = \mathbb{E}\left[v_2^2\right] = 1$, and covariance

$$\mathbb{E}\left[v_1 v_2\right] = \frac{N - 2\left\lfloor Nx \right\rfloor}{N} \underset{N \to +\infty}{\longrightarrow} 1 - 2x.$$

Let us consider the normalized logarithm of the first moment,

$$F\left(x, K, \alpha\right) = \lim_{N \to \infty} \frac{1}{N} \ln \mathbb{E}\left[\mathcal{Z}_{y=2}\left(x, K, \boldsymbol{\xi}\right)\right] = \ln 2 + H_2\left(x\right) + \alpha \ln f_1\left(x, K\right),$$

where as before $H_2\left(x\right) = -x \ln x - \left(1 - x\right) \ln\left(1 - x\right)$ is the two-state entropy function while $f_1\left(x, K\right)$ is defined as follows. Denote with Σ_2 the covariance matrix of the Gaussian random vector $\vec{v} = \left(v_1, v_2\right)$ whose components have covariance equal to $1 - 2x$ and variances equal to one. We define $f_1\left(x, K\right)$ as the probability that this random vector takes values in the box $\left[-K, K\right]^2$:

$$\begin{aligned} f_1\left(x, K\right) &= \frac{1}{2\pi \left|\Sigma_2\right|^{1/2}} \int_{-K}^{K} \int_{-K}^{K} dv_1 dv_2 e^{-\vec{v}^T \Sigma_2^{-1} \vec{v}} \\ &= \int_{-K}^{K} du_1 \frac{e^{-u_1^2/2}}{\sqrt{2\pi}} \int_{\frac{-K-(1-2x)u_1}{2\sqrt{x(1-x)}}}^{\frac{K-(1-2x)u_1}{2\sqrt{x(1-x)}}} du_2 \frac{e^{-u_2^2/2}}{\sqrt{2\pi}}. \end{aligned} \tag{2.8}$$

From the inequality (2.7), $F\left(x, K, \alpha\right) < 0$ implies that $\lim_{N \to \infty} \mathbb{P}\left[\mathcal{Z}_{y=2}\left(x, K, \xi\right) > 0\right] = 0$. In turn this provides the upper bound in the statement of the theorem. □

Notice that the first moment computation for $Z_{y=2}(x)$ is similar to the second moment computation for $\mathcal{Z}_{y=1}$ in ref. [Aubin et al., 2019]: in the former x enters as an external constraint, in the latter as an order parameter to be optimized.

The upper bound that we obtained for $K = 1$ and as a function of x is shown in fig. 2.2. For $x = 0$ the upper bound trivially reduces to the one for a single replica as found in ref. [Aubin et al., 2019]. The same happens also for $x = 1/2$, as the two constrained replicas behave as independent systems in the large N limit.

2.3.2 Lower bound: the second moment method

We compute the lower bound to the critical capacity using the second moment method, which is a direct consequence of the Cauchy-Schwarz inequality:

Lemma 2 (Second moment method). *If X is a non-negative random variable, then*

$$\mathbb{P}\left[X > 0\right] \geq \frac{\mathbb{E}\left[X\right]^2}{\mathbb{E}\left[X^2\right]}. \tag{2.9}$$

From the results of section 2.3.1 we have

$$\mathbb{E}\left[\mathcal{Z}_{y=2}\left(x, K, \boldsymbol{\xi}\right)\right] = 2^N \binom{N}{\lfloor Nx \rfloor} f_1\left(\frac{\lfloor Nx \rfloor}{N}, K\right)^M, \tag{2.10}$$

where $f_1\left(x, K\right)$ is defined like in eq. (2.8). The second moment of the random variable $\mathcal{Z}_{y=2}$ follows from simple combinatorics and reads

$$\begin{aligned}
&\mathbb{E}\left[\mathcal{Z}_{y=2}^2\left(x, K, \boldsymbol{\xi}\right)\right] \\
&= \sum_{\left\{\mathbf{w}^1\right\}} \sum_{\{\mathbf{w}^2\}} \sum_{\left\{\tilde{\mathbf{w}}^1\right\}} \sum_{\{\tilde{\mathbf{w}}^2\}} \mathbb{1}\left(d_H\left(\mathbf{w}^1, \mathbf{w}^2\right) = \lfloor Nx \rfloor\right) \mathbb{1}\left(d_H\left(\tilde{\mathbf{w}}^1, \tilde{\mathbf{w}}^2\right) = \lfloor Nx \rfloor\right) \\
&\quad \prod_{\mu=1}^{M} \mathbb{E}\left[\mathbb{1}\left(w^1 \cdot \xi^\mu \in I_K\right) \mathbb{1}\left(w^2 \cdot \xi^\mu \in I_K\right) \mathbb{1}\left(\tilde{w}^1 \cdot \xi^\mu \in I_K\right) \mathbb{1}\left(\tilde{w}^2 \cdot \xi^\mu \in I_K\right)\right] \\
&= 2^N \sum_{\mathbf{a} \in V_{N,x} \cap \{0,1/N,2/N,\ldots,1\}^8} \frac{N!}{\prod_{i=0}^{7}\left(Na_i\right)!} f_2\left(\mathbf{a}, x, K\right)^M,
\end{aligned} \tag{2.11}$$

where we have adopted the following conventions.

- $\mathbf{a}$ is an 8-component vector giving the proportion of each type of quadruplets $\left(w_i^1, w_i^2, \tilde{w}_i^1, \tilde{w}_i^2\right)$ as described in the table below, where we have arbitrarily (but without loss of generality) fixed $\mathbf{w}^1$ to $(1, \ldots, 1)$. Fixing the vector $\mathbf{a}$ entails fixing all the possible overlaps between the vectors $w^1, w^2, \tilde{w}^1$ and $\tilde{w}^2$ and consequently the covariances of the random variables $z_1 := w^1 \cdot \xi$, $z_2 := w^2 \cdot \xi$, $\tilde{z}_1 := \tilde{w}^1 \cdot \xi$ and $\tilde{z}_2 := \tilde{w}^2 \cdot \xi$ with $\xi_i \sim \mathcal{N}\left(0, 1/N\right)$ i.i.d. These covariances as functions of $\mathbf{a}$ are made explicit in eq. (2.12).

	a_0	a_1	a_2	a_3	a_4	a_5	a_6	a_7
w_i^1	+	+	+	+	+	+	+	+
w_i^2	+	+	+	+	−	−	−	−
$\tilde{w}_i^1$	+	+	−	−	+	+	−	−
$\tilde{w}_i^2$	+	−	+	−	+	−	+	−

- $f_2\left(\mathbf{a}, x, K\right)$ has the expression

$$f_2\left(\mathbf{a}, x, K\right) = \mathbb{P}\left[z_1 \in I_K, z_2 \in I_K, \tilde{z}_1 \in I_K, \tilde{z}_2 \in I_K\right] .$$

where $\mathbf{z}^T := (z_1, z_2, \tilde{z}_1, \tilde{z}_2)$ is a 4-dimensional Gaussian vector, with the following set of covariances:

$$\Sigma = \begin{pmatrix} 1 & q_1 & q_{01} & q_{02} \\ q_1 & 1 & q_{03} & q_{04} \\ q_{01} & q_{03} & 1 & q_1 \\ q_{02} & q_{04} & q_1 & 1 \end{pmatrix} \quad \text{where} \quad \begin{cases} q_1 = 1 - 2\frac{\lfloor Nx \rfloor}{N} \\ q_{01} = 1 - 2\left(a_2 + a_3 + a_6 + a_7\right) \\ q_{02} = 1 - 2\left(a_1 + a_3 + a_5 + a_7\right) \\ q_{03} = 1 - 2\left(a_2 + a_3 + a_4 + a_5\right) \\ q_{04} = 1 - 2\left(a_1 + a_3 + a_4 + a_6\right) \end{cases} . \tag{2.12}$$

Therefore $f_2(\mathbf{a}, x, K)$ can be simply written as the following Gaussian integral

$$f_2(\mathbf{a}, x, K) = \int_{I_K^4} dz_1 dz_2 d\tilde{z}_1 d\tilde{z}_2 \frac{1}{(2\pi)^2 |\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T \Sigma^{-1} \mathbf{z}}. \tag{2.13}$$

- The set $V_{N,x} \subset [0,1]^8$ is a simplex specified by:

$$\begin{cases} \lfloor N\left(a_4 + a_5 + a_6 + a_7\right) \rfloor = \lfloor Nx \rfloor \\ \lfloor N\left(a_1 + a_2 + a_5 + a_6\right) \rfloor = \lfloor Nx \rfloor \\ \sum_{i=0}^7 a_i = 1 \end{cases} . \tag{2.14}$$

These three conditions correspond to the normalization of the proportions and to the enforcement of the conditions $d_{\mathbf{w}^1\mathbf{w}^2} = \lfloor Nx \rfloor$, $d_{\tilde{\mathbf{w}}^1\tilde{\mathbf{w}}^2} = \lfloor Nx \rfloor$. When $N \to \infty$, $V_x = \bigcap_{N \in \mathbb{N}} V_{N,x}$ defines a five-dimensional simplex described by the three hyperplanes:

$$\begin{cases} a_4 + a_5 + a_6 + a_7 = x \\ a_1 + a_2 + a_5 + a_6 = x \\ \sum_{i=0}^7 a_i = 1 \end{cases} . \tag{2.15}$$

In order to yield an asymptotic estimate of $\mathbb{E}\left[\mathcal{Z}_{y=2}^2\right]$ we first use the following known result, which comes from the approximation of integrals by sums (proof in Appendix 2.6.2):

Lemma 3. *Let $\psi(\mathbf{a})$ be a real, positive, continuous function of $\mathbf{a}$, and let $V_{N,x}$, V_x be as defined above. Then for any given x there exists a constant C_0 such that for sufficiently large N:*[6]

$$\sum_{\mathbf{a} \in V_{N,x} \cap \{0,1/N,2/N,\ldots,1\}^8} \frac{N!}{\prod_{i=0}^7 (Na_i)!} \psi(\mathbf{a})^N \leq C_0 N^{3/2} \int_{V_x} d\mathbf{a}\, e^{N[H_8(\mathbf{a}) + \ln \psi(\mathbf{a})]}, \tag{2.16}$$

where $H_8(\mathbf{a}) = -\sum_{i=0}^7 a_i \ln a_i$.

The bound for the second moment then reads:

$$\mathbb{E}\left[\mathcal{Z}_{y=2}^2(x, K, \boldsymbol{\xi})\right] \leq C_0 N^{3/2} \int_{V_x} d\mathbf{a}\, e^{N[\ln 2 + H_8(\mathbf{a}) + \alpha \ln f_2(\mathbf{a}, x, K)]}, \tag{2.17}$$

[6] Here and below this 8-dimensional integration is to be understood as being performed with a uniform measure in the 5-dimensional subspace V_x, i.e. $\int_{V_x} d\mathbf{a} \equiv \int_{[0,1]^8} d\mathbf{a}\, \delta\left(a_4 + a_5 + a_6 + a_7 - x\right) \delta\left(a_1 + a_2 + a_5 + a_6 - x\right) \delta\left(\sum_{i=0}^7 a_i - 1\right)$, where δ is a Dirac delta, cf. eq. (2.15).

which is obtained from substitution of eq. (2.3.2) into Lemma 3. The number of components of the vector $\mathbf{a}$ is eight, but we can reduce their number to five with a change of variables and rewrite the integral in a particularly simple form where f_2 just depends on four of them. This is done in Appendix 2.6.2. Here we give just the final expression where the new integration variables are η (a scalar) and $\vec{q}_0 = (q_{01}, q_{02}, q_{03}, q_{04})$. The bound becomes

$$\mathbb{E}\left[\mathcal{Z}^2_{y=2}(x, K, \boldsymbol{\xi})\right] \leq C_0 N^{3/2} \int_{\tilde{V}_x} d\vec{q}_0 \, d\eta \, e^{N[\ln 2 + H_8(\vec{q}_0, \eta, x) + \alpha \ln f_2(\vec{q}_0, x, K)]}, \tag{2.18}$$

where:

- $f_2(\vec{q}_0, x, K)$ has the expression

$$f_2(\vec{q}_0, x, K) = \int_{I_K^4} dz_1 dz_2 d\tilde{z}_1 d\tilde{z}_2 \frac{1}{(2\pi)^2 |\Sigma|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T \Sigma^{-1} \mathbf{z}},$$

  where Σ is the covariance matrix of eq. (2.12) with $q_1 = 1 - 2x$ and where the components of $\vec{q}_0$ are considered as independent variables.

- $H_8(\vec{q}_0, \eta, x)$ is defined as the Shannon entropy of a probability mass function with masses corresponding to the components of the following vector:

$$\begin{pmatrix} \frac{1}{4}(q_{02} + q_{03} + 2 - 4x) + \eta \\ \frac{1}{4}(q_{01} - q_{02} + 2x) - \eta \\ \frac{1}{4}(-q_{03} + q_{04} + 2x) - \eta \\ \frac{1}{4}(2 - q_{01} - q_{04} - 4x) + \eta \\ \frac{1}{4}(q_{01} - q_{03} + 2x) - \eta \\ \eta \\ \frac{1}{4}(-q_{01} + q_{02} + q_{03} - q_{04}) + \eta \\ \frac{1}{4}(-q_{02} + q_{04} + 2x) - \eta \end{pmatrix}; \tag{2.19}$$

- $\tilde{V}_x$ is the new domain of integration specified by the inequalities

$$\begin{cases} \frac{1}{4}(q_{01} - q_{02} + 2x - 4) \leq \eta \leq \frac{1}{4}(q_{01} - q_{02} + 2x) \\ \frac{1}{4}(-q_{03} + q_{04} + 2x - 4) \leq \eta \leq \frac{1}{4}(-q_{03} + q_{04} + 2x) \\ \frac{1}{4}(q_{01} + q_{04} + 4x - 2) \leq \eta \leq \frac{1}{4}(q_{01} + q_{04} + 4x + 2) \\ \frac{1}{4}(q_{01} - q_{03} + 2x - 4) \leq \eta \leq \frac{1}{4}(q_{01} - q_{03} + 2x) \\ 0 \leq \eta \leq 1 \\ \frac{1}{4}(q_{01} - q_{02} - q_{03} + q_{04}) \leq \eta \\ \frac{1}{4}(-q_{02} + q_{04} + 2x - 4) \leq \eta \leq \frac{1}{4}(-q_{02} + q_{04} + 2x) \\ \frac{1}{4}(-q_{02} - q_{03} + 4x - 2) \leq \eta \end{cases}, \tag{2.20}$$

some of which are already contained in eq. (2.19).

Proposition 1. *For each K, x, define:*

$$\Phi_{x,K,\alpha}(\vec{q}_0, \eta) = H_8(\vec{q}_0, \eta, x) - \ln 2 - 2H_2(x) + \alpha \ln f_2(\vec{q}_0, x, K) - 2\alpha \ln f_1(x, K) \ . \tag{2.21}$$

and let $(\vec{q}_0^{\,M}, \eta^M) \in \tilde{V}_x$ *be the global maximum of* $\Phi_{x,K,\alpha}$ *restricted to* $\tilde{V}_x$*. Then there exists a* x, K*-dependent constant* $C > 0$ *such that, for* N *sufficiently large,*

$$\frac{\mathbb{E}\left[\mathcal{Z}_{y=2}(x, K, \boldsymbol{\xi})\right]^2}{\mathbb{E}\left[\mathcal{Z}_{y=2}^2(x, K, \boldsymbol{\xi})\right]} \geq C \exp\left(-N\Phi_{x,K,\alpha}\left(\vec{q}_0^{\,M}, \eta^M\right)\right) . \tag{2.22}$$

Proof. Applying Laplace method to the integral in eq. (2.18), for some constant C_1 and for N large enough we obtain

$$\mathbb{E}\left[\mathcal{Z}_{y=2}^2(x, K, \boldsymbol{\xi})\right] \leq C_1 N^{-1} e^{N\left[\ln 2 + H_8\left(\vec{q}_0^{\,M}, \eta, x\right) + \alpha \ln f_2\left(\vec{q}_0^{\,M}, x, K\right)\right]} ,$$

where the factor $N^{-1} = N^{\frac{3}{2}-\frac{5}{2}}$ stems from the Gaussian fluctuations around the 5-dimensional saddle point. For the first moment instead, a simple application of Stirling formula to eq. (2.7) leads, for some constant c_1 and N large enough, to

$$\mathbb{E}\left[\mathcal{Z}_{y=2}(x, K, \boldsymbol{\xi})\right]^2 \geq c_1 N^{-1} e^{2N[\ln 2 + H_2(x) + \alpha \ln f_1(x,K)]} .$$

Combining the two expressions, the proposition follows. □

Theorem 2. *For each* K *and* $0 < x < 1$*, and for all* α *such that*

$$\alpha < \alpha_{LB}(x, K) \equiv \inf_{(\vec{q}_0, \eta) \in \tilde{V}_x^+} \frac{\ln 2 + 2H_2(x) - H_8(\vec{q}_0, \eta, x)}{\ln f_2(\vec{q}_0, x, K) - 2 \ln f_1(x, K)} \tag{2.23}$$

we have that there is a positive probability of finding SAT-x-pairs of solutions, namely

$$\liminf_{N \to \infty} \mathbb{P}\left[\mathcal{Z}_{y=2}(x, K, \boldsymbol{\xi}) > 0\right] > 0.$$

Proof. Given that $\Phi_{x,K,\alpha}\left(\vec{q}_0^{\,M}, \eta^M\right) \geq 0$, the second moment method gives a useful bound just when $\Phi_{x,K,\alpha}\left(\vec{q}_0^{\,M}, \eta^M\right) = 0$. If instead $\Phi_{x,K,\alpha}\left(\vec{q}_0^{\,M}, \eta^M\right) > 0$, the probability is bounded above zero (included) and the bound is non-informative.

For a particular point $(\vec{q}_0^{\,\star}, \eta^\star) \in \tilde{V}_x$, which can be interpreted intuitively as capturing the situation where the two pairs of solutions are uncorrelated, we have that $\Phi_{x,K,\alpha}\left(\vec{q}_0^{\,\star}, \eta^\star\right) = 0$ for all values of α. This point $(\vec{q}_0^{\,\star}, \eta^\star)$ is specified by the following equations,

$$q_{01}^\star = 0,\ q_{02}^\star = 0,\ q_{03}^\star = 0,\ q_{04}^\star = 0,\ \eta^\star = \frac{x^2}{2}.$$

In that case, we have the following properties:

- $H_8\left(\vec{q}_0^{\,\star}, \eta^\star, x\right) = \ln 2 + 2H_2(x)$,
- $f_2\left(\vec{q}_0^{\,\star}, x, K\right) = f_1(x, K)^2$.

Therefore, α_{LB} is the largest value of α such that $(\vec{q}_0^{\,\star}, \eta^\star)$ is a global maximum, i.e. such that there exists no $(\vec{q}_0, \eta) \in \tilde{V}_x$ with $\Phi_{x,K,\alpha}(\vec{q}_0, \eta) > 0$. In particular, for $\alpha = 0$ the second moment bound holds (proof in Appendix 2.6.2):

$$\Phi_{x,K,\alpha=0}(\vec{q}_0, \eta) = H_8(\vec{q}_0, \eta, x) - \ln 2 - 2H_2(x) \leq 0 \quad \forall (\vec{q}_0, \eta) \in \tilde{V}_x. \tag{2.24}$$

Now, let us split $\tilde{V}_x$ in the following way:

$$\tilde{V}_x^+ := \left\{ (\vec{q}_0, \eta) \in \tilde{V}_x \mid f_2(\vec{q}_0, x, K) > f_1^2(x, K) \right\} \quad \text{and}$$

$$\tilde{V}_x^- := \left\{ (\vec{q}_0, \eta) \in \tilde{V}_x \mid f_2(\vec{q}_0, x, K) \leq f_1^2(x, K) \right\}.$$

It follows that for all $(\vec{q}_0, \eta) \in \tilde{V}_x^-$ and $\alpha > 0$ we have

$$\Phi_{x,K,\alpha}(\vec{q}_0, \eta) \leq \Phi_{x,K,\alpha=0}(\vec{q}_0, \eta) \leq 0.$$

As already discussed, α_{LB} is the largest value of α such that

$$\max_{(\vec{q}_0,\eta)\in\tilde{V}_x} \Phi_{x,K,\alpha}(\vec{q}_0, \eta) = 0.$$

From the previous observation

$$\max_{(\vec{q}_0,\eta)\in\tilde{V}_x} \Phi_{x,K,\alpha}(\vec{q}_0, \eta) = \sup_{(\vec{q}_0,\eta)\in\tilde{V}_x^+} \Phi_{x,K,\alpha}(\vec{q}_0, \eta),$$

and therefore α_{LB} is the largest value of α such that

$$\sup_{(\vec{q}_0,\eta)\in\tilde{V}_x^+} \Phi_{x,K,\alpha}(\vec{q}_0, \eta) = 0.$$

Then, α_{LB} is the largest value of α such that there exists no $(\vec{q}_0, \eta) \in \tilde{V}_x^+$ with $\Phi_{x,K,\alpha}(\vec{q}_0, \eta) > 0$, which is true if and only if

$$H_8(\vec{q}_0, \eta, x) - \ln 2 - 2H_2(x) + \alpha \ln f_2(\vec{q}_0, x, K) - 2\alpha \ln f_1(x, K) \leq 0, \tag{2.25}$$

for all $(\vec{q}_0, \eta) \in \tilde{V}_x^+$ and for all $\alpha \leq \alpha_{LB}$. Therefore, eq. (2.25) implies that for $\alpha \leq \alpha_{LB}$ the following condition must hold as well:

$$\alpha \leq \frac{\ln 2 + 2H_2(x) - H_8(\vec{q}_0, \eta, x)}{\ln f_2(\vec{q}_0, x, K) - 2 \ln f_1(x, K)} \quad \forall\ (\vec{q}_0, \eta) \in \tilde{V}_x^+. \tag{2.26}$$

□

The optimization in (2.23) can be simplified further by slicing the set $\tilde{V}_x^+$ in the two "directions" $\vec{q}_0$ and η. We define a $\vec{q}_0$-slice as $\left(\tilde{V}_x^+\right)_{\vec{q}_0} := \left\{\eta \mid (\vec{q}_0, \eta) \in \tilde{V}_x^+\right\}$ and the natural projection of the set $\tilde{V}_x^+$ on the $\vec{q}_0$-subspace as $\pi_{\vec{q}_0}\left(\tilde{V}_x^+\right) = \left\{\vec{q}_0 \mid \exists\, \eta \text{ s.t. } (\vec{q}_0, \eta) \in \tilde{V}_x^+\right\}$. With this notation, eq. (2.23) becomes:

$$\alpha_{LB}(x, K) = \inf_{\vec{q}_0 \in \pi_{\vec{q}_0}\left(\tilde{V}_x^+\right)} \frac{\ln 2 + 2H_2(x) - \sup_{\eta\in\left(\tilde{V}_x^+\right)_{\vec{q}_0}} H_8(\vec{q}_0, \eta, x)}{\ln f_2(\vec{q}_0, x, K) - 2 \ln f_1(x, K)}. \tag{2.27}$$

The optimization in η is easy because the function $H_8(\vec{q}_0, \eta)$ is concave in η for each $\vec{q}_0$. This is not necessarily true for the optimization in $\vec{q}_0$. In fact, it is crucial that we find the global optimum of the objective function because this gives the correct value for the lower bound. To this purpose we have devised two computational strategies. First

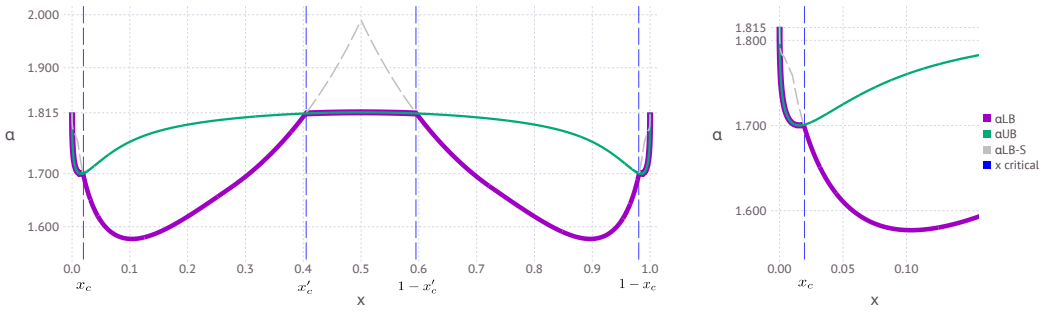

Figure 2.2: Lower and upper bounds for the RBP with $K = 1$. (Left) Lower and upper bounds on the whole range $x \in [0, 1]$. These bounds are symmetric around the vertical axis that passes by $x = 0.5$. In correspondence of the SB solution, the lower bound prediction from the S point (gray line) is larger than the upper bound and therefore patently wrong. This is what happens in the regions $x \leq x_c$, $x \geq 1 - x_c$ and $x'_c \leq x \leq 1 - x'_c$, where the two critical values $x_c \simeq 0.195\ldots$ and $x'_c \simeq 0.405\ldots$ are highlighted by the blue vertical lines on the left of the symmetry axis. In the regions $x_c \leq x \leq x'_c$ and $1 - x'_c \leq x \leq 1 - x_c$, there is a gap between the lower (purple line) and the upper bound (green line) where the S solution is indeed valid. (Right) Zoom of the figure on the left, in the region around x_c. Here, for $x \leq x_c$ the S solution fails. This is evident from the fact that the symmetric lower bound becomes bigger than the upper bound (gray line). In this region instead, the true lower bound perfectly matches the upper bound since the optimum of eq. (2.31) is in correspondence of the SB solution.

we evaluated the objective function on a 4-dimensional grid with increasing number of points. Then we have also implemented a simple gradient descent starting from the points of the grid. The different strategies are discussed in Appendix 2.6.2.

The bounds that we obtain in fig. 2.2 are symmetric around the value $x = 0.5$ and there are two critical values $x_c, x'_c \in [0, 0.5]$ (plus the symmetric ones, $1 - x_c$ and $1 - x'_c$) that delimit regions characterized by two different phases. For values of x such that $x_c \leq x \leq x'_c$ or $1 - x'_c \leq x \leq 1 - x_c$ all four entries of $\vec{q}_0$ take the same value. We use the subscript S to denote this kind of solution. Instead for $x \leq x_c$, $x \geq 1 - x_c$ and $x'_c \leq x \leq 1 - x'_c$, this symmetry is broken and the optimum is achieved on a different point that we call Symmetry Broken (SB) solution. This point has the property that the two pairs of binary vectors of solutions $(\mathbf{w}_1, \mathbf{w}_2)$ and $(\tilde{\mathbf{w}}_1, \tilde{\mathbf{w}}_2)$ coincide, as can be seen from the structure of the covariance matrix. We report below the covariance structure of the two solutions, where we adopted the convention $q_1 := 1 - 2x$. The symmetric covariance matrix is the following:

$$\Sigma_S = \begin{pmatrix} 1 & q_1 & q_0 & q_0 \\ q_1 & 1 & q_0 & q_0 \\ q_0 & q_0 & 1 & q_1 \\ q_0 & q_0 & q_1 & 1 \end{pmatrix}, \tag{2.28}$$

while the one corresponding to point SB is the following:

$$\Sigma_{SB} = \begin{pmatrix} 1 & q_1 & 1 & q_1 \\ q_1 & 1 & q_1 & 1 \\ 1 & q_1 & 1 & q_1 \\ q_1 & 1 & q_1 & 1 \end{pmatrix}. \tag{2.29}$$

This is a degenerate covariance matrix, and in correspondence of the SB solution it follows from the previous equations that the lower bound and the upper bound coincide.

The physical meaning of what happens is qualitatively different for these two phases. Let us take $x < x_c$, where the bounds are tight, and let's start with low α and progressively increase it. In this regime the typical overlap between pairs of solutions is zero, i.e. the two pairs of solutions are independent and there is a positive probability of finding SAT-x-pairs since we are below α_{LB}. When we reach $\alpha = \alpha_{LB} = \alpha_{UB}$ there is a transition to a regime where w.h.p. there exists no pair of solutions to the problem. When this happens the point $(\vec{q}_0^M, \eta^M) \in \tilde{V}_x$ that optimizes (2.21) is the SB point. For $x_c < x < x_c'$, the bounds are no longer tight and we can only identify a region between the two bounds where a SAT/UNSAT transition occurs. Again, for $x_c' < x \leq 0.5$ the bounds are tight. For $x > 0.5$ the behavior is symmetric to the one that we have just described.

2.4 Multiplets of solutions ($y > 2$)

In the previous section we were able to derive rigorous expressions for the upper bound $\alpha_{UB}(x)$, in eq. (2.6), and the lower bound $\alpha_{LB}(x)$, in eq. (2.27), obtained by first and second moment calculations, such that w.h.p. no pairs of solutions at distance x exist for load $\alpha > \alpha_{UB}(x)$ and at least one pair exists for $\alpha < \alpha_{LB}(x)$. It would be then natural to try to generalize the derivation to sets of y solutions at pairwise distance x (multiplets) and in particular asses the existence of a small α regime where such sets can be found for any value of y and for small enough x. This result would rigorously confirm the existence of a dense region of solutions as derived in sec. 2.2, which in turn has been non-rigorously advocated as a necessary condition for the existence of efficient learning algorithms [Baldassi et al., 2015].

Unfortunately, it is technically unfeasible to carry out the rigorous derivation for $y > 2$ as we have done above for the case $y = 2$. Therefore, in this section, after giving an rigorous expression for the first moment upper bound limited to the cases $y = 3$ and $y = 4$, we will derive compact expressions for the first and second moment bound using non-rigorous field theoretical calculations and a replica symmetric ansatz. We find that the non-rigorous results match the rigorous ones when available, although we expect the prediction to break down at large values of y due to replica symmetry breaking effects (see the discussion in the Introduction).

2.4.1 Rigorous first moment upper bounds

In the following we derive the rigorous expressions for the first moment bound in two additional cases: the existence of triplets and quadruplets of solutions at fixed pairwise distance x.

Triplets ($y = 3$)

Let us define the symbol $\cong$ as equivalence up to sub-exponential terms as $N \to \infty$, that is for any two sequences $(a_N)_N$ and $(b_N)_N$ we write $a_N \cong b_N$ iff $\lim_{N\to+\infty} \frac{\ln a_N}{\ln b_N} = 1$. The first moment of the triplets partition function has the following asymptotic expression:

$$\begin{aligned}\mathbb{E}\left[\mathcal{Z}_{y=3}\left(x, K, \boldsymbol{\xi}\right)\right] &\cong 2^{N\binom{N}{\frac{Nx}{2}, \frac{Nx}{2}, \frac{Nx}{2}, N\left(1-\frac{3}{2}x\right)}} \mathbb{P}\left[v_1 \in I_K, v_2 \in I_K, v_3 \in I_K\right]^M \\ &\cong e^{N\left(\ln(2) + H_4(x) + \alpha \ln f_1^{y=3}(x,K)\right)},\end{aligned}$$

where $H_4\left(x\right) = -\frac{3}{2}x \ln\left(\frac{x}{2}\right) - \left(1 - \frac{3}{2}x\right) \ln\left(1 - \frac{3}{2}x\right)$ and we get the geometric condition $0 < x < \frac{2}{3}$, and $f_1^{y=3}\left(x, K\right)$ is the probability that a zero mean Gaussian random vector $\vec{v}_3 = (v_1, v_2, v_3)$, whose covariance matrix Σ_3 has ones on the diagonal and $1 - 2x$ off-diagonal, takes values in the box $[-K, K]^3$, that is

$$f_1^{y=3}\left(x, K\right) = \frac{1}{\left(2\pi\right)^{\frac{3}{2}} \left|\Sigma_3\right|^{1/2}} \int_{[-K,K]^3} dv_1 dv_2 dv_3 \, e^{-\vec{v}_3^T \Sigma_3^{-1} \vec{v}_3}.$$

An equivalent argument to the case $y = 2$ gives the following upper bound for the existence of clusters of three solutions:

$$\alpha_{UB}^{y=3}\left(x, K\right) = -\frac{\ln 2 + H_4\left(x\right)}{\ln f_1^{y=3}\left(x, K\right)}. \tag{2.30}$$

Quadruplets ($y = 4$)

For quadruplets of solutions, we have

$$\mathbb{E}\left[\mathcal{Z}_{y=4}\left(x, K, \boldsymbol{\xi}\right)\right] \cong 2^N \sum_{\mathbf{a} \in V_{N,x}^{y=4} \cap \{0,1/N,2/N,\ldots,1\}^8} \frac{N!}{\prod_{i=0}^7 \left(Na_i\right)!} \left[f_1^{y=4}\left(x, K\right)\right]^M, \tag{2.31}$$

where:

- In complete analogy with the previous case $f_1^{y=4}\left(x, K\right)$ is the probability that a zero mean Gaussian random vector $\vec{v}_4 = (v_1, v_2, v_3, v_4)$, whose covariance matrix Σ_4 has ones on the diagonal and $1 - 2x$ off-diagonal, takes values in the box $[-K, K]^4$, that is

$$f_1^{y=4}\left(x, K\right) = \frac{1}{\left(2\pi\right)^2 \left|\Sigma_4\right|^{1/2}} \int_{[-K,K]^4} d\vec{v}_4 \, e^{-\vec{v}_4^T \Sigma_4^{-1} \vec{v}_4}. \tag{2.32}$$

- The summation is restricted to the set $V_{N,x}^{y=4} \subseteq [0,1]^8$, specified by:

$$\begin{cases} \lfloor N\left(a_4+a_5+a_6+a_7\right)\rfloor = \lfloor Nx\rfloor \\ \lfloor N\left(a_1+a_2+a_5+a_6\right)\rfloor = \lfloor Nx\rfloor \\ \lfloor N\left(a_2+a_3+a_6+a_7\right)\rfloor = \lfloor Nx\rfloor \\ \lfloor N\left(a_1+a_3+a_5+a_7\right)\rfloor = \lfloor Nx\rfloor \\ \lfloor N\left(a_2+a_3+a_4+a_5\right)\rfloor = \lfloor Nx\rfloor \\ \lfloor N\left(a_1+a_3+a_4+a_6\right)\rfloor = \lfloor Nx\rfloor \\ \sum_{i=0}^{7} a_i = 1 \end{cases} . \tag{2.33}$$

In the limit $N \to \infty$, due to the 7 constraints in eq. (2.33), the summation over elements in the box $[0,1]^8$ in eq. (2.31) can be replaced by an integral over the interval

$$\mathcal{B}_x \equiv \left[0,\ \min\left\{\frac{x}{2}, 1-\frac{3}{2}x\right\}\right] \qquad \text{for } x < \frac{2}{3}, \tag{2.34}$$

while for $x > \frac{2}{3}$ the constraints admit no solutions and $\mathbb{E}\left[\mathcal{Z}_{y=4}\left(x,K,\boldsymbol{\xi}\right)\right] \cong 0$.

Therefore, for $x < \frac{2}{3}$, we can write

$$\begin{aligned} &\mathbb{E}\left[\mathcal{Z}_{y=4}\left(x,K,\boldsymbol{\xi}\right)\right] \\ &\quad\cong 2^N \int_{\mathcal{B}_x} db \binom{N}{N\left(1-b-\frac{3}{2}x\right), Nb, Nb, N\left(\frac{x}{2}-b\right), Nb, N\left(\frac{x}{2}-b\right), N\left(\frac{x}{2}-b\right), Nb} \\ &\qquad f_1^{y=4}\left(x,K\right) \\ &\quad\cong \int_{\mathcal{B}_x} db\, e^{N\left(\ln 2 + H_8(x,b) + \ln f_1^{y=4}(x,K)\right)} \\ &\quad\cong e^{N\left(\ln 2 + H_8(x,b^*(x)) + \ln f_1^{y=4}(x)\right)}, \end{aligned}$$

where in the last line we estimated the integral with its saddle point contribution at $b^\star\left(x\right) = \text{argmax}_{b\in\mathcal{B}_x} H_8\left(x,b\right)$. The function $H_8\left(x,b\right)$ is the Shannon entropy of an eight-states discrete probability distribution with masses given by the components of the following vector

$$\left(1-b-3/2\ x, b, b, x/2-b, b, x/2-b, x/2-b, b\right) .$$

It follows that the first moment upper bound to the storage capacity for quadruplets of solutions at a fixed distance x is given by

$$\alpha_{UB}^{y=4}\left(x,K\right) = -\frac{\ln 2 + H_8\left(x, b^\star\left(x\right)\right)}{\ln f_1^{y=4}\left(x,K\right)}.$$

The numerical evaluation of the two upper bounds, $y=3$ and $y=4$, can be found in fig. 2.3 along with the predictions for the upper bound from non-rigorous calculations for larger y's.

2.4.2 Upper bounds under symmetric assumption for saddle point

Since a rigorous expression for the upper bound $\alpha_{UB}^{y}(x,K)$ for $y>4$ is hard to derive, due to highly non-trivial combinatorial factors, we resort to non-rigorous field theoretical techniques and replica symmetric ansatz to obtain an expression that we believe to be exact for low values of y but is likely slightly incorrect for very large y due to replica symmetry breaking effects. The generic computation of the n-th moment of the partition function, $\mathbb{E}\left[\mathcal{Z}_y^n\right]$, is shown in Appendix 2.6.3. Here we present the final result for the first moment bound, i.e. the case $n=1$.

In what follows, we denote with SP the saddle point operation, and we use the overlap between solutions $q_1 \equiv 1-2x$ as our control parameter instead of the distance x to match the usual notation of replica theory calculations. Up to subleading terms in N as $N\to\infty$ we have:

$$\mathbb{E}\left[\mathcal{Z}_y(q_1,K,\boldsymbol{\xi})\right] \;\cong\; e^{N\left(\underset{\hat{q}_1}{\mathrm{SP}}\left\{G_{IS}^{n=1,y}(q_1,\hat{q}_1)\right\}+\alpha G_E^{n=1,y,K}(q_1)\right)},$$

where

$$\begin{aligned} G_{IS}^{n=1,y}(q_1,\hat{q}_1) &= -q_1\hat{q}_1\frac{y(y-1)}{2}-\frac{\hat{q}_1 y}{2}+\ln\int Dt\left(2\cosh\left(t\sqrt{\hat{q}_1}\right)\right)^y \\ G_E^{n=1,y,K}(q_1) &= \ln\int Dz\left[\sum_{s=\pm1} s\,H\left(\frac{-s\,K}{\sqrt{1-q_1}}+\frac{\sqrt{q_1}z}{\sqrt{1-q_1}}\right)\right]^y \end{aligned}$$

where we have used the shorthand notation for standard Gaussian integrals $Dz \equiv dz\frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}}$, and the definition $H(x)=\int_x^\infty Dz=\frac{1}{2}\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right)$.

The first moment bound therefore implies that in the limit $N\to\infty$ there are no SAT-x multiplets of y solutions if

$$\underset{\hat{q}_1}{\mathrm{SP}}\left\{G_{IS}^{n=1,y}(q_1,\hat{q}_1)\right\}+\alpha G_E^{n=1,y,K}(q_1)<0. \tag{2.35}$$

This leads to an estimation $\alpha_{UB,S}^{y}$ given by the symmetric saddle point of the true upper bound α_{UB}^{y} that takes the form

$$\alpha_{UB,S}^{y}(q_1,K)\equiv-\frac{\underset{\hat{q}_1}{\mathrm{SP}}\left\{G_{IS}^{n=1,y}(q_1,\hat{q}_1)\right\}}{G_E^{n=1,y,K}(q_1)}. \tag{2.36}$$

These expressions are derived under a symmetric ansatz (i.e. we restrict the search for the saddle point to a particular subset of the region of integration) and thus are not rigorous; yet the results in the cases $y=2,3,4$ agree with the rigorous ones. The corresponding curves are shown in fig. 2.3. Notice that for some values and y and x, the second moment upper bound is larger than the critical value for the single (and less

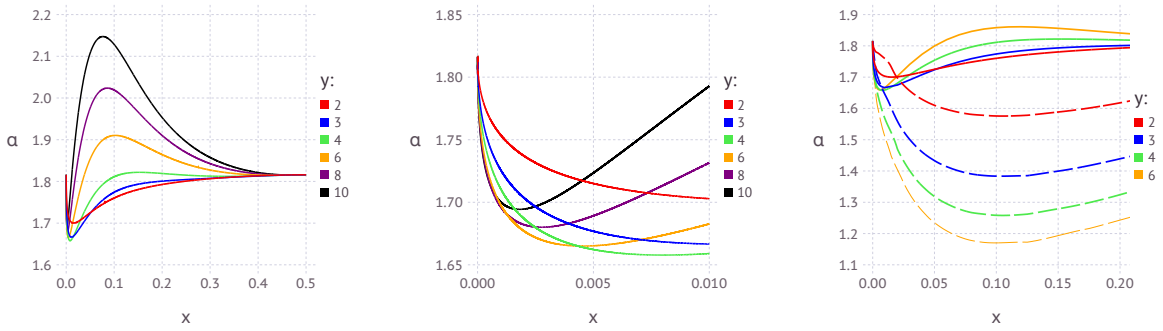

Figure 2.3: (Left) Upper bound $\alpha^{y}_{UB}(x, K=1)$ to the SAT/UNSAT threshold for the RBP problem with y replicas constrained at pairwise distance x. Curves are given by rigorous derivation ($y = 2, 3, 4$) or by non-rigorous field theoretical calculations (2.36) ($y > 4$) . (Center) Zoom of the figure on the left. Close to $x = 0$ the curves corresponding to different y intersect. (Right) The upper bounds (solid lines) are compared to the S point predictions (2.40) for the lower bounds (dashed lines).

constrained) system , $\alpha^{y}_{UB}(x) > \alpha_c$. Since the replicated system critical value, if exist, is such that $\alpha^{y}_{c}(x) \leq \alpha_c$, in that parameter region the upper bound is not tight.

As one can see, the curves intersect in a nontrivial way. Let's take for example the curves for $y = 2$ and $y = 3$. If the bounds were tight for all values of x, the curve at $y = 3$ should always stay below the curve for $y = 2$. This follows directly from the fact that if we have no way of accommodating pairs of solutions then we do not have a way to accommodate triplets solutions either. Instead, the fact that the curves intersect means that for values of x smaller than the intersection point the bounds stop being tight. This straightforward argument, generalized to higher values of y, therefore we can define a tighter upper bound, that we call $\tilde{\alpha}^{y}_{UB}(x, K)$, for the existence of sets of y constrained solutions:

$$\tilde{\alpha}^{y}_{UB}(x, K) = \min\left\{\alpha^{y'}_{UB}(x, K) \,:\, y' \in \mathbb{N},\, 2 \leq y' \leq y\right\}. \tag{2.37}$$

2.4.3 Lower bounds under symmetric assumption for the saddle point

We compute the partition function moments needed for the lower bounds in Appendix 2.6.3. The final result of the replica calculation is given by

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{Z}_y^2\left(q_1, K, \boldsymbol{\xi}\right)\right] \\
&\quad \cong \exp\left(N \cdot \underset{q_0\ \hat{q}_0\ \hat{q}_1}{\mathrm{SP}} \left\{ -\hat{q}_1 y - y\left[y q_0 \hat{q}_0 + (y-1) q_1 \hat{q}_1\right] \right.\right. \\
&\qquad + \ln \int Dz \left[\int Dt \left(2\cosh\left(\sqrt{\hat{q}_0} z + \sqrt{\hat{q}_1 - \hat{q}_0}\, t\right)\right)^y\right]^2 \\
&\qquad \left.\left. + \alpha \ln \int Dz \left[\int Dt \left[\sum_{s=\pm 1} s\, H\left(\frac{-s\,K}{\sqrt{1-q_1}} + \frac{\sqrt{q_0} z + \sqrt{q_1 - q_0}\, t}{\sqrt{1-q_1}}\right)\right]^y\right]^2 \right\}\right) \\
&\quad = \exp\left(N \cdot \underset{q_0\ \hat{q}_0\ \hat{q}_1}{\mathrm{SP}} \left\{ G_I^{n=2,y}\left(q_0, \hat{q}_0, q_1, \hat{q}_1\right) + G_S^{n=2,y}\left(\hat{q}_0, \hat{q}_1\right) + \alpha G_E^{n=2,y,K}\left(q_0, q_1\right) \right\}\right) ,
\end{aligned}
$$

where

$$
\begin{aligned}
G_I^{n=2,y}\left(q_0, \hat{q}_0, q_1, \hat{q}_1\right) &= -\hat{q}_1 y - y\left[y q_0 \hat{q}_0 + (y-1) q_1 \hat{q}_1\right] \\
G_S^{n=2,y}\left(\hat{q}_0, \hat{q}_1\right) &= \ln \int Dz \left[\int Dt \left(2\cosh\left(\sqrt{\hat{q}_0} z + \sqrt{\hat{q}_1 - \hat{q}_0}\, t\right)\right)^y\right]^2 \\
G_E^{n=2,y,K}\left(q_0, q_1\right) &= \ln \int Dz \left[\int Dt \left[\sum_{s=\pm 1} s\, H\left(\frac{-s\,K}{\sqrt{1-q_1}} + \frac{\sqrt{q_0} z + \sqrt{q_1 - q_0}\, t}{\sqrt{1-q_1}}\right)\right]^y\right]^2 .
\end{aligned}
$$

Performing the saddle points over the variables $\hat{q}_0$ and $\hat{q}_1$, these expressions reduce to

$$
\mathbb{E}\left[\mathcal{Z}_y^2\left(q_1, K, \boldsymbol{\xi}\right)\right] \ \simeq\ e^{N\left(\max_{q_0}\left\{G_{IS}^{\mathrm{opt},n=2,y}\left(q_0,q_1\right) + \alpha G_E^{n=2,y,K}\left(q_0,q_1\right)\right\}\right)}, \tag{2.38}
$$

where

$$
G_{IS}^{\mathrm{opt},n=2,y}\left(q_0, q_1\right) = \underset{\hat{q}_0\ \hat{q}_1}{\mathrm{SP}} \left\{ G_I^{n=2,y}\left(q_0, \hat{q}_0, q_1, \hat{q}_1\right) + G_S^{n=2,y}\left(\hat{q}_0, \hat{q}_1\right) \right\} . \tag{2.39}
$$

For fixed α, if the optimum in eq. (2.38) is at $q_0 = 0$ we have

$$
\mathbb{E}\left[\mathcal{Z}_y^2\left(q_1, K, \boldsymbol{\xi}\right)\right] \cong \mathbb{E}\left[\mathcal{Z}_y\left(q_1, K, \boldsymbol{\xi}\right)\right]^2 ,
$$

and from the second moment inequality, eq. (2.9), we have that there is positive probability of finding multiplets of y solutions at distance $x = \frac{1}{2}\left(1 - q_1\right)$. This in turn implies that the lower bound is valid for all α's such that

$$
\underset{q_0}{\mathrm{argmax}} \left\{ G_{IS}^{opt,n=2,y}\left(q_0, q_1\right) + \alpha G_E^{n=2,y,K}\left(q_0, q_1\right) \right\} = 0 .
$$

In particular the symmetric saddle point prediction for the lower bound is given by

$$\alpha^{y}_{LB,S}(q_1) = \sup\left\{\alpha \geq 0 \,\middle|\, \underset{q_0}{\text{argmax}}\left\{G^{opt,n=2,y}_{IS}(q_0,q_1) + \alpha G^{n=2,y,K}_{E}(q_0,q_1)\right\} = 0\right\}. \tag{2.40}$$

The results for $y = 2,3,4,5$ are summarized in fig. 2.3 on the right. In fig. 2.4 we plot an enlargement of the small-distance region around $x = 0$, corresponding to $q_1 = 1$. We find that in all cases there is an inconsistency region $[0, x_c(y)]$ in which the symmetric lower and upper bounds switch roles, similarly to what happened in the case of $y = 2$ (see fig. 2.2). The true lower bound cannot thus be symmetric in this region: the configuration in which the two SAT-x multiplets of y solutions are collapsed on a single multiplet always gives a better saddle point, resulting in a lower bound equal to the upper bound. We thus conjecture that for $x < x_c(y)$ the bounds are tight, like in the $y = 2$ case. The symmetry of lower and upper bounds with respect to x on the interval $[0,1]$ (or the corresponding symmetry for q_1) which holds for $y = 2$ does not apply to general y. In our numerical exploration presented in fig. 2.3, we focused on the region of small x. We also notice that the lower bounds for increasing y's decrease monotonically, and in the limit $y \to \infty$ the limiting curve seem to exhibit a vertical asymptote for $x = 0$. Furthermore, the intersection point $x_c(y)$ seem to decrease monotonically with y and to approach zero. It is also worth noting that, for all the y that we tested, we found that in the region $[0, x_c(y)]$ we have $\tilde{\alpha}^{y}_{UB} = \alpha^{y}_{UB}$, which is consistent with the conjecture that the bounds are tight in this region.

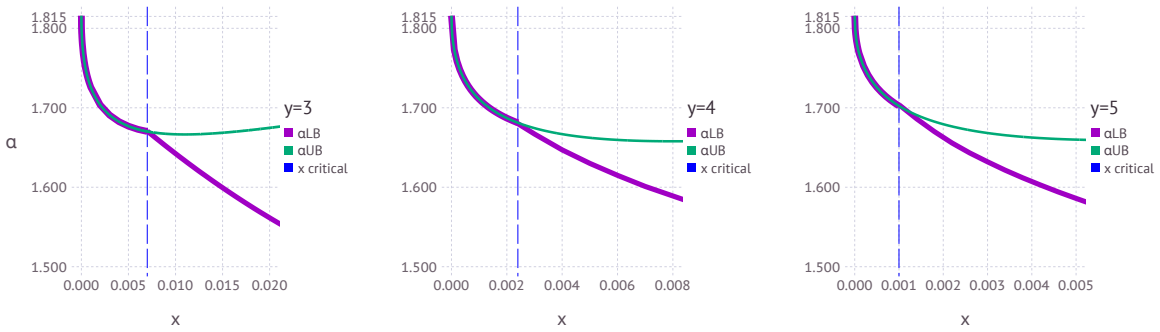

Figure 2.4: Lower and upper bounds for the RBP with $K = 1$ and for different values of $y = 3,4,5$, in the region of small x. Like in the case of $y = 2$, for x larger than the critical value $x_c(y)$ (blue vertical line) there is a gap between the symmetric lower bound (purple line) and the upper bound (green line). This gap closes in correspondence of the SB solution for $x \leq x_c(y)$ and the two bounds coincide.

2.5 Conclusions

We have presented an investigation of the geometry of the solutions space for the binary symmetric perceptron model storing random patterns. According to the non-rigorous analysis conducted with the replica method, this model exhibits the same qualitative phenomenology as the more standard non-symmetric counterpart. In particular, we focused on signatures for the presence of rare dense regions of solutions, which are

of particular interest since according to previous studies they appear to be crucially connected to the existence of efficient learning algorithms [Baldassi et al., 2015, 2016a]. The analogous structures for continuous models (of the kind used for deep learning applications) are wide flat minima, which have also been related to training efficiency and generalization capabilities [Baldassi et al., 2020b].

Compared to standard models, the symmetry in the model used for this paper simplifies the analytical treatment, as was first shown in ref. [Aubin et al., 2019]. Thanks to this, we have been able to show rigorously (up to a numerical optimization step) that in the large N limit there exist an exponential number of pairs of solution at arbitrary $O(N)$ Hamming distance. A further analysis led us to conjecture that this scenario extends to multiplets of more than 2 solutions at fixed distance. These results are highly non-trivial, and consistent with the replica analysis; a complete and rigorous confirmation will presumably require different tools or alternative approaches however, and thus remains as an open problem. Besides this, several other important problems related to the dense regions, with potentially far-fetching practical and theoretical implications, remain open: in particular, obtaining a detailed description of their geometry, and a complete characterization of their accessibility by efficient algorithms.

2.6 Appendix

2.6.1 $y \to \infty$ limit

In this Section we derive the large y limit of the entropy

$$\phi_y(x, K, \alpha) = \lim_{N\to\infty} \frac{1}{yN} \mathbb{E}_\xi \ln \mathcal{Z}_y(x, K, \boldsymbol{\xi})$$

within RS assumptions. For convenience of notation we will use the overlap $q_1 = 1 - 2x$ instead of x. As explained in ref. [Baldassi et al., 2020b], the computation of $\phi_y(x)$ is formally equivalent to that of a single replica in the 1RSB ansatz with Parisi parameter y, except for the fact that q_1 is fixed externally instead of being optimized as usual. We obtain the following entropy for the Random Binary Perceptron (RBP) with y real replicas:

$$\begin{aligned}\phi_y(x, K, \alpha) = \underset{q_0\ \hat{q}_0\ \hat{q}_1}{\mathrm{SP}} \Bigg\{ & -\frac{\hat{q}_1}{2}(1 - q_1) + \frac{y}{2}(q_0\hat{q}_0 - q_1\hat{q}_1) \\ & + \frac{1}{y}\int Dz_0 \ln \int Dz_1 \left[2\cosh\left(\sqrt{\hat{q}_0}z_0 + \sqrt{\hat{q}_1 - \hat{q}_0}z_1\right)\right]^y \\ & + \frac{\alpha}{y}\int Dz_0 \ln \int Dz_1 \left[\sum_{s=\pm 1} s\, H\left(\frac{-s\,K}{\sqrt{1-q_1}} + \frac{\sqrt{q_0}z_0 + \sqrt{q_1 - q_0}z_1}{\sqrt{1-q_1}}\right)\right]^y \Bigg\}.\end{aligned}$$

We want to take the limit $y \to \infty$ in the previous expression. By looking at the entropic and energetic parts we derive the appropriate scalings

$$\hat{q}_0 = \hat{q}_1 - \frac{\delta\hat{q}}{y}, \quad q_0 = q_1 - \frac{\delta q}{y}, \tag{2.41}$$

and the previous equation becomes

$$\phi_{y=\infty}(q_1, K, \alpha) = \underset{\delta q \hat{q}_1 \delta\hat{q}}{\mathrm{SP}} \left\{ -\frac{\hat{q}_1}{2}(1-q_1) - \frac{1}{2}(\delta q \hat{q}_1 + \delta\hat{q} q_1) + \int Dz_0\, A^\star(z_0) + \alpha \int Dz_0\, B^\star(z_0) \right\},$$

where

$$A^\star(z_0) = \ln 2 - \min_{z_1} \left\{ \frac{z_1^2}{2} - \ln\cosh\left(\sqrt{\hat{q}_1} z_0 + \sqrt{\delta\hat{q}} z_1\right) \right\}, \tag{2.42}$$

$$B^\star(z_0) = -\min_{z_1} \left\{ \frac{z_1^2}{2} - \ln\left[\sum_{s=\pm 1} s\, H\left(\frac{-s\,K}{\sqrt{1-q_1}} + \frac{\sqrt{q_1} z_0 + \sqrt{\delta q} z_1}{\sqrt{1-q_1}} \right) \right] \right\}. \tag{2.43}$$

The results are shown in fig. 2.1. The behavior of $\phi_{y=\infty}(q_1)$ close to $q_1 = 1$, where it approaches the maximum volume curve, reveals the existence of a dense cluster of solutions. Furthermore, the maximum volume curve coincides with the curve for $\alpha = 0$, which means that there are no constraints to impose and the function $\phi_{y=\infty}(q_1, K, 0) = H_2\left(\left(1+\sqrt{q_1}\right)/2\right)$. We expect the value obtained within the RS ansatz for $\phi_{y=\infty}(q_1, K, \alpha)$ to not be the correct one, at least for α above some critical value where spin glass instabilities arise. In fact $\phi_{y=\infty}(q_1, K, \alpha)$ yields a SAT/UNSAT transition that is wrong, since it is above the known one for the standard $y = 1$ model. Therefore this scenario should be checked within a 1RSB calculation, where we also expect the dense cluster prediction to remain true. We refer to [Baldassi et al., 2015] for an in-depth analysis of a similar model which takes also into account replica symmetry breaking corrections.

2.6.2 Derivation of the lower bound

Change of integration variables in second moment bound

The bound in eq. (2.17) depends on the 8 variables $\mathbf{a}$. We want now to reduce the number of from 8 to 5 using the constraints in eq. (2.15). We choose to write a_0, a_6, a_7 as functions of the other variables

$$\begin{cases} a_0 = 1 - a_1 - a_2 - a_3 - x \\ a_6 = x - a_1 - a_2 - a_5 \\ a_7 = a_1 + a_2 - a_4 \end{cases}. \tag{2.44}$$

The integration set V_x is then reparametrized as a function of the variables $\vec{a}$ which are defined as $\vec{a} := (a_1, a_2, a_3, a_4, a_5)$. We indicate with $\mathbf{a}(\vec{a}, x)$ the immersion from $\mathbb{R}^5$ to $\mathbb{R}^8$ whose components from a_1 to a_5 are mapped in themselves while the remaining ones are specified by the equations in (2.44). This makes the expression of V_x more explicit and lets us rewrite the integral in an equivalent way. The integration set becomes

$V_x' \subseteq [0,1]^5$ and it is specified by the following set of inequalities:

$$\begin{cases} 0 \leq a_i \leq 1 \ \forall \ i = 1, \dots, 5 \\ 0 \leq a_1 + a_2 - a_4 \leq 1 \\ a_1 + a_2 + a_5 \leq x \\ a_1 + a_2 + a_3 \leq 1 - x \end{cases} . \tag{2.45}$$

With this change of variables eq. (2.17) becomes:

$$\mathbb{E}\left[\mathcal{Z}_{y=2}^2(x, K, \boldsymbol{\xi})\right] \leq C_0 N^{3/2} \int_{V_x'} d\vec{a} \, e^{N[\ln 2 + H_8(\vec{a}, x) + \alpha \ln f_2(\vec{a}, x, K)]}, \tag{2.46}$$

where we defined $H_8(\vec{a}, x) := H_8(\mathbf{a}(\vec{a}, x))$ and $f_2(\vec{a}, x, K) := f_2(\mathbf{a}(\vec{a}, x), x, K)$. The covariance matrix in the Gaussian integral $f_2(\vec{a}, x, K)$ is reparameterized in the following way (cf. eq. (2.12)):

$$\Sigma = \begin{pmatrix} 1 & q_1 & q_{01} & q_{02} \\ q_1 & 1 & q_{03} & q_{04} \\ q_{01} & q_{03} & 1 & q_1 \\ q_{02} & q_{04} & q_1 & 1 \end{pmatrix} \quad \text{where} \quad \begin{cases} q_1 = 1 - 2x \\ q_{01} = 1 - 2(x + a_2 + a_3 - a_4 - a_5) \\ q_{02} = 1 - 2(2a_1 + a_2 + a_3 - a_4 + a_5) \\ q_{03} = 1 - 2(a_2 + a_3 + a_4 + a_5) \\ q_{04} = 1 - 2(x - a_2 + a_3 + a_4 - a_5) \end{cases} . \tag{2.47}$$

The next and final reparametrization of the integral is suggested by the form of the covariance matrix. In particular we would like to express the four possible overlaps between the two pairs of solution using the four parameters q_{01}, q_{02}, q_{03}, q_{04} and group them in a four dimensional vector $\vec{q}_0$. However, since our integration domain is 5-dimensional, we need an additional parameter that we call η. Inverting the under-parametrized system of eqs. (2.47), we obtain the vectors $\vec{a}^\star$ that lie in the vector space below, for $\eta \in \mathbb{R}$:

$$\begin{cases} a_1^\star = \frac{1}{4}(q_{01} - q_{02} + 2x) - \eta \\ a_2^\star = \frac{1}{4}(-q_{03} + q_{04} + 2x) - \eta \\ a_3^\star = \frac{1}{4}(2 - q_{01} - q_{04} - 4x) + \eta \\ a_4^\star = \frac{1}{4}(q_{01} - q_{03} + 2x) - \eta \\ a_5^\star = \eta \end{cases} . \tag{2.48}$$

By constraining the solutions $\vec{a}^\star$ in their natural domain V_x' we find how the domain is transformed in the new coordinates $\vec{q}_0$ and η:

$$\begin{cases} \frac{1}{4}(q_{01} - q_{02} + 2x - 4) \leq \eta \leq \frac{1}{4}(q_{01} - q_{02} + 2x) \\ \frac{1}{4}(-q_{03} + q_{04} + 2x - 4) \leq \eta \leq \frac{1}{4}(-q_{03} + q_{04} + 2x) \\ \frac{1}{4}(q_{01} + q_{04} + 4x - 2) \leq \eta \leq \frac{1}{4}(q_{01} + q_{04} + 4x + 2) \\ \frac{1}{4}(q_{01} - q_{03} + 2x - 4) \leq \eta \leq \frac{1}{4}(q_{01} - q_{03} + 2x) \\ 0 \leq \eta \leq 1 \\ \frac{1}{4}(q_{01} - q_{02} - q_{03} + q_{04}) \leq \eta \\ \frac{1}{4}(-q_{02} + q_{04} + 2x - 4) \leq \eta \leq \frac{1}{4}(-q_{02} + q_{04} + 2x) \\ \frac{1}{4}(-q_{02} - q_{03} + 4x - 2) \leq \eta \end{cases} , \tag{2.49}$$

where we have expressed all inequalities in terms of the variable η. This set of inequalities specifies a new integration domain in eq. (2.46), this time in the new variables η and $\vec{q}_0$, that we call $\tilde{V}_x$ and that depends on x. Again, we can express the vector of solutions $\vec{a}^\star$ as a function of the pair $(\vec{q}_0, \eta)$. The integral (2.46) is rewritten as:

$$\mathbb{E}\left[\mathcal{Z}^2_{y=2}(x,K,\xi)\right] \leq C_0 N^{3/2} \int_{\tilde{V}_x} d\vec{q}_0\, d\eta\, e^{N[\ln 2 + H_8(\vec{q}_0,\eta,x) + \alpha \ln f_2(\vec{q}_0,x,K)]}, \tag{2.50}$$

where we adopt the convention that $f_2(\vec{q}_0, x, K) := f_2(\vec{a}^\star(\vec{q}_0,\eta), x, K)$ and $H_8(\vec{q}_0,\eta,x) := H_8(\vec{a}^\star(\vec{q}_0,\eta), x)$.

Proof of Lemma 3

Proof of Lemma 3. From eq. (2.14) we obtain the following inequalities:

$$\begin{cases} |-a_0+1-a_1-a_2-a_3-x| < \frac{3}{N} \\ |a_6-x+a_1+a_2+a_5| < \frac{1}{N} \\ |a_7-a_1-a_2+a_4| < \frac{2}{N} \end{cases}. \tag{2.51}$$

In the limit $N \to \infty$ these inequalities determine three of the parameters as a function of the other five:

$$\begin{cases} a_0^\star = 1-a_1-a_2-a_3-x \\ a_6^\star = x-a_1-a_2-a_5 \\ a_7^\star = a_1+a_2-a_4 \end{cases}. \tag{2.52}$$

Notice that the summation on the left hand side of eq. (2.16) is taken for

$$\mathbf{a} \in \{0, 1/N, 2/N, \ldots, 1\}^8 .$$

If we fix the five components vector $\vec{a} := (a_1, \ldots, a_5) \in V'_x \cap \{0, 1/N, 2/N, \ldots, 1\}^5$ where V'_x is defined as in eq. (2.45), then, independently from this 5-dimensional vector, there exist at most a fixed number of $\mathbf{a}$'s that satisfy the inequalities in eq. (2.51) (for every N and $x \in [0,1]$). This is sufficient to conclude that for large enough N there exists a positive constant F_0 such that

$$\sum_{\mathbf{a} \in V_{N,x} \cap \{0,1/N,2/N,\ldots,1\}^8} \binom{N}{Na_0\ \ldots\ Na_7} \psi(\mathbf{a})^N$$
$$\leq F_0 \sum_{\vec{a} \in V'_x \cap \{0,1/N,2/N,\ldots,1\}^5} \binom{N}{\lfloor Na_0^\star \rfloor\ Na_1\ \ldots\ Na_5\ \lfloor Na_6^\star \rfloor\ Na_7^\star} \psi(a_0^\star, a_1, \ldots, a_5, a_6^\star, a_7^\star)^N .$$

where V'_x is defined by the system of eqs. (2.45).

From Stirling's approximation, the expression for large N and fixed a_i of the multinomial factor is

$$\binom{N}{Na_0\ \ldots\ Na_m} = e^{NH(\mathbf{a}) - \frac{m-1}{2}\ln N + \mathcal{O}(1)} \leq G_0 e^{NH(\mathbf{a}) - \frac{m-1}{2}\ln N}$$

where G_0 is some positive constant and $H(\mathbf{a})$ is the Shannon entropy of the discrete probability distribution with masses $\{a_0, \ldots, a_m\}$.

Putting all together we have

$$\begin{aligned}
&\sum_{\mathbf{a}\in V_{N,x}\cap\{0,1/N,2/N,\ldots,1\}^8} \binom{N}{Na_0\ \ldots\ Na_7} \psi(\mathbf{a})^N \\
&\leq F_0 \sum_{\vec{a}\in V'_x\cap\{0,1/N,2/N,\ldots,1\}^5} \binom{N}{\lfloor Na_0^\star\rfloor\ Na_1\ \ldots\ Na_5\ \lfloor Na_6^\star\rfloor\ Na_7^\star} \psi\left(a_0^\star, a_1, \ldots, a_6^\star, a_7^\star\right)^N \\
&\leq \frac{F_0 G_0}{N^{\frac{7}{2}}} \sum_{\vec{a}\in V'_x\cap\{0,1/N,2/N,\ldots,1\}^5} e^{NH_8\left(a_0^\star,a_1,\ldots,a_5,a_6^\star,a_7^\star\right)-\frac{7}{2}\ln N} \psi\left(a_0^\star, a_1, \ldots, a_5, a_6^\star, a_7^\star\right)^N \\
&< C_0 N^{\frac{3}{2}} \int_{V_x} d\mathbf{a}\, e^{N[H_8(\mathbf{a})+\ln\psi(\mathbf{a})]},
\end{aligned}$$

where we have used the limit of Riemann sums in the last step and $C_0 > F_0 G_0$ is a positive constant that does not depend on N but depends on x. The integral in the last line is defined as in the footnote for Lemma 3. □

Proof of eq. (2.24)

For finite N we define $\mathcal{N}_2(x)$ and $\mathcal{N}_4(x, \mathbf{a})$ as follows. First,

$$\mathcal{N}_2(x) \equiv \sum_{\{\mathbf{w}^1\}} \sum_{\{\mathbf{w}^2\}} \mathbb{1}\left(d_H\left(\mathbf{w}^1, \mathbf{w}^2\right) = \lfloor Nx\rfloor\right),$$

which implies that

$$\begin{aligned}
\left(\mathcal{N}_2(x)\right)^2 &= \left(\sum_{\{\mathbf{w}^1\}} \sum_{\{\mathbf{w}^2\}} \mathbb{1}\left(d_H\left(\mathbf{w}^1, \mathbf{w}^2\right) = \lfloor Nx\rfloor\right)\right)^2 \\
&= \sum_{\{\mathbf{w}^1\}} \sum_{\{\mathbf{w}^2\}} \sum_{\{\tilde{\mathbf{w}}^1\}} \sum_{\{\tilde{\mathbf{w}}^2\}} \mathbb{1}\left(d_H\left(\mathbf{w}^1, \mathbf{w}^2\right) = \lfloor Nx\rfloor\right)\ \mathbb{1}\left(d_H\left(\tilde{\mathbf{w}}^1, \tilde{\mathbf{w}}^2\right) = \lfloor Nx\rfloor\right).
\end{aligned}$$

Then, for $\mathbf{a} \in V_{N,x}$ we have:

$$\begin{aligned}
\mathcal{N}_4(x, \mathbf{a}) \equiv \sum_{\{\mathbf{w}^1\}} \sum_{\{\mathbf{w}^2\}} \sum_{\{\tilde{\mathbf{w}}^1\}} \sum_{\{\tilde{\mathbf{w}}^2\}} &\mathbb{1}\left(d_H\left(\mathbf{w}^1, \mathbf{w}^2\right) = \lfloor Nx\rfloor\right)\ \mathbb{1}\left(d_H\left(\tilde{\mathbf{w}}^1, \tilde{\mathbf{w}}^2\right) = \lfloor Nx\rfloor\right) \\
&\cdot \mathbb{1}\left(d_H\left(\mathbf{w}^1, \tilde{\mathbf{w}}^1\right) = \lfloor N(a_2 + a_3 + a_6 + a_7)\rfloor\right) \\
&\cdot \mathbb{1}\left(d_H\left(\mathbf{w}^1, \tilde{\mathbf{w}}^2\right) = \lfloor N(a_1 + a_3 + a_5 + a_7)\rfloor\right) \\
&\cdot \mathbb{1}\left(d_H\left(\mathbf{w}^2, \tilde{\mathbf{w}}^1\right) = \lfloor N(a_2 + a_3 + a_4 + a_5)\rfloor\right) \\
&\cdot \mathbb{1}\left(d_H\left(\mathbf{w}^2, \tilde{\mathbf{w}}^2\right) = \lfloor N(a_1 + a_3 + a_4 + a_6)\rfloor\right).
\end{aligned}$$

From the definitions it follows that $\mathcal{N}_4(x, \mathbf{a}) \leq (\mathcal{N}_2(x))^2$ and computing the summations gives

$$2^N \frac{N!}{\prod_{i=0}^{7}(Na_i)!} \leq \left(2^N \binom{N}{\lfloor Nx \rfloor}\right)^2, \quad \forall \boldsymbol{a} \in V_{N,x} \cap \left\{0, \frac{1}{N}, \ldots, 1\right\}^8.$$

Taking the logarithm on both sides, dividing by N and taking the limit for $N \to \infty$, gives the following inequality

$$\ln 2 + H_8(\mathbf{a}) \leq 2 \log 2 + 2H_2(x) \quad \forall \mathbf{a} \in V_{N,x}.$$

If we apply now the same change of variable of Appendix 2.6.2 the result is

$$H_8(\vec{q}_0, \eta, x) \leq \ln 2 + 2H_2(x) \quad \forall (\vec{q}_0, \eta) \in \tilde{V}_x.$$

Numerical optimization

We performed the optimization in expression (2.27) numerically. We empirically find the objective function to be ridden by many local minima, therefore we implemented 3 different strategies to partition the search space and obtain a numerical estimate of the global one.

A first strategy consists in constructing a 4-dimensional uniformly-spaced grid for the values of $\vec{q}_0$, and then performing Gradient Descent (GD) starting from these points and selecting the overall minimum obtained. The downside of this approach is that the the optimization is very time-consuming. We simulated grids with up to $m = 100^4$ number of points. We restrict the experiment to the region of small x, in particular $x < x_c'$. The results are shown in fig. 2.5. While for $x > x_c$, and already for a low numbers of points m, the numerical estimate coincides with the symmetric point prediction, for $x < x_c$ instead, where we predict the broken symmetry point to yield the true value of α_{LB}, only with the two finest grid spacing we are able to get close to the theoretical prediction. Overall, the results for this numerical experiment are in good agreement with theoretical value predicted for the saddle point by symmetry arguments, supporting our conclusion that for $x < x_c$ lower and upper bounds coincide.

Another approach is to restrict the search space to a lower dimensional manifold, containing both the symmetric (S) and the symmetry broken (SB) points. The lower dimensionality (2 instead of 4) allows us to use as starting points of our GD procedure grids with smaller spacings. Therefore, we restrict the search space to points of the type $\vec{q}_0 = (q_a, q_b, q_b, q_a)$. The corresponding covariance matrix in this case is given by

$$\Sigma_{SB} = \begin{pmatrix} 1 & q_1 & q_a & q_b \\ q_1 & 1 & q_b & q_a \\ q_a & q_b & 1 & q_1 \\ q_b & q_a & q_1 & 1 \end{pmatrix}. \tag{2.53}$$

The optimization over this submanifold is done by multiple restarts of GD from a 2-dimensional grid corresponding of values for q_a and q_b. The results are reported in fig. 2.6 (Left). Again, while GD quickly finds the global minima for $x > x_c$, the S point, for $x < x_c$ the global minima SB is more difficult to approach, and the restriction to the

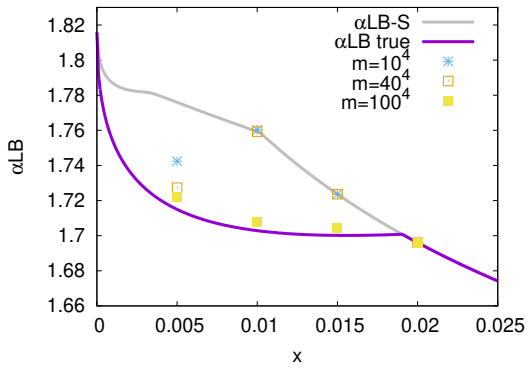

Figure 2.5: Numerical lower bounds $\alpha_{LB,y=2}\left(x, K=1\right)$ obtained by multiple restarts of GD from a 4d grids with m points, for different values of m, along with theoretical predictions from the symmetric point S (that we know to be wrong for $x < x_c$) and the true lower bound (point S for $x > x_c$, point SB for $x < x_c$).

2d submanifold doesn't seem to provide a computational advantage, possibly due to the presence of further spurious minima in this restricted space.

A further approach is to just evaluate the objective function in eq. (2.27) on the points of the increasingly refined 2d-grid, without any GD refinement, and take the lowest of the values obtained. With this approach, we evaluated grids of up to $m = 5000^2$ points. Results are presented in fig. 2.6 (Right).

All of the 3 approaches are in good agreement with each other and with theoretical predictions.

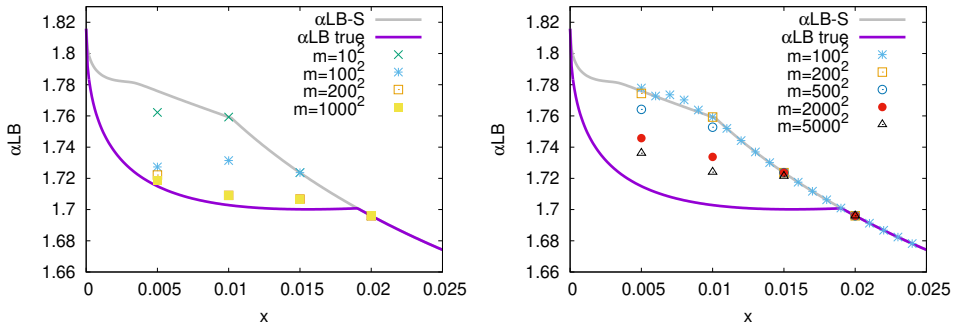

Figure 2.6: (Left) Numerical and theoretical estimates for $\alpha_{LB,y=2}\left(x, K=1\right)$ as in fig. (2.5) but with GD in 2-dimensional space and multiple restarts from grids of m points. (Right) Evaluation of the points in 2d grids of different sizes m with no GD refinement.

Computation of $f_2\left(\vec{q}_0, x, K\right)$

The computation in an efficient and precise way of the quantity $f_2\left(\vec{q}_0, x, K\right)$ is crucial for the numerical results. We use the Cholesky decomposition of matrix $\Sigma = C_L C_L^T$ where C_L is lower triangular and $C_L^{-1} = C_L^T$. Then it is natural to use the change of variable

$\mathbf{y} = C_L^{-1}\mathbf{z}$, in matrix form

$$\begin{pmatrix} z_1 \\ z_2 \\ \tilde{z}_1 \\ \tilde{z}_2 \end{pmatrix} = \begin{pmatrix} c_{11} & 0 & 0 & 0 \\ c_{21} & c_{22} & 0 & 0 \\ c_{31} & c_{32} & c_{33} & 0 \\ c_{41} & c_{42} & c_{43} & c_{44} \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \tilde{y}_1 \\ \tilde{y}_2 \end{pmatrix}$$

the integral is transformed in the following way:

$$\begin{aligned} & f_2\left(\vec{q}_0, x, K\right) \\ & = \int_{I_K^4} \frac{dz_1 dz_2 d\tilde{z}_1 d\tilde{z}_2}{(2\pi)^2 \left|\Sigma\right|^{1/2}} e^{-\frac{1}{2}\mathbf{z}^T\Sigma^{-1}\mathbf{z}} \\ & = \frac{1}{(2\pi)^2} \int_{-\frac{K}{c_{11}}}^{\frac{K}{c_{11}}} dy_1 \int_{\frac{(-K-c_{21}y_1)}{c_{22}}}^{\frac{(K-c_{21}y_1)}{c_{22}}} dy_2 \int_{\frac{(-K-c_{31}y_1-c_{32}y_2)}{c_{33}}}^{\frac{(K-c_{31}y_1-c_{32}y_2)}{c_{33}}} d\tilde{y}_1 \int_{\frac{(-K-c_{41}y_1-c_{42}y_2-c_{43}\tilde{y}_1)}{c_{44}}}^{\frac{(K-c_{41}y_1-c_{42}y_2-c_{43}\tilde{y}_1)}{c_{44}}} d\tilde{y}_2 e^{-\frac{\mathbf{y}^T\mathbf{y}}{2}} \\ & = \frac{1}{(2\pi)^{\frac{3}{2}}} \int_{-\frac{K}{c_{11}}}^{\frac{K}{c_{11}}} dy_1 \int_{\frac{(-K-c_{21}y_1)}{c_{22}}}^{\frac{(K-c_{21}y_1)}{c_{22}}} dy_2 \int_{\frac{(-K-c_{31}y_1-c_{32}y_2)}{c_{33}}}^{\frac{(K-c_{31}y_1-c_{32}y_2)}{c_{33}}} d\tilde{y}_1 e^{-\frac{y_1^2+y_2^2+\tilde{y}_1^2}{2}} \\ & \sum_{s=\pm 1} s\, H\left(\frac{-sK - c_{41}y_1 - c_{42}y_2 - c_{43}\tilde{y}_1}{c_{44}}\right), \end{aligned}$$

where in the last line we have performed the integral over $\tilde{y}_2$, using the definition $H\left(x\right) = \frac{1}{2}\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right)$.

2.6.3 n-th moment of y-solutions multiplet using Replica Ansatz

Let us define $\mathcal{Z}_y$ to be the number of configurations of y vectors of binary weights each satisfying the CSP eq. (2.2) and whose mutual distance is x. In the following we will use the overlap $q_1 = 1 - 2x$ between solutions as an external control parameter. We also introduce for convenience of notation the indicator functions $\varphi_K\left(z\right) = \mathbb{1}\left(\left|z\right| \leq K\right)$ and $\delta\left(z\right) = \mathbb{1}\left(z = 0\right)$. We denote with δ_D the Dirac's delta distribution. With these definitions we have:

$$\begin{aligned} \mathcal{Z}_y\left(q_1, K, \boldsymbol{\xi}\right) &= \sum_{\{\mathbf{w}^a\}_{a=1}^{y}} \prod_{a=1}^{y} \mathbb{X}_{\xi,K}\left(\mathbf{w}^a\right) \prod_{a<b}^{y} \delta\left(\sum_i w_i^a w_i^b - \lfloor N q_1 \rfloor\right) \\ &= \sum_{\{\mathbf{w}^a\}_{a=1}^{y}} \prod_{a=1}^{y} \prod_{\mu=1}^{M} \varphi_K\left(\sum_i w_i^a \xi_i^\mu\right) \prod_{a<b}^{y} \delta\left(\sum_i w_i^a w_i^b - \lfloor N q_1 \rfloor\right). \end{aligned}$$

We want to take the expectation of the n-th moment of this partition function:

$$
\begin{aligned}
&\mathcal{Z}_y^n\left(q_1, K, \boldsymbol{\xi}\right) \\
&= \sum_{\{\mathbf{w}_\alpha^a\}} \prod_{a,\alpha,\mu} \varphi_K\left(\sum_i w_{\alpha,i}^a \xi_i^\mu\right) \prod_{\alpha,a<b} \delta\left(\sum_i w_{\alpha,i}^a w_{\alpha,i}^b - \lfloor N q_1 \rfloor\right) \\
&= \sum_{\{\mathbf{w}_\alpha^a\}} \int \prod_{a,\alpha,\mu} d\lambda_{\alpha,\mu}^a \varphi_K\left(\lambda_{\alpha,\mu}^a\right) \delta_D\left(\lambda_{\alpha,\mu}^a - \sum_i w_{\alpha,i}^a \xi_i^\mu\right) \prod_{\alpha,a<b} \delta\left(\sum_i w_{\alpha,i}^a w_{\alpha,i}^b - \lfloor N q_1 \rfloor\right) \\
&= \sum_{\{\mathbf{w}_\alpha^a\}} \int \prod_{a,\alpha,\mu} \frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi} \varphi_K\left(\lambda_{\alpha,\mu}^a\right) e^{i\hat{\lambda}_{\alpha,\mu}^a \lambda_{\alpha,\mu}^a - i\hat{\lambda}_{\alpha,\mu}^a \sum_i w_{\alpha,i}^a \xi_i^\mu} \prod_{\alpha,a<b} \delta\left(\sum_i w_{\alpha,i}^a w_{\alpha,i}^b - \lfloor N q_1 \rfloor\right).
\end{aligned}
$$

Now we can take the average over the quenched disorder (in the large N limit, up to the leading exponential order):

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{Z}_y^n\left(q_1, K, \boldsymbol{\xi}\right)\right] \\
&\quad= \sum_{\{\mathbf{w}_\alpha^a\}} \int \prod_{a,\alpha,\mu} \left(\frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi} \varphi_K\left(\lambda_{\alpha,\mu}^a\right) e^{i\hat{\lambda}_{\alpha,\mu}^a \lambda_{\alpha,\mu}^a}\right) \mathbb{E}\left[e^{\sum_{\mu,i} \xi_i^\mu \sum_{a,\alpha}\left(-i\hat{\lambda}_{\alpha,\mu}^a w_{\alpha,i}^a\right)}\right] \\
&\qquad \prod_{\alpha,a<b} \delta\left(\sum_i w_{\alpha,i}^a w_{\alpha,i}^b - \lfloor N q_1 \rfloor\right) \\
&\quad\cong \sum_{\{\mathbf{w}_\alpha^a\}} \int \prod_{a,\alpha,\mu} \left(\frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi} \varphi_K\left(\lambda_{\alpha,\mu}^a\right) e^{i\hat{\lambda}_{\alpha,\mu}^a \lambda_{\alpha,\mu}^a}\right) e^{-\frac{1}{2N}\sum_{\mu,i}\left(\sum_{a,\alpha} \hat{\lambda}_{\alpha,\mu}^a w_{\alpha,i}^a\right)^2} \\
&\qquad \prod_{\alpha,a<b} \delta\left(\sum_i w_{\alpha,i}^a w_{\alpha,i}^b - \lfloor N q_1 \rfloor\right) \\
&\quad= \sum_{\{\mathbf{w}_\alpha^a\}} \int \prod_{a,\alpha,\mu} \left(\frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi} \varphi_K\left(\lambda_{\alpha,\mu}^a\right)\right) e^{i\sum_{\alpha,a,\mu} \hat{\lambda}_{\alpha,\mu}^a \lambda_{\alpha,\mu}^a - \frac{1}{2}\sum_\mu \sum_{a,b} \sum_{\alpha,\beta} \hat{\lambda}_{\alpha,\mu}^a \hat{\lambda}_{\beta,\mu}^b \left(\frac{\sum_i w_{\alpha,i}^a w_{\beta,i}^b}{N}\right)} \\
&\qquad \prod_{\alpha,a<b} \delta\left(\sum_i w_{\alpha,i}^a w_{\alpha,i}^b - \lfloor N q_1 \rfloor\right).
\end{aligned}
$$

Next, we introduce the overlaps $q_{\alpha\beta}^{ab} = \frac{\sum_i w_{\alpha,i}^a w_{\beta,i}^b}{N}$ via Dirac deltas:

$$
\begin{aligned}
&\mathbb{E}\left[\mathcal{Z}_y^n\left(q_1, K, \boldsymbol{\xi}\right)\right] \\
&= \sum_{\{\mathbf{w}_\alpha^a\}} \int \prod_{a,\alpha,\mu} \left(\frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi} \varphi_K\left(\lambda_{\alpha,\mu}^a\right)\right) \int \prod_{\alpha<\beta;a,b} dq_{\alpha\beta}^{ab} \int \prod_{\alpha;a<b} dq_{\alpha\alpha}^{ab} e^{i\sum_{\alpha,a,\mu}\hat{\lambda}_{\alpha,\mu}^a\lambda_{\alpha,\mu}^a} \\
&\quad e^{-\sum_\mu \sum_{a,b,\alpha<\beta}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b q_{\alpha\beta}^{ab}} e^{-\sum_\mu\sum_{\alpha,a<b}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b q_1 - \frac{1}{2}\sum_\mu\sum_{a,\alpha}\left(\hat{\lambda}_{\alpha,\mu}^a\right)^2} \\
&\quad \prod_{\alpha<\beta;a,b} \delta_D\left(\frac{\sum_i w_{\alpha,i}^a w_{\beta,i}^b}{N} - q_{\alpha\beta}^{ab}\right) \prod_{\alpha,a<b} \delta_D\left(\frac{\sum_i w_{\alpha,i}^a w_{\alpha,i}^b}{N} - q_{\alpha\alpha}^{ab}\right) \delta\left(Nq_{\alpha\alpha}^{ab} - \lfloor Nq_1\rfloor\right) \\
&\cong \sum_{\{\mathbf{w}_\alpha^a\}} \int \prod_{\alpha<\beta;a,b} \frac{dq_{\alpha\beta}^{ab} d\hat{q}_{\alpha\beta}^{ab}}{2\pi} \int \prod_{\alpha;a<b} \frac{d\hat{q}_{\alpha\alpha}^{ab}}{2\pi} \int \prod_{a,\alpha,\mu} \left(\frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi} \varphi_K\left(\lambda_{\alpha,\mu}^a\right)\right) \\
&\quad e^{i\sum_{\alpha,a,\mu}\hat{\lambda}_{\alpha,\mu}^a\lambda_{\alpha,\mu}^a - \sum_\mu\sum_{a,b,\alpha<\beta}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b q_{\alpha\beta}^{ab} - \sum_\mu\sum_{\alpha,a<b}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b q_1 - \frac{1}{2}\sum_\mu\sum_{a,\alpha}\left(\hat{\lambda}_{\alpha,\mu}^a\right)^2 - N\sum_{\alpha<\beta;a,b}\hat{q}_{\alpha\beta}^{ab}q_{\alpha\beta}^{ab}} \\
&\quad e^{\sum_{\alpha<\beta;a,b}\hat{q}_{\alpha\beta}^{ab}\sum_i w_{\alpha,i}^a w_{\beta,i}^b - Nq_1\sum_{\alpha,a<b}\hat{q}_\alpha^{ab} + \sum_{\alpha,a<b}\hat{q}_{\alpha\alpha}^{ab}\sum_i w_{\alpha,i}^a w_{\alpha,i}^b} \\
&= \int \prod_{\alpha<\beta;a,b} \frac{dq_{\alpha\beta}^{ab} d\hat{q}_{\alpha\beta}^{ab}}{2\pi} \prod_{\alpha;a<b} \frac{d\hat{q}_{\alpha\alpha}^{ab}}{2\pi} e^{N\left(G_I(q,\hat{q}) + G_S(\hat{q}) + \alpha G_E(q)\right)},
\end{aligned}
$$

where we have introduced the interaction, entropic and energetic terms:

$$
\begin{aligned}
G_I^{n,y}\left(q,\hat{q}\right) &= -\sum_{\alpha<\beta;a,b}\hat{q}_{\alpha\beta}^{ab}q_{\alpha\beta}^{ab} - q_1\sum_{\alpha,a<b}\hat{q}_\alpha^{ab}, \\
G_S^{n,y}\left(\hat{q}\right) &= \frac{1}{N}\ln\sum_{\{\mathbf{w}_\alpha^a\}} e^{\sum_{\alpha<\beta;a,b}\hat{q}_{\alpha\beta}^{ab}\sum_i w_{\alpha,i}^a w_{\beta,i}^b + \sum_{\alpha,a<b}\hat{q}_{\alpha\alpha}^{ab}\sum_i w_{\alpha,i}^a w_{\alpha,i}^b}, \\
G_E^{n,y,K}\left(q\right) &= \frac{1}{\alpha N}\ln\int \prod_{a,\alpha,\mu}\left(\frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi}\varphi_K\left(\lambda_{\alpha,\mu}^a\right)\right) e^{i\sum_{\alpha,a,\mu}\hat{\lambda}_{\alpha,\mu}^a\lambda_{\alpha,\mu}^a - \sum_\mu\sum_{a,b,\alpha<\beta}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b q_{\alpha\beta}^{ab}} \\
&\qquad e^{-\sum_\mu\sum_{\alpha,a<b}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b q_1 - \frac{1}{2}\sum_\mu\sum_{a,\alpha}\left(\hat{\lambda}_{\alpha,\mu}^a\right)^2}.
\end{aligned}
$$

We introduce a replica-symmetric ansatz on the matrices $Q_{\alpha\beta}$ and $\hat{Q}_{\alpha\beta}$ which is specified by the following set of equations:

$$
Q_{\alpha\beta}^{ab} = \begin{cases} 1 & \text{if } \alpha = \beta \text{ and } a = b \\ q_0 & \text{if } \alpha \neq \beta \\ q_1 & \text{if } \alpha = \beta \text{ and } a \neq b \end{cases} \qquad \hat{Q}_{\alpha\beta}^{ab} = \begin{cases} 0 & \text{if } \alpha = \beta \text{ and } a = b \\ \hat{q}_0 & \text{if } \alpha \neq \beta \\ \hat{q}_1 & \text{if } \alpha = \beta \text{ and } a \neq b \end{cases}.
$$

In the case $y = 3$ and $n = 2$ they look as follows:

$$
Q = \begin{pmatrix} 1 & q_1 & q_1 & q_0 & q_0 & q_0 \\ q_1 & 1 & q_1 & q_0 & q_0 & q_0 \\ q_1 & q_1 & 1 & q_0 & q_0 & q_0 \\ q_0 & q_0 & q_0 & 1 & q_1 & q_1 \\ q_0 & q_0 & q_0 & q_1 & 1 & q_1 \\ q_0 & q_0 & q_0 & q_1 & q_1 & 1 \end{pmatrix} \qquad \hat{Q} = \begin{pmatrix} 0 & \hat{q}_1 & \hat{q}_1 & \hat{q}_0 & \hat{q}_0 & \hat{q}_0 \\ \hat{q}_1 & 0 & \hat{q}_1 & \hat{q}_0 & \hat{q}_0 & \hat{q}_0 \\ \hat{q}_1 & \hat{q}_1 & 0 & \hat{q}_0 & \hat{q}_0 & \hat{q}_0 \\ \hat{q}_0 & \hat{q}_0 & \hat{q}_0 & 0 & \hat{q}_1 & \hat{q}_1 \\ \hat{q}_0 & \hat{q}_0 & \hat{q}_0 & \hat{q}_1 & 0 & \hat{q}_1 \\ \hat{q}_0 & \hat{q}_0 & \hat{q}_0 & \hat{q}_1 & \hat{q}_1 & 0 \end{pmatrix}.
$$

We now compute the interaction, entropic and energetic terms using this ansatz:

$$G_I^{n,y}\left(q_0,q_1,\hat{q}_0,\hat{q}_1\right) \;=\; -y^2\frac{n\left(n-1\right)}{2}q_0\hat{q}_0 - n\frac{y\left(y-1\right)}{2}q_1\hat{q}_1 - \frac{yn}{2}\hat{q}_1\,, \qquad (2.54)$$

$$\begin{aligned}
G_S^{n,y}\left(\hat{q}_0,\hat{q}_1\right) =& \frac{1}{N}\ln\sum_{\{\mathbf{w}_\alpha^a\}}\prod_i e^{\sum_{\alpha<\beta;a,b}\hat{q}_0 w_{\alpha,i}^a w_{\beta,i}^b+\sum_{\alpha,a<b}\hat{q}_1 w_{\alpha,i}^a w_{\alpha,i}^b}\\
=& -\frac{ny\hat{q}_1}{2}+\ln\sum_{\{w_\alpha^a\}} e^{\frac{1}{2}\hat{q}_0\left(\sum_{a\alpha}w_\alpha^a\right)^2+\frac{\hat{q}_1-\hat{q}_0}{2}\sum_\alpha\left(\sum_a w_\alpha^a\right)^2}\\
=& -\frac{ny\hat{q}_1}{2}+\ln\sum_{\{w_\alpha^a\}}\int Dz\, e^{z\sqrt{\hat{q}_0}\sum_{a\alpha}w_\alpha^a}\int\prod_\alpha Dt_\alpha e^{\sqrt{\hat{q}_1-\hat{q}_0}\sum_\alpha t_\alpha\sum_a w_\alpha^a}\\
=& -\frac{ny\hat{q}_1}{2}+\ln\int Dz\left[\int Dt\left(2\cosh\left(\sqrt{\hat{q}_0}z+\sqrt{\hat{q}_1-\hat{q}_0}t\right)\right)^y\right]^n\,, \qquad (2.55)
\end{aligned}$$

$$\begin{aligned}
G_E^{n,y,K}\left(q_0,q_1\right) =& \frac{1}{\alpha N}\ln\int\prod_{a,\alpha,\mu}\left(\frac{d\lambda_{\alpha,\mu}^a d\hat{\lambda}_{\alpha,\mu}^a}{2\pi}\varphi_K\left(\lambda_{\alpha,\mu}^a\right)\right)e^{i\sum_{\alpha,a,\mu}\hat{\lambda}_{\alpha,\mu}^a\lambda_{\alpha,\mu}^a-\sum_\mu q_0\sum_{a,b,\alpha<\beta}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b}\\
& e^{-\sum_\mu\sum_{\alpha,a<b}\hat{\lambda}_{\alpha,\mu}^a\hat{\lambda}_{\beta,\mu}^b q_1-\frac{1}{2}\sum_\mu\sum_{a,\alpha}\left(\hat{\lambda}_{\alpha,\mu}^a\right)^2}\\
=& \ln\int\prod_{a,\alpha}\left(\frac{d\lambda_\alpha^a d\hat{\lambda}_\alpha^a}{2\pi}\varphi_K\left(\lambda_\alpha^a\right)\right)e^{i\sum_{\alpha,a}\hat{\lambda}_\alpha^a\lambda_\alpha^a-\frac{1}{2}q_0\left(\sum_{a\alpha}\hat{\lambda}_\alpha^a\right)^2-\frac{q_1-q_0}{2}\sum_\alpha\left(\sum_a\hat{\lambda}_\alpha^a\right)^2-\frac{1-q_1}{2}\sum_{a\alpha}\left(\hat{\lambda}_\alpha^a\right)^2}\\
=& \ln\int Dz\int\prod_\alpha Dt_\alpha\int\prod_{a\alpha}\left(\frac{d\lambda_\alpha^a d\hat{\lambda}_\alpha^a}{2\pi}\varphi_K\left(\lambda_\alpha^a\right)\right)e^{i\sum_{\alpha,a}\hat{\lambda}_\alpha^a\lambda_\alpha^a}\\
& e^{iz\sqrt{q_0}\sum_{a\alpha}\hat{\lambda}_\alpha^a+i\sqrt{q_1-q_0}\sum_\alpha t_\alpha\sum_a\hat{\lambda}_\alpha^a-\frac{1-q_1}{2}\sum_{a\alpha}\left(\hat{\lambda}_\alpha^a\right)^2}\\
=& \ln\int Dz\left[\int Dt\left[\int\frac{d\lambda d\hat{\lambda}}{2\pi}\varphi_K\left(\lambda\right)e^{i\hat{\lambda}\lambda+iz\sqrt{q_0}\hat{\lambda}+i\sqrt{q_1-q_0}t\hat{\lambda}-\frac{1-q_1}{2}\hat{\lambda}^2}\right]^y\right]^n\\
=& \ln\int Dz\left[\int Dt\left[\int\frac{d\lambda}{\sqrt{2\pi\left(1-q_1\right)}}\varphi_K\left(\lambda\right)e^{-\frac{\left(\lambda+\sqrt{q_0}z+\sqrt{q_1-q_0}t\right)^2}{2\left(1-q_1\right)}}\right]^y\right]^n\\
=& \ln\int Dz\left[\int Dt\left[\sum_{s=\pm1}s\,H\left(\frac{-s\,K}{\sqrt{1-q_1}}+\frac{\sqrt{q_0}z+\sqrt{q_1-q_0}t}{\sqrt{1-q_1}}\right)\right]^y\right]^n\,.
\end{aligned}$$

In the last line, as in the main text, the function $H\left(x\right)$ is defined as $H\left(x\right)\equiv\int_x^\infty Dz\equiv\int_x^\infty\frac{dz}{\sqrt{2\pi}}e^{-z^2/2}=\frac{1}{2}\mathrm{erfc}\left(\frac{x}{\sqrt{2}}\right)$.

Chapter 3

An Efficient Algorithm for Cooperative Semi-Bandits

3.1 Abstract

We consider the problem of asynchronous online combinatorial optimization on a network of communicating agents. At each time step, some of the agents are stochastically activated, requested to make a prediction, and the system pays the corresponding loss. Then, neighbors of active agents receive semi-bandit feedback and exchange some succinct local information. The goal is to minimize the network regret, defined as the difference between the cumulative loss of the predictions of active agents and that of the best action in hindsight, selected from a combinatorial decision set. The main challenge in such a context is to control the computational complexity of the resulting algorithm while retaining minimax optimal regret guarantees. We introduce Coop-FTPL, a cooperative version of the well-known Follow The Perturbed Leader algorithm, that implements a new loss estimation procedure generalizing the Geometric Resampling of Neu and Bartók [2013] to our setting. Assuming that the elements of the decision set are k-dimensional binary vectors with at most m non-zero entries and α_1 is the independence number of the network, we show that the expected regret of our algorithm after T time steps is of order $Q\sqrt{mkT\log(k)(k\alpha_1/Q+m)}$, where Q is the total activation probability mass. Furthermore, we prove that this is only $\sqrt{k\log k}$-away from the best achievable rate and that Coop-FTPL has a state-of-the-art $T^{3/2}$ worst-case computational complexity.

3.2 Introduction

Distributed online settings with communication constraints arise naturally in large-scale learning systems. For example, in domains such as finance or online advertising, agents often serve high volumes of prediction requests and have to update their local models in an online fashion. Bandwidth and computational constraints may therefore preclude a central processor from having access to all the observations from all sessions and synchronizing all local models at the same time. With these motivations in mind, we introduce and analyze a new online learning setting in which a network of agents solves

efficiently a common nonstochastic combinatorial semi-bandit problem by sharing information only with their network neighbors. At each time step t, some agents v belonging to a communication network $\mathcal{G}$ are asked to make a prediction $x_t(v)$ belonging to a subset $\mathcal{A}$ of $\{0,1\}^k$ and pay a (linear) loss $\langle x_t(v), \ell_t \rangle$ where $\ell_t \in [0,1]^k$ is chosen adversarially by an oblivious environment. Then, any such agent v receives the feedback $\big(x_t(1,v)\ell_t(i), \dots, x_t(k,v)\ell_t(k)\big)$, which is shared, together with some local information, to its first neighbors in the graph. The goal is to minimize the network regret after T time steps

$$R_T = \max_{a \in \mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \langle x_t(v), \ell_t \rangle - \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \langle a, \ell_t \rangle\right], \tag{3.1}$$

where $\mathcal{S}_t$ is the set of agents v that made a prediction at time t. In words, this is the difference between the cumulative loss of the "active" agents and the loss that they would have incurred had they consistently made the best prediction in hindsight.

For this setting, we design a distributed algorithm that we call Coop-FTPL (Algorithm 1), and prove that its regret is upper bounded by $Q\sqrt{mkT\log(k)(k\alpha_1/Q + m)}$ (Theorem 3), where α_1 is the independence number of the network $\mathcal{G}$ and Q is the sum over all agents of the probability that the agent is active during a time step. Our algorithm employs an estimation technique that we call Cooperative Geometric Resampling (Coop-GR, Algorithm 2). It is an extension of a similar procedure appearing in [Neu and Bartók, 2013] that relies on the fact that the reciprocal of the probability of an event can be estimated by measuring the reoccurrence time. We can leverage this idea in the context of cooperative learning thanks to some statistical properties of the minimum of a family of geometric random variables (see Lemmas 4–6). Our algorithm has a state-of-the-art dependence on time of order $T^{3/2}$ for the worst-case computational complexity (Proposition 2). Moreover, we show with a lower bound (Theorem 4) that no algorithm can achieve a regret smaller than $Q\sqrt{mkT\alpha_1/Q}$ on all cooperative semi-bandit instances. Thus, our Coop-FTPL is at most a multiplicative factor of $\sqrt{k\log k}$-away from the minimax result.

To the best of our knowledge, ours is the first computationally efficient near-optimal learning algorithm for the problem of cooperative learning with nonstochastic combinatorial bandits, where not all agents are necessarily active at all time steps.

3.3 Related work and further applications

Single-agent combinatorial bandits find applications in several fields, such as path planning, ranking and matching problems, finding minimum-weight spanning trees, cut sets, and multitask bandits. An efficient algorithm for this setting is Follow-The-Perturbed-Leader (FTPL), which was first proposed by Hannan [1957] and later rediscovered by Kalai and Vempala [2005]. Neu and Bartók [2013] show that combining FTPL with a loss estimation procedure called Geometric Resampling (GR) leads to a computationally efficient solution for this problem. More precisely, the solution is efficient given that the offline optimization problem of finding

$$a^\star = \operatorname*{argmin}_{a \in \mathcal{A}} \langle a, y \rangle\,, \qquad \forall y \in [0,+\infty)^k \tag{3.2}$$

admits a computationally efficient algorithm. This assumption is minimal, in the sense that if the offline problem in Eq. (3.2) is hard to approximate, then any algorithm with low regret must also be inefficient.[1] Grötschel et al. [2012] and Lee et al. [2018] give some sufficient conditions for the validity of this assumption. They essentially rely on having an efficient membership oracle for the convex hull $\mathrm{co}(\mathcal{A})$ of $\mathcal{A}$ and an evaluation oracle for the linear function to optimize. Audibert et al. [2014] note that Online Stochastic Mirror Descent (OSMD) or Follow The Regularized Leader (FTRL)-type algorithms can be efficiently implemented by convex programming if the convex hull of the decision set can be described by a polynomial number of constraints. Suehiro et al. [2012] investigate the details of such efficient implementations and design an algorithm with k^6 time-complexity, which might still be unfeasible in practice. Methods based on the exponential weighting of each decision vector can be implemented efficiently only in a handful of special cases —see, e.g., [Koolen et al., 2010] and [Cesa-Bianchi and Lugosi, 2012] for some examples.

The study of cooperative nonstochastic online learning on networks was pioneered by Awerbuch and Kleinberg [2008], who investigated a bandit setting in which the communication graph is a clique, agents belong to clusters characterized by the same loss, and some agents may be non-cooperative. In our multi-agent setting, the end goal is to control the total network regret (4.1). This objective was already studied by Cesa-Bianchi et al. [2019a] in the full-information case. A similar line of work was pursued by Cesa-Bianchi et al. [2019b], where the authors consider networks of learning agents that cooperate to solve the same nonstochastic bandit problem. In their setting, all agents are simultaneously active at all time steps, and the feedback propagates throughout the network with a maximum delay of d time steps, where d is a parameter of the proposed algorithm. The authors introduce a cooperative version of Exp3 that they call Exp3-COOP with regret of order $\sqrt{(d+1+K\alpha_d/N)(T\log K)}$ where K is the number of arms in the nonstochastic bandit problem, N is the total number of agents in the network, and α_d is the independence number of the d-th power of the communication network. The case $d=1$ corresponds to information that arrives with one round of delay and communication limited to first neighbors. In this setting Exp3-COOP has regret of order $\sqrt{(1+K\alpha_1/N)(T\log K)}$. Thus, our work can be seen as an extension of this setting to the case of combinatorial bandits with stochastic activation of agents. Finally, we point out that if the network consists of a single node, our cooperative setting collapses into a single-agent combinatorial semi-bandit problem. In particular, when the number of arms is k and $m=1$, this becomes the well-known adversarial multiarmed bandit problem (see [Auer et al., 2002]). Hence, ours is a proper generalization of all the settings mentioned above.

Finally, the reader may wonder what kind of results could be achieved if the agents are activated adversarially rather than stochastically. Cesa-Bianchi et al. [2019a] showed that in this setting no learning can occur, not even in with full-information feedback.

[1] A slight relaxation in this direction would be assuming that Eq. (3.2) can be approximated accurately and efficiently.

3.4 Cooperative semi-bandit setting

In this section, we present our cooperative semi-bandit protocol and we introduce all relevant definitions and notation.

We say that $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ is a *communication network* over N agents if it is an undirected graph over a set $\mathcal{V}$ with cardinality $|\mathcal{V}| = N$, whose elements we refer to as *agents*. Without loss of generality, we assume that $\mathcal{V} = \{1, \dots, N\}$. For any agent $v \in \mathcal{V}$, we denote by $\mathcal{N}(v)$ the set of agents containing v and the neighborhood $\{w \in \mathcal{V} : (v, w) \in \mathcal{E}\}$. We say that α_1 is the *independence number* of the network $\mathcal{G}$ if is the largest cardinality of an *independent set* of $\mathcal{G}$, where an independent set of $\mathcal{G}$ is a subset of agents, no two of which are neighbors.

We study the following cooperative online combinatorial optimization protocol. Initially, hidden from the agents, the environment draws a sequence of subsets $\mathcal{S}_1, \mathcal{S}_2, \dots \subset \mathcal{V}$ of agents, that we call *active*, and a sequence of *loss vectors* $\ell_1, \ell_2, \dots \in \mathbb{R}^k$. We assume that each agent v has a probability $q(v)$ of being activated, which need only be known by v. The set of active agents $\mathcal{S}_t$ at time $t \in \{1, 2, \dots\}$ is then determined by drawing, for each agent $v \in \mathcal{V}$, a Bernoulli random variable $X_t(v)$ with bias $q(v)$, independently of the past, and $\mathcal{S}_t$ consists exclusively of agents $v \in \mathcal{V}$ for which $X_t(v) = 1$. The *decision set* is a subset $\mathcal{A}$ of $\left\{a \in \{0,1\}^k : \sum_{i=1}^k a_i \leq m\right\}$, for some $m \in \{1, \dots, k\}$.

For each time step $t \in \{1, 2, \dots\}$:

1. each active agent $v \in \mathcal{S}_t$ predicts with $x_t(v) \in \mathcal{A}$ (possibly drawn at random);
2. each neighbor $v \in \mathcal{N}(w)$ of an active agent $w \in \mathcal{S}_t$ receives the feedback
$$f_t(w) := \big(x_t(1, w)\ell_t(1), \dots, x_t(k, w)\ell_t(k)\big) \; ; \tag{3.3}$$
3. each agent $v \in \bigcup_{w \in \mathcal{S}_t} \mathcal{N}(w)$ receives some local information from its neighbors in $\mathcal{N}(v)$;
4. the system incurs the loss $\sum_{v \in \mathcal{S}_t} \langle x_t(v), \ell_t \rangle$.

The goal is to minimize the expected *network regret* as a function of the *time horizon* T, defined by
$$R_T := \max_{a \in \mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \langle x_t(v), \ell_t \rangle - \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \langle a, \ell_t \rangle\right] , \tag{3.4}$$
where the expectation is taken with respect to the draws of $\mathcal{S}_1, \dots, \mathcal{S}_T$ and (possibly) the randomization of the learners. In the next sections we will also denote by $\mathbb{P}_t$ the probability conditioned to the history up to and including round $t-1$, and by $\mathbb{E}_t$ the corresponding expectation.

The nature of the local information exchanged by neighbors of active agents will be clarified in the next section. In short, they share succinct representations of the current state of their local prediction model.

3.5 Coop-FTPL and upper bound

In this section we introduce and analyze our efficient Coop-FTPL algorithm (Algorithm 1) for cooperative online combinatorial optimization.

Coop-FTPL takes as input a decision set $\mathcal{A} \subset \{0,1\}^k$, a time horizon $T \in \mathbb{N}$, a learning rate $\eta > 0$, a truncation parameter β, and an exploration distribution ζ. At each time step t, all active agents v make a FTPL prediction $x_t(v)$ (line 5) with an i.i.d. perturbation sampled from ζ (line 4), then they receive some feedback $f_t(v)$ and share it with their first neighbors (line 6). Afterwards, each agent who received some feedback this round, requests from its neighbors w a vector $K_t(w)$ of geometric random samples (line 7) which is efficiently computed by Algorithm 2 and will be described in detail later. With these geometric samples, each agent v computes an estimated loss $\widehat{\ell}_t(v)$ (line 8) and updates the cumulative loss estimate $\widehat{L}_t(v)$ (line 9).

Algorithm 1: Follow the perturbed leader for cooperative semi-bandits (Coop-FTPL)

Input: decision set $\mathcal{A}$, horizon T, learning rate η, truncation parameter β, exploration pdf ζ

Initialization: $\widehat{L}_0 = 0 \in \mathbb{R}^k$

1 **for** *each time* $t = 1, 2, \ldots$ **do**

2 **for** *each agent* $v \in \bigcup_{w \in \mathcal{S}_t} \mathcal{N}(w)$ **do** `// neighbors of active agents`

3 **if** $v \in \mathcal{S}_t$ **then** `// active agents`

4 sample $Z_t(v) \sim \zeta$, independently of the past

5 make the prediction $x_t(v) = \operatorname{argmax}_{a \in \mathcal{A}} \langle a, Z_t(v) - \eta \widehat{L}_{t-1}(v) \rangle$

6 receive the feedback $f_t(w)$ (Eq. (3.3)) from each active neighbor $w \in \mathcal{N}(v) \cap \mathcal{S}_t$

7 receive $K_t(i, w)$ ($\forall\, i \in \{1, \ldots, k\}$) from each neighbor $w \in \mathcal{N}(v)$ using Algorithm 2

8 compute $\widehat{\ell}_t(i, v) = \ell_t(i)\, B_t\,(i, v) \min_{w \in \mathcal{N}(v)} \{K_t(i, w)\}$, $\forall i \in \{1, \ldots, k\}$ (Eq. (3.6)–(3.7))

9 update $\widehat{L}_t(v) = \widehat{L}_{t-1}(v) + \widehat{\ell}_t(v)$

Algorithm 2: Geometric resampling for cooperative semi-bandits (Coop-GR)

Input: time step t, component i, agent w, truncation parameter β, exploration pdf ζ

1 **for** $s = 1, 2, \ldots$ **do**

2 sample $z'_s \sim \zeta$, independently of the past

3 let x'_s be the i-th component of the vector $\operatorname{argmax}_{a \in \mathcal{A}} \langle a, z'_s - \eta \widehat{L}_{t-1}(w) \rangle$

4 sample y'_s from a Bernoulli distribution with parameter $q(w)$, independently of the past

5 if ($x'_s = 1$ **and** $y'_s = 1$) **or** $s = \beta$, **break**

6 **return** $K_t(i, w) = s$

Before describing $K_t(w)$ and $\widehat{\ell}_t(v)$, we make a connection between FTPL and the Online Stochastic Mirror Descent algorithm (OSMD)[2] that will be crucial for our analysis.[3]

Fix any time step t and an agent v. As we mentioned above, if v is active, it makes the following FTPL prediction (line 5)

$$x_t(v) = \operatorname*{argmin}_{a \in \mathcal{A}} \langle a, \eta \widehat{L}_{t-1}(v) - Z_t(v) \rangle \,,$$

[2]For a brief overview of some key convex analysis and OSMD facts, see Appendices 3.8.1 and 3.8.2

[3]For a similar approach in the single-agent case, see [Lattimore and Szepesvári, 2020].

where $Z_t(v) \in \mathbb{R}^k$ is sampled i.i.d. from ζ (the random perturbations introduce the exploration, which for an appropriate choice of ζ is sufficient to guarantee small regret). On the other hand, given a Legendre potential F with $\mathrm{dom}\,(\nabla F) = \mathrm{int}\,(\mathrm{co}\,(\mathcal{A}))$, an OSMD algorithm makes the prediction

$$\overline{x}_t(v) = \operatorname*{argmin}_{x \in \mathrm{co}(\mathcal{A})} \Big(\big\langle x, \eta\, \widehat{\ell}_{t-1}(v) \big\rangle + \mathcal{B}_F(x, \overline{x}_{t-1}(v)) \Big) \,,$$

where $\mathcal{B}_F$ is the Bregman divergence induced by F and $\mathrm{co}(\mathcal{A})$ is the convex hull of $\mathcal{A}$. Using the fact that $\mathrm{dom}(\nabla F) = \mathrm{int}(\mathrm{co}(\mathcal{A}))$, the argmin above can be computed in a standard way by studying when the gradient of its argument is equal to zero, and proceeding inductively, we obtain the two identities $\nabla F(\overline{x}_t(v)) = \nabla F(\overline{x}_{t-1}(v)) - \eta\, \widehat{\ell}_{t-1}(v) = -\eta \widehat{L}_{t-1}(v)$. By duality this implies that $\overline{x}_t(v) = \nabla F^*\big(-\eta \widehat{L}_{t-1}(v)\big)$. We now want to relate $x_t(v)$ and $\overline{x}_t(v)$ so that

$$\overline{x}_t(v) = \mathbb{E}_t\big[x_t(v)\big] = \mathbb{E}_t \left[\operatorname*{argmin}_{a \in \mathcal{A}} \big\langle a, \eta \widehat{L}_{t-1}(v) - Z_t(v) \big\rangle \right] \,, \tag{3.5}$$

where the conditional expectation $\mathbb{E}_t$ (given the history up to time $t-1$) is taken with respect to $Z_t(v)$. Thus, in order to view FTPL as an instance of OSMD, it suffices to find a Legendre potential F with $\mathrm{dom}\,(\nabla F) = \mathrm{int}\,(\mathrm{co}\,(\mathcal{A}))$ such that $\nabla F^*\big(-\eta \widehat{L}_{t-1}(v)\big) = \mathbb{E}_t\big[\mathrm{argmax}_{a \in \mathcal{A}} \big\langle a, Z_t(v) - \eta \widehat{L}_{t-1}(v) \big\rangle\big]$. In order to satisfy this condition, we need that for any $x \in \mathbb{R}^k$, the Fenchel conjugate F^* of F enjoys $\nabla F^*(x) = \int_{\mathbb{R}^k} \mathrm{argmax}_{a \in \mathcal{A}} \langle a, z - x \rangle\, \zeta(z)\, \mathrm{d}z$. Then, we define $h(x) := \mathrm{argmax}_{a \in \mathcal{A}} \langle a, x \rangle$ for any $x \in \mathbb{R}^k$, where $h(x)$ is chosen to be an arbitrary maximizer if multiple maximizers exist. From convex analysis, if the convex hull $\mathrm{co}(\mathcal{A})$ of $\mathcal{A}$ had a smooth boundary, then the support function $x \mapsto \phi(x) := \max_{a \in \mathrm{co}(\mathcal{A})} \langle a, x \rangle = \max_{a \in \mathcal{A}} \langle a, x \rangle$, of $\mathrm{co}(\mathcal{A})$ would satisfy $\nabla \phi(x) = h(x)$. For combinatorial bandits, $\mathrm{co}(\mathcal{A})$ is non-smooth, but, being ζ a density with respect to Lebesgue measure, one can prove (see, e.g., Lattimore and Szepesvári [2020]) that $\nabla \int_{\mathbb{R}^k} \phi\,(x+z)\, \zeta(z)\, \mathrm{d}z = \int_{\mathbb{R}^k} h\,(x+z)\, \zeta(z)\, \mathrm{d}z$, for all $x \in \mathbb{R}^k$. This shows that FTPL can be interpreted as OSMD with a potential F defined implicitly by its Fenchel conjugate

$$F^*(x) := \int_{\mathbb{R}^k} \phi\,(x+z)\, \zeta(z)\, \mathrm{d}z \,, \qquad \forall x \in \mathbb{R}^k \,.$$

Thus, recalling (3.5), we can think of the update $\overline{x}_t(v)$ of OSMD as the average of a random component-wise draw $\overline{x}_t(i, v) = \sum_{a \in \mathcal{A}} P_t(a, v)\, a(i)$ (for all $i \in \{1, \dots, k\}$), with respect to a distribution $P_t(v)$ on $\mathcal{A}$ defined in terms of the distribution of Z_t, as

$$P_t(a, v) = \mathbb{P}_t\Big[h\big(Z_t(v) - \eta \widehat{L}_{t-1}(v)\big) = a \Big] \,, \qquad \forall a \in \mathcal{A} \,,$$

where $\mathbb{P}_t$ is the probability conditioned of the history up to time $t-1$.

For the understanding of the definitions and analyses of $K_t(w)$ and $\widehat{\ell}_t(v)$, we introduce three useful lemmas on geometric distributions. We defer their proofs to Appendix 3.8.3.

Lemma 4. *Let $Y_1, \dots, Y_m$ be m independent random variables such that each Y_j has a geometric distribution with parameter $p_j \in [0, 1]$. Then, the random variable $Z := \min_{j \in \{1, \dots, m\}} Y_j$ has a geometric distribution with parameter $1 - \prod_{j=1}^m (1 - p_j)$.*

Lemma 5. *Let G be a geometric random variable with parameter $q \in (0,1]$ and $\beta > 0$. Then, the expectation of the random variable $\min\{G,\beta\}$ satisfies $\mathbb{E}\big[\min\{G,\beta\}\big] = \big(1-(1-q)^{\beta}\big)/q$.*

Lemma 6. *For all $v \in \mathcal{V}$, fix two arbitrary numbers $p_1(v), p_2(v) \in [0,1]$. Consider a collection $\big\{X_s(v), Y_s(v)\big\}_{s\in\mathbb{N}, v\in\mathcal{V}}$ of independent Bernoulli random variables such that $\mathbb{E}\big[X_s(v)\big] = p_1(v)$ and $\mathbb{E}\big[Y_s(v)\big] = p_2(v)$ for any $s \in \mathbb{N}$ and all $v \in \mathcal{V}$. Then, the random variables $\{G(v)\}_{v\in\mathcal{V}}$ defined for all $v \in \mathcal{V}$ by $G(v) := \inf\big\{s \in \mathbb{N} : X_s(v)\, Y_s(v) = 1\big\}$ are all independent and they have a geometric distribution with parameter $p_1(v)\, p_2(v)$.*

Fix now any time step t, agent v, and component $i \in \{1,\dots,k\}$. The loss estimator $\widehat{\ell}_t(i,v)$ depends on the algorithmic definition of $K_t(i,w)$ in Algorithm 2, where $w \in \mathcal{N}(v)$. By Lemma 6, we have that for any w, conditionally on the history up to time $t-1$, the random variable $K_t(i,w)$, has a truncated geometric distribution with success probability equal to $\overline{x}_t(i,w)q(w)$ and truncation parameter β. The loss estimator of v is then defined as

$$\widehat{\ell}_t(i,v) := \ell_t(i)\, B_t\,(i,v) \min_{w\in\mathcal{N}(v)} \big\{K_t(i,w)\big\}, \tag{3.6}$$

where

$$B_t(i,v) = \mathbb{I}\big\{\exists w \in \mathcal{N}(v) : w \in \mathcal{S}_t,\, x_t(i,w) = 1\big\}\,, \qquad K_t(i,w) = \min\big\{G_t(i,w), \beta\big\}\,, \tag{3.7}$$

and given the history up to time $t-1$, for each $i \in \{1,\dots,k\}$, the family $\big\{G_t(i,w)\big\}_{w\in\mathcal{V}}$ consists of independent geometric random variables with parameter $\overline{x}_t(i,w)q(w)$. Note that the geometric random variables $G_t(i,w)$ are actually never computed by Algorithm 2 which efficiently computes only their truncations $K_t(i,w)$, with truncation parameter β. Nevertheless, as it will be apparent later, they are a useful tool for the theoretical analysis. Note that, by Eq. (3.5), we have

$$\mathbb{P}_t\big[x_t(i,w) = 1\big] = \mathbb{E}_t[x_t(i,w)] = \overline{x}_t\,(i,w)\,,$$

therefore

$$\overline{B}_t(i,v) := \mathbb{E}_t\,[B_t(i,v)] = 1 - \prod_{w\in\mathcal{N}(v)} \big(1-\overline{x}_t(i,w)\,q(w)\big) = \frac{1}{\mathbb{E}_t\Big[\min_{w\in\mathcal{N}(v)} G_t(i,w)\Big]}\,,$$

where the last identity follows by Lemma 4. Moreover from Lemma 5, we have

$$\mathbb{E}_t\,[K_t(i,w)] = \frac{1-\prod_{w\in\mathcal{N}(v)}\big(1-\overline{x}_t(i,w)\,q(w)\big)^{\beta}}{\overline{B}_t(i,v)}\,.$$

The following key lemma gives an upper bound on the expected estimated loss.

Lemma 7. *For any time t, component i, agents v, and truncation parameter β, the expectation of the loss estimator in (3.6), given the history up to time $t-1$, satisfies*

$$\mathbb{E}_t\Big[\widehat{\ell}_t(i,v)\Big] = \ell_t(i)\left(1-\left(\prod_{w\in\mathcal{N}(v)}\big(1-\overline{x}_t(i,w)\,q(w)\big)\right)^{\beta}\right) \le \ell_t(i)\,.$$

Proof. Using the fact that, conditioned on the history up to time $t-1$, the random variable $\min_{w\in\mathcal{N}(v)} G_t(i,w)$ has a geometric distribution with parameter $\overline{B}_t(i,v)$ (Lemmas 4-6), we get

$$\begin{aligned}
\mathbb{E}_t\left[\widehat{\ell}_t(i,v)\right] &= \mathbb{E}_t\left[\ell_t(i)B_t(i,v)\min_{w\in\mathcal{N}(v)}\left\{\min\left\{G_t(i,w),\beta\right\}\right\}\right] \\
&= \mathbb{E}_t\left[\ell_t(i)B_t(i,v)\min\left\{\min_{w\in\mathcal{N}(v)} G_t(i,w),\beta\right\}\right] \\
&= \ell_t(i)\mathbb{E}_t\left[B_t(i,v)\right]\mathbb{E}_t\left[\min\left\{\min_{w\in\mathcal{N}(v)} G_t(i,w),\beta\right\}\right] \\
&= \ell_t(i)\overline{B}_t(i,v)\frac{\left(1-\left(1-\overline{B}_t(i,v)\right)^{\beta}\right)}{\overline{B}_t(i,v)} \\
&= \ell_t(i)\left(1-\left(1-\overline{B}_t(i,v)\right)^{\beta}\right) \\
&= \ell_t(i)\left(1-\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right)\,,
\end{aligned}$$

where we plugged in the definition of $\overline{B}_t(i,v)$ in the last equation. From the fact that $\overline{x}_t(i,w)\,q(w)\in[0,1]$ and $\beta>0$ follows that $\mathbb{E}_t\left[\widehat{\ell}_t(i,v)\right]\leq\ell_t(i)$. □

We can finally state our upper bound on the regret of Coop-FTPL.

Theorem 3. *If ζ is the Laplace density $z\mapsto\zeta(z)=2^{-k}\exp\left(-\left\|z\right\|_1\right)$ and the parameters η,β are chosen as follows*

$$\beta=\left\lfloor\frac{1}{k\eta}\right\rfloor \quad \textit{and} \quad \eta=\sqrt{\frac{2m\log(k)}{5kT\left(\frac{k}{Q}\alpha_1+m\right)}}\,, \qquad \textit{where} \qquad Q=\sum_{v\in\mathcal{V}}q(v)\,, \tag{3.8}$$

then the regret of our Coop-FTPL algorithm (Algorithm 1) satisfies

$$R_T\leq 2Q\sqrt{10mkT\log(k)\left(\frac{k}{Q}\alpha_1+m\right)}\,.$$

We now present a detailed sketch of the proof of our main result (full proof in Appendix 3.8.4).

Proof sketch. For the sake of convenience, we define the expected individual regret of an agent $v\in\mathcal{V}$ in the network with respect to a fixed action $a\in\mathcal{A}$ by

$$R_T(a,v):=\mathbb{E}\left[\sum_{t=1}^{T}\left\langle x_t(v),\ell_t\right\rangle-\sum_{t=1}^{T}\left\langle a,\ell_t\right\rangle\right]\,,$$

where the expectation is taken with respect to the internal randomization of the agent, but not to its activation probability $q(v)$. With this definition the total regret on the

network in Eq. (3.4) can be decomposed as

$$R_T = \max_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{S}_t}\Big(\langle x_t(v),\ell_t\rangle - \langle a,\ell_t\rangle\Big)\right] = \max_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\sum_{v\in\mathcal{S}_t}\Big(\langle x_t(v),\ell_t\rangle - \langle a,\ell_t\rangle\Big)\right]\right]$$
$$= \max_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} q(v)\mathbb{E}_t\Big[\langle x_t(v),\ell_t\rangle - \langle a,\ell_t\rangle\Big]\right] = \max_{a\in\mathcal{A}} \sum_{v\in\mathcal{V}} q(v) R_T(a,v)\,. \tag{3.9}$$

The proof then proceeds by isolating the bias in the loss estimators. For each $a\in\mathcal{A}$ we have

$$R_T(a,v) = \mathbb{E}\left[\sum_{t=1}^{T}\Big\langle \overline{x}_t(v) - a, \widehat{\ell}_t(v)\Big\rangle\right] + \mathbb{E}\left[\sum_{t=1}^{T}\Big\langle \overline{x}_t(v) - a, \ell_t - \widehat{\ell}_t(v)\Big\rangle\right]\,.$$

Exploiting the analogy that we established between FTPL and OSMD, we begin by using the standard bound for the regret of OSMD in the first term of the previous equation. For the reader's convenience, we restate it in Appendix 6, Theorem 6. This leads to

$$R_T(a,v) \le \underbrace{\frac{F(\overline{x}_1(v)) - F(a)}{\eta}}_{\text{(I)}} + \underbrace{\mathbb{E}\left[\frac{1}{\eta}\sum_{t=1}^{T}\mathcal{B}_F\left(\overline{x}_t(v),\overline{x}_{t+1}(v)\right)\right]}_{\text{(II)}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\Big\langle \overline{x}_t(v) - a, \ell_t - \widehat{\ell}_t(v)\Big\rangle\right]}_{\text{(III)}}.$$

The three terms are studied separately and in detail in Appendix 3.8.4. Here, we provide a sketch of the bounds.

For the first term (I), we use the fact that the regularizer F satisfies, for all $a\in\mathcal{A}$,

$$F(a) \ge -m\big(1+\log(k)\big), \tag{3.10}$$

which follows by the definition of F, the properties of the perturbation distribution , and the fact that $\|a\|_1 \le m$ for any $a\in\mathcal{A}$. One can also show that $F(a)\le 0$ for all $a\in\mathcal{A}$, and this, combined with the previous equation, leads to

$$\text{(I)} \le \frac{m(1+\log k)}{\eta}\,.$$

For the second term (II), we have

$$\begin{aligned}\mathcal{B}_F\left(\overline{x}_t(v),\overline{x}_{t+1}(v)\right) &= \mathcal{B}_{F^*}\left(\nabla F\left(\overline{x}_{t+1}(v)\right), \nabla F\left(\overline{x}_t(v)\right)\right)\\ &= \mathcal{B}_{F^*}\Big(-\eta\widehat{L}_{t-1}(v) - \eta\widehat{\ell}_t(v), -\eta\widehat{L}_{t-1}(v)\Big)\\ &= \frac{\eta^2}{2}\left\|\widehat{\ell}_t(v)\right\|^2_{\nabla^2 F^*(\xi(v))}\,,\end{aligned} \tag{3.11}$$

where the first equality is a standard property of Bregmann divergence (see Theorem 5 in Appendix 3.8.1), the second follows from the definitions of the updates and the last by Taylor's theorem, where $\xi(v) = -\eta\widehat{L}_{t-1}(v) - \alpha\eta\widehat{\ell}_t(v)$, for some $\alpha\in[0,1]$. The estimation of the entries of the Hessian are non trivial (but tedious); the interested

reader can find them in Appendix 3.8.4. Exploiting our assumption that $\beta \leq 1/(\eta k)$, we get, for all $i, j \in \{1, \dots, k\}$,

$$\nabla^2 F^*\big(\xi(v)\big)_{ij} \leq e\, \overline{x}_t(i,v) \ .$$

Plugging this estimate in Eq. (3.11) yields

$$\begin{aligned}
\frac{\eta^2}{2} \left\| \widehat{\ell}_t(v) \right\|^2_{\nabla^2 F^*(\xi(v))} &= \frac{\eta^2}{2} \sum_{i=1}^k \sum_{j=1}^k \nabla^2 F^*\big(\xi(v)\big)_{i,j} \widehat{\ell}_t(i,v) \widehat{\ell}_t(j,v) \\
&\leq \frac{\eta^2 e}{2} \sum_{i=1}^k \sum_{j=1}^k \overline{x}_t(i,v) \widehat{\ell}_t(i,v) \widehat{\ell}_t(j,v) \\
&\leq \frac{\eta^2 e}{2} \sum_{i=1}^k \sum_{j=1}^k \overline{x}_t(i,v) B_t(i,v) \min_{w \in \mathcal{N}(v)} \{G_t(i,w)\}\, B_t(j,v) \min_{w \in \mathcal{N}(v)} \{G_t(j,w)\} \,,
\end{aligned}$$

where the last inequality follows by neglecting the truncation with β. Hence multiplying (II) by $q(v)$ and summing over $v \in \mathcal{V}$ yields

$$\begin{aligned}
&\sum_{v \in \mathcal{V}} q(v) \mathbb{E}\left[\frac{\eta}{2} \sum_{t=1}^T \left\| \widehat{\ell}_t(v) \right\|^2_{\nabla^2 F^*(\xi(v))} \right] = \sum_{v \in \mathcal{V}} q(v) \frac{\eta}{2} \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_t \left[\left\| \widehat{\ell}_t(v) \right\|^2_{\nabla^2 F^*(\xi(v))} \right] \right] \\
&\leq \sum_{v \in \mathcal{V}} q(v) \frac{\eta e}{2} \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_t \left[\sum_{i,j=1}^k \overline{x}_t(i,v) B_t(i,v) \min_{w \in \mathcal{N}(v)} \{G_t(i,w)\}\, B_t(j,v) \min_{w \in \mathcal{N}(v)} \{G_t(j,w)\} \right] \right] ,
\end{aligned}$$

which, making use of Lemmas 4–6, gives

$$\begin{aligned}
&\sum_{v \in \mathcal{V}} q(v) \frac{\eta e}{2} \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_t \left[\sum_{i,j=1}^k \overline{x}_t(i,v) B_t(i,v) \min_{w \in \mathcal{N}(v)} \{G_t(i,w)\}\, B_t(j,v) \min_{w \in \mathcal{N}(v)} \{G_t(j,w)\} \right] \right] \\
&= \sum_{v \in \mathcal{V}} q(v) \frac{\eta e}{2} \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_t \left[\sum_{i=1}^k \sum_{j=1}^k \overline{x}_t(i,v) B_t(i,v) \tilde{G}_t(i,v) B_t(j,v) \tilde{G}_t(j,v) \right] \right] \\
&= \sum_{v \in \mathcal{V}} q(v) \frac{\eta e}{2} \mathbb{E}\left[\sum_{t=1}^T \sum_{i=1}^k \sum_{j=1}^k \overline{x}_t(i,v) \mathbb{E}_t\left[B_t(i,v) B_t(j,v) \right] \mathbb{E}_t\left[\tilde{G}_t(i,v) \right] \mathbb{E}_t\left[\tilde{G}_t(j,v) \right] \right] =: (\star) \ ,
\end{aligned}$$

where in the first equality we defined $\tilde{G}_t(i,v) = \min_{w \in \mathcal{N}(v)} \big\{ G_t(i,w) \big\}$ and, analogously, $\tilde{G}_t(j,v) = \min_{w \in \mathcal{N}(v)} \big\{ G_t(j,w) \big\}$, while the second follows by the conditional independence of the three terms $\big(B_t(i,v), B_t(j,v) \big)$, $\tilde{G}_t(i,v)$, and $\tilde{G}_t(j,v)$ given the history up to time $t-1$. Further upper bounding, we get

$$\begin{aligned}
(\star) &= \sum_{v \in \mathcal{V}} q(v) \frac{\eta e}{2} \mathbb{E}\left[\sum_{t=1}^T \mathbb{E}_t \left[\sum_{i=1}^k \sum_{j=1}^k \frac{\overline{x}_t(i,v)}{\overline{B}_t(i,v)} B_t(i,v) \frac{B_t(j,v)}{\overline{B}_t(j,v)} \right] \right] \\
&\leq \frac{\eta e k}{2} \mathbb{E}\left[\sum_{t=1}^T \sum_{i=1}^k \sum_{v \in \mathcal{V}} \frac{\overline{x}_t(i,v) q(v)}{\overline{B}_t(i,v)} \right] \leq \frac{\eta e k T}{2\,(1 - e^{-1})} \left(k \alpha_1 + m Q \right) ,
\end{aligned}$$

where the first equality uses the expected value of the geometric random variables $\tilde{G}$, the first inequality is obtained neglecting the indicator function $B_t(i,v)$ and taking the

conditional expectation of $B_t(j,v)$, and the last inequality follows by a known upper bound involving independence numbers appearing, for example in Cesa-Bianchi et al. [2019a,b]. For the sake of convenience, we add this result to Appendix 3.8.5, Lemma 9. We now consider the last term (III). Since $\ell_t \geq \mathbb{E}_t\left[\widehat{\ell}_t(v)\right]$ by Lemma 7, we have

$$\begin{aligned}(\text{III}) &= \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\left\langle \overline{x}_t(v) - a, \ell_t - \widehat{\ell}_t(v)\right\rangle\right]\right] \leq \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\left\langle \overline{x}_t(v), \ell_t - \widehat{\ell}_t(v)\right\rangle\right]\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\ell_t(i)\overline{x}_t(i,v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right].\end{aligned}$$

Multiplying (III) by $q(v)$ and summing over the agents, we now upper bound $\ell_t(i)$ with 1 and use the facts that $1-x \leq e^{-x}$ for all $x \in [0,1]$ and $e^{-y} \leq 1/y$ for all $y > 0$, to obtain

$$\begin{aligned}&\sum_{v\in\mathcal{V}} q(v)\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\ell_t(i)\overline{x}_t(i,v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right] \\ &\quad\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{v\in\mathcal{V}}\overline{x}_t(i,v)\,q(v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right] \\ &\quad= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{\substack{v\in\mathcal{V}\\ \overline{x}_t(i,v)\,q(v)>0}}\overline{x}_t(i,v)\,q(v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right] \\ &\quad\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{\substack{v\in\mathcal{V}\\ \overline{x}_t(i,v)\,q(v)>0}}\overline{x}_t(i,v)q(v)\exp\left(-\beta\sum_{w\in\mathcal{N}(v)}\overline{x}_t(i,w)\,q(w)\right)\right] \\ &\quad\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{\substack{v\in\mathcal{V}\\ \overline{x}_t(i,v)\,q(v)>0}}\frac{\overline{x}_t(i,v)\,q(v)}{\beta\sum_{w\in\mathcal{N}(v)}\overline{x}_t(i,w)\,q(w)}\right] \leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\alpha_1}{\beta}\right] = \frac{\alpha_1\,k\,T}{\beta}\end{aligned}$$

where the last inequality follows by a known upper bound involving independence numbers appearing, for example in [Alon et al., 2017, Lemma 10]. For the sake of convenience, we add this result to Appendix 3.8.5, Lemma 8.

Putting everything together and recalling that $\beta = \lfloor 1/(k\eta)\rfloor \geq 1/(2k\eta)$, we can finally

conclude that

$$\begin{aligned}
R_T &\leq Q\frac{m\,(1+\log(k))}{\eta} + Q\frac{\eta e k T}{2\,(1-e^{-1})}\left(\frac{k}{Q}\alpha_1 + m\right) + \frac{\alpha_1\, k\, T}{\beta} \\
&\leq Q\frac{m\,(1+\log(k))}{\eta} + Q\frac{\eta e k T}{2\,(1-e^{-1})}\left(\frac{k}{Q}\alpha_1 + m\right) + 2\eta\alpha_1 k^2 T \\
&= Q\frac{m\,(1+\log(k))}{\eta} + \eta Q k T\left(\frac{e}{2\,(1-e^{-1})}\left(\frac{k}{Q}\alpha_1 + m\right) + 2\alpha_1\frac{k}{Q}\right) \\
&\leq Q\frac{m\,(1+\log(k))}{\eta} + 5\eta Q k T\left(\frac{k}{Q}\alpha_1 + m\right) \\
&\leq 2Q\sqrt{10mkT\log(k)\left(\frac{k}{Q}\alpha_1 + m\right)}\,.
\end{aligned}$$

□

We conclude this section by discussing the computational complexity of our Coop-FTPL algorithm. The next result shows that the total number of elementary operations performed by Coop-FTPL over T time-steps scales with $T^{3/2}$ in the worst-case. To the best of our knowledge, no known algorithm attains a lower worst-case computational complexity.

Proposition 2. *If the optimization problem (3.2) can be solved with at most $c \in \mathbb{N}$ elementary operations, the worst-case computational complexity $\gamma_{\text{Coop-FTPL}}$ of each agent $v \in \mathcal{V}$ running our Coop-FTPL algorithm with the optimal tuning (3.8) for T rounds is*

$$\gamma_{\text{Coop-FTPL}} = \mathcal{O}\left(T^{3/2} c\sqrt{\frac{\alpha_1/Q+1}{m}}\right)\,.$$

Proof. The result follows immediately by noting that the number of elementary operations performed by each agent v at each time step t is at most

$$c(\beta+1) \leq c\frac{1}{k\eta} = c\frac{1}{k}\sqrt{\frac{5kT(k\alpha_1/Q+m)}{2m\log k}} = c\sqrt{\frac{5T(\alpha_1/Q+m/k)}{2m\log k}}\,.$$

□

3.6 Lower bound

In this section we show that no cooperative semi-bandit algorithm can beat the $Q\sqrt{mkT\alpha_1/Q}$ rate. The key idea for constructing the lower bound is simple: if the activation probabilities $q(v)$ are non-zero only for agents v belonging to an independent set with cardinality α_1, then the problem is reduced to α_1 independent instances of single-agent semi-bandits, whose minimax rate is known.

Theorem 4. *For each communication network $\mathcal{G}$ with independence number α_1 there exist cooperative semi-bandit instances for which the regret of any learning algorithm satisfies*

$$R_T = \Omega\big(Q\sqrt{mkT\alpha_1/Q}\big)\,.$$

Proof. Let $\mathcal{W} = \{w_1, \dots, w_{\alpha_1}\} \subset \mathcal{V}$ be an independent set with cardinality α_1. Furthermore, let $q \in (0, 1]$ be a positive probability and for all agents $v \in \mathcal{V}$, let

$$q(v) = q\mathbb{I}\{v \in \mathcal{W}\} \ .$$

In words, only agents belonging to an independent set with largest cardinality are activated (with positive probability), and all with the same probability. Thus, only agents in $\mathcal{W}$ contribute to the expected regret and their total mass $Q = \sum_{v \in \mathcal{V}} q(v)$ is equal to $\alpha_1 q$. Moreover, note that being non-adjacent, agents in $\mathcal{W}$ never exchange any information. Each agent $w \in \mathcal{W}$ is therefore running an independent single-agent online linear optimization problem with semi-bandit feedback for an average of qT rounds. Since for single-agent semi-bandits, the worst-case lower bound on the regret after T' time steps is known to be $\Omega\big(\sqrt{mkT'}\big)$ (see, e.g., Audibert et al. [2014], Lattimore et al. [2018]) and the cardinality of $\mathcal{W}$ is α_1, the regret of any cooperative semi-bandit algorithm run on this instance satisfies

$$R_T = \Omega\big(\alpha_1\sqrt{mk\,qT}\big) = \Omega\big(\alpha_1 q\sqrt{mkT/q}\big) = \Omega\big(Q\sqrt{mkT\alpha_1/Q}\big) \ ,$$

where we used $Q = \alpha_1 q$. This concludes the proof. □

In the previous section we showed that the expected regret of our Coop-FTPL algorithm can always be upper bounded by $Q\sqrt{mkT\log(k)(k\alpha_1/Q + m)}$ (ignoring constants). Thus, Theorem 4 shows that, up to the additive m term inside the rightmost bracket, the regret of Coop-FTPL is at most $\sqrt{k\log k}$-away from the minimax optimal rate.

3.7 Conclusions and open problems

Motivated by spatially distributed large-scale learning systems, we introduced a new cooperative setting for adversarial semi-bandits in which only some of the agents are active at any given time step. We designed and analyzed an efficient algorithm that we called Coop-FTPL for which we proved near-optimal regret guarantees with state-of-the-art computational complexity costs. Our analysis relies on the fact that agents are aware of their activation probabilities, and they have some prior knowledge about the connectivity of the graph. Two interesting new lines of research are investigating if either of these assumptions could be lifted while retaining low regret and good computational complexity. In particular, removing the need for prior knowledge of the independence number would represent a significant theoretical and practical improvement, given that computing α_1 is NP-hard in the worst-case. Unfortunately, existing techniques that address this problem in similar settings (e.g., Cesa-Bianchi et al. [2019b]) rely heavily on agents being active at all time steps, and they are unlikely to yield any results in our general case. We believe that entirely new ideas will be required to deal with this issue. We leave these intriguing problems open for future work.

3.8 Appendix

3.8.1 Legendre functions and Fenchel conjugates

In this section, we briefly recall a few known definitions and facts in convex analysis.

Definition 1 (Interior, boundary, and convex hull). *For any subset E of $\mathbb{R}^k$, we denote its* topological interior *by* $\mathrm{int}(E)$*, its* boundary *by* ∂E*, and its* convex hull *by* $\mathrm{co}(E)$.

Definition 2 (Effective domain). *The* effective domain *of a convex function $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ is*

$$\mathrm{dom}(F) := \left\{x \in \mathbb{R}^k : F(x) < +\infty\right\}. \tag{3.12}$$

With a slight abuse of notation, we will denote with the same symbol f a convex function $f\colon \to \mathbb{R} \cup \{+\infty\}$ and its restriction $\widetilde{f}\colon \mathrm{dom}(f) \to \mathbb{R}$ to its effective domain.

Definition 3 (Legendre function). *A convex function $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ is* Legendre *if*

1. $\mathrm{int}\big(\mathrm{dom}(F)\big)$ *is non-empty;*
2. *F is differentiable and strictly convex on* $\mathrm{int}\big(\mathrm{dom}(F)\big)$*;*
3. *for all $x_0 \in \partial\left[\mathrm{int}\big(\mathrm{dom}(F)\big)\right]$, if $x \in \mathrm{int}\big(\mathrm{dom}(F)\big), x \to x_0$, then $\|\nabla F(x)\|_2 \to +\infty$.*

Definition 4 (Fenchel conjugate). *Let $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ be a convex function. The* Fenchel conjugate *F^* of F is defined as the function*

$$\begin{aligned} F^*\colon \mathbb{R}^k &\to \mathbb{R} \cup \{+\infty\} \\ z &\mapsto F^*(z) := \sup_{x \in \mathbb{R}^k} \big(\langle x, z\rangle - F(x)\big). \end{aligned}$$

Definition 5 (Bregman divergence). *Let $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ a convex function with non-empty* $\mathrm{int}\big(\mathrm{dom}(F)\big)$ *that is differentiable on* $\mathrm{int}\big(\mathrm{dom}(F)\big)$*. The* Bregman divergence *induced by F is*

$$\begin{aligned} \mathcal{B}_F\colon \mathbb{R}^k \times \mathrm{int}\big(\mathrm{dom}(F)\big) &\to \mathbb{R} \cup \{+\infty\} \\ (x, y) &\mapsto \mathcal{B}_F(x, y) := F(x) - F(y) - \big\langle \nabla F(y), x - y\big\rangle. \end{aligned}$$

The following results are taken from [Lattimore and Szepesvári, 2020, Theorem 26.6 and Corollary 26.8].

Theorem 5. *Let $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ be a Legendre function. Then:*

1. *the Fenchel conjugate F^* of F is Legendre;*
2. *$\nabla F\colon \mathrm{int}\big(\mathrm{dom}(F)\big) \to \mathrm{int}\big(\mathrm{dom}(F^*)\big)$ is bijective with inverse $(\nabla F)^{-1} = \nabla F^*$;*
3. *$\mathcal{B}_F(x, y) = \mathcal{B}_{F^*}\big(\nabla F(y), \nabla F(x)\big)$, for all $x, y \in \mathrm{int}\big(\mathrm{dom}(f)\big)$.*

Corollary 1. *If $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ is a Legendre function and $x \in \mathrm{argmin}_{x \in \mathrm{dom}(F)} F(x)$, then $x \in \mathrm{int}\big(\mathrm{dom}(F)\big)$.*

3.8.2 Online Stochastic Mirror Descent (OSMD)

In this section, we briefly recall the standard Online Stochastic Mirror Descent algorithm (OSMD) (Algorithm 3) and its analysis.

For an overview on some basic convex analysis definitions and results, we refer the reviewer to the previous Appendix 3.8.1. For a convex function $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$ that is differentiable on the non-empty interior $\mathrm{int}\big(\mathrm{dom}(F)\big) \neq \varnothing$ of its effective domain

$\mathrm{dom}(F)$, we denote by $\mathcal{B}_F$ the Bregman divergence induced by F (Definition 5). Following the existing convention, we refer to the input function F of OSMD as a *potential*. Furthermore, given a measure P on a subset of $\mathbb{R}^k$, we say that a vector $x \in \mathbb{R}^k$ is the *mean of the measure* P if x is the component-wise expectation of a $\mathbb{R}^k$-valued random variable with distribution P. For any time step $t \in \{1, 2, \dots\}$, we denote by $\mathbb{E}_t$ the expectation conditioned to the history up to and including round $t-1$.

Algorithm 3: Online Stochastic Mirror Descent (OSMD)

Input: Legendre potential $F\colon \mathbb{R}^k \to \mathbb{R} \cup \{+\infty\}$, compact action set $\mathcal{A} \subset \mathbb{R}^k$ with $\mathrm{int}\big(\mathrm{dom}(F)\big) \cap \mathrm{co}(\mathcal{A}) \neq \varnothing$, learning rate $\eta > 0$

Initialization: $\overline{x}_1 = \mathrm{argmin}_{x \in \mathrm{dom}(F) \cap \mathrm{co}(\mathcal{A})} F(x)$

1 **for** $t = 1, 2, \dots$ **do**

2 choose a measure P_t on $\mathcal{A}$ with mean $\overline{x}_t$

3 make a prediction x_t drawn from $\mathcal{A}$ according to P_t and suffer the loss $\langle x_t, \ell_t \rangle$

4 compute an estimate $\widehat{\ell}_t$ of the loss vector ℓ_t

5 make the update: $\overline{x}_{t+1} = \mathrm{argmin}_{x \in \mathrm{dom}(F) \cap \mathrm{co}(\mathcal{A})} \eta \langle x, \widehat{\ell}_t \rangle + \mathcal{B}_F(x, \overline{x}_t)$

It is known that since $\mathrm{co}(\mathcal{A})$ is convex and compact, $\mathrm{int}\big(\mathrm{dom}(F)\big) \cap \mathrm{co}(\mathcal{A}) \neq \varnothing$, and F is Legendre, then, all the argmin's exist in Algorithm 3 and $\overline{x}_t \in \mathrm{int}\big(\mathrm{dom}(F)\big) \cap \mathrm{co}(\mathcal{A})$ for all $t \in \{1, 2, \dots\}$ (see, e.g., [Lattimore and Szepesvári, 2020, Exercise 28.2]).

The following result is taken from [Lattimore and Szepesvári, 2020, Theorem 28.10] and gives an upper bound on the regret of OSMD.

Theorem 6. *Suppose that OSMD (Algorithm 3) is run with input $F, \mathcal{A}, \eta$. If the estimates $\widehat{\ell}_t$ computed at line 4 satisfy $\mathbb{E}_t\big[\widehat{\ell}_t\big] = \ell_t$ for all $t \in \{1, 2, \dots\}$, then, for all $x \in \mathrm{co}(\mathcal{A})$,*

$$\mathbb{E}\left[\sum_{t=1}^{T} \langle \overline{x}_t, \ell_t \rangle - \sum_{t=1}^{T} \langle x, \ell_t \rangle\right] \leq \mathbb{E}\left[\frac{F(x) - F(\overline{x}_1)}{\eta} + \sum_{t=1}^{T} \langle \overline{x}_t - \overline{x}_{t+1}, \widehat{\ell}_t \rangle - \frac{1}{\eta} \sum_{t=1}^{T} \mathcal{B}_F(\overline{x}_{t+1}, \overline{x}_t)\right] .$$

Furthermore, letting

$$\widetilde{x}_{t+1} = \operatorname*{argmin}_{x \in \mathrm{dom}(F)} \eta \langle x, \widehat{\ell}_t \rangle + \mathcal{B}_F(x, \overline{x}_t)$$

and assuming that $-\eta \widehat{\ell}_t + \nabla F(x) \in \nabla F\big(\mathrm{dom}(F)\big)$ for all $x \in \mathrm{co}(\mathcal{A})$ almost surely, then

$$\sup_{x \in \mathrm{co}(\mathcal{A})} \mathbb{E}\left[\sum_{t=1}^{T} \langle \overline{x}_t, \ell_t \rangle - \sum_{t=1}^{T} \langle x, \ell_t \rangle\right] \leq \frac{\mathrm{diam}_F\big(\mathrm{co}(\mathcal{A})\big)}{\eta} + \frac{1}{\eta} \sum_{t=1}^{T} \mathbb{E}\big[\mathcal{B}_F(\overline{x}_t, \widetilde{x}_{t+1})\big] ,$$

where $\mathrm{diam}_F\big(\mathrm{co}(\mathcal{A})\big) := \sup_{x,y \in \mathrm{co}(\mathcal{A})} \big(F(x) - F(y)\big)$ *is the* diameter *of* $\mathrm{co}(\mathcal{A})$ *with respect to* F.

3.8.3 Proofs of lemmas on geometric distributions

In this section we provide all missing proofs on geometric distributions that we stated in Section 3.5.

Proof of Lemma 4. For all $j \in \{1, \dots, m\}$, the cumulative distribution function (c.d.f.) of Y_j is given, for all $n \in \mathbb{N}$, by

$$\mathbb{P}\left[Y_j \leq n\right] = p_j \sum_{i=1}^{n} \left(1 - p_j\right)^{i-1} = 1 - \left(1 - p_j\right)^n \ .$$

The c.d.f. of Z is given, for all $n \in \mathbb{N}$, by

$$\begin{aligned}
\mathbb{P}\left[Z \leq n\right] = \mathbb{P}\left[\min_{j \in \{1,\dots,m\}} Y_j \leq n\right] &= 1 - \prod_{j=1}^{m} \mathbb{P}\left[Y_j > n\right] = 1 - \prod_{j=1}^{m} \left(1 - \mathbb{P}\left[Y_j \leq n\right]\right) \\
&= 1 - \prod_{j=1}^{m} \left(1 - \left(1 - \left(1 - p_j\right)^n\right)\right) = 1 - \left(\prod_{j=1}^{m} \left(1 - p_j\right)\right)^n \\
&= 1 - \left(1 - \left[1 - \prod_{j=1}^{m} \left(1 - p_j\right)\right]\right)^n \ ,
\end{aligned}$$

and this is the c.d.f. of a geometric random variable with parameter $\left(1 - \prod_{j=1}^{m} \left(1 - p_j\right)\right.$ $\left.\right)$. □

Proof of Lemma 5. By elementary calculations,

$$\begin{aligned}
\mathbb{E}\left[\min\left\{G, \beta\right\}\right] &= \sum_{n=1}^{\infty} n \left(1 - q\right)^{n-1} q - \sum_{n=\beta}^{\infty} \left(n - \beta\right) \left(1 - q\right)^{n-1} q \\
&= \sum_{n=1}^{\infty} n \left(1 - q\right)^{n-1} q - \left(1 - q\right)^{\beta} \sum_{n=\beta}^{\infty} \left(n - \beta\right) \left(1 - q\right)^{n-\beta-1} q \\
&= \left(1 - \left(1 - q\right)^{\beta}\right) \sum_{n=1}^{\infty} n \left(1 - q\right)^{n-1} q = \frac{\left(1 - \left(1 - q\right)^{\beta}\right)}{q} \ .
\end{aligned}$$

□

Proof of Lemma 6. The proof follows immediately from the fact that $\left\{X_s(v)\, Y_s(v)\right\}_{s \in \mathcal{I}, v \in \mathcal{V}}$ is a collection of independent Bernoulli random variables with expectation $\mathbb{E}\left[X_s(v)\, Y_s(v)\right] = p_1(v)\, p_2(v)$ for any $s \in \mathbb{N}$ and all $v \in \mathcal{V}$. □

3.8.4 Proof of Theorem 3

In this section, we present a complete proof of Theorem 3.

Proof of Theorem 3. For the sake of convenience, we define the expected individual regret of an agent $v \in \mathcal{V}$ in the network with respect to a fixed action $a \in \mathcal{A}$ by

$$R_T(a, v) := \mathbb{E}\left[\sum_{t=1}^{T} \left\langle x_t(v), \ell_t \right\rangle - \sum_{t=1}^{T} \left\langle a, \ell_t \right\rangle\right] \ ,$$

where the expectation is taken with respect to the internal randomization of the agent, but not its activation probability $q(v)$. With this definition the total regret on the network in Eq. (3.4) can be decomposed as

$$R_T = \max_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{S}_t}\Big(\langle x_t(v),\ell_t\rangle - \langle a,\ell_t\rangle\Big)\right] = \max_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\sum_{v\in\mathcal{S}_t}\Big(\langle x_t(v),\ell_t\rangle - \langle a,\ell_t\rangle\Big)\right]\right]$$
$$= \max_{a\in\mathcal{A}} \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} q(v)\mathbb{E}_t\Big[\langle x_t(v),\ell_t\rangle - \langle a,\ell_t\rangle\Big]\right] = \max_{a\in\mathcal{A}} \sum_{v\in\mathcal{V}} q(v) R_T(a,v)\,. \tag{3.13}$$

The proof then proceeds by isolating the bias in the loss estimators. For each $a\in\mathcal{A}$ we get

$$\begin{aligned}
&R_T(a,v)\\
&= \mathbb{E}\left[\sum_{t=1}^{T}\langle x_t(v)-a,\ell_t\rangle\right] = \mathbb{E}\left[\mathbb{E}_t\left[\sum_{t=1}^{T}\langle x_t(v)-a,\ell_t\rangle\right]\right] = \mathbb{E}\left[\sum_{t=1}^{T}\langle \overline{x}_t(v)-a,\ell_t\rangle\right]\\
&= \mathbb{E}\left[\sum_{t=1}^{T}\Big\langle \overline{x}_t(v)-a,\widehat{\ell}_t(v)\Big\rangle\right] + \mathbb{E}\left[\sum_{t=1}^{T}\Big\langle \overline{x}_t(v)-a,\ell_t-\widehat{\ell}_t(v)\Big\rangle\right]\\
&\le \underbrace{\frac{F(\overline{x}_1(v))-F(a)}{\eta}}_{\text{(I)}} + \underbrace{\mathbb{E}\left[\frac{1}{\eta}\sum_{t=1}^{T}\mathcal{B}_F\left(\overline{x}_t(v),\overline{x}_{t+1}(v)\right)\right]}_{\text{(II)}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\Big\langle \overline{x}_t(v)-a,\ell_t-\widehat{\ell}_t(v)\Big\rangle\right]}_{\text{(III)}}
\end{aligned}$$

where the inequality follows by the standard analysis of OMD. We bound the three terms separately. For the first term (I), we have

$$\begin{aligned}
F(a) &= \sup_{x\in\mathbb{R}^k}\left(\langle a,x\rangle - F^*(x)\right) = \sup_{x\in\mathbb{R}^k}\left(\langle a,x\rangle - \mathbb{E}\left[\max_{a\in\mathcal{A}}\langle a,x+Z\rangle\right]\right)\\
&\ge -\mathbb{E}\left[\max_{a\in\mathcal{A}}\langle a,Z\rangle\right] \ge -m\mathbb{E}\left[\|Z\|_\infty\right] = -m\sum_{i=1}^{k}\frac{1}{i} \ge -m\left(1+\log(k)\right),
\end{aligned} \tag{3.14}$$

where the first inequality follows by choosing $x=0$, the second follows from Hölder's inequality and $\|a\|_1 \le m$ for any $a\in\mathcal{A}$, and the last equality is Exercise 30.4 in Lattimore and Szepesvári [2020]. It follows that

$$F\big(\overline{x}_1(v)\big) - F(a) \le m\big(1+\log(k)\big)\,,$$

where we use the fact that $F(a)\le 0$ for all $a\in\mathcal{A}$ and this follows from the first line of Eq. (3.14) by the convexity of the maximum, using Jensen's inequality and the fact that the random variable Z is centered. Thus

$$\text{(I)} \le \frac{m(1+\log k)}{\eta}\,.$$

We now study the second term (II). We have

$$\begin{aligned}\mathcal{B}_F\left(\overline{x}_t(v), \overline{x}_{t+1}(v)\right) &= \mathcal{B}_{F^*}\left(\nabla F\left(\overline{x}_{t+1}(v)\right), \nabla F\left(\overline{x}_t(v)\right)\right) \\ &= \mathcal{B}_{F^*}\left(-\eta \widehat{L}_{t-1}(v) - \eta \widehat{\ell}_t(v), -\eta \widehat{L}_{t-1}(v)\right) \\ &= \frac{\eta^2}{2}\left\|\widehat{\ell}_t(v)\right\|^2_{\nabla^2 F^*(\xi(v))}\,, \end{aligned} \tag{3.15}$$

where the first equality is a standard property of Bregmann divergence, the second follows from the definitions of the updates and the last by Taylor's theorem, where $\xi(v) = -\eta \widehat{L}_{t-1}(v) - \alpha\eta\widehat{\ell}_t(v)$, for some $\alpha \in [0,1]$. To calculate the Hessian we use a change of variable to avoid applying the gradient to the (possibly) non-differentiable argmax and we get:

$$\begin{aligned}\nabla^2 F^*(x) &= \nabla\left(\nabla F^*(x)\right) = \nabla \mathbb{E}[h(x+Z)] = \nabla \int_{\mathbb{R}^k} h(x+z)\zeta(z)dz \\ &= \nabla \int_{\mathbb{R}^k} h(u)\zeta(u-x)du = \int_{\mathbb{R}^k} h(u)(\nabla\zeta(u-x))^\top du \\ &= \int_{\mathbb{R}^k} h(u)\mathrm{sign}(u-x)^\top \zeta(u-x)du = \int_{\mathbb{R}^k} h(x+z)\mathrm{sign}(z)^\top\zeta(z)dz\end{aligned}$$

Using the definition of $\xi(v)$ and the fact that $h(x)$ is nonnegative,

$$\begin{aligned}\nabla^2 F^*(\xi(v))_{ij} &= \int_{\mathbb{R}^k} h(\xi(v)+z)_i \mathrm{sign}(z)_j \zeta(z)dz \\ &\le \int_{\mathbb{R}^k} h(\xi(v)+z)_i \zeta(z)dz \\ &= \int_{\mathbb{R}^k} h\left(z - \eta\widehat{L}_{t-1} - \alpha\eta\widehat{\ell}_t\right)_i \zeta(z)dz \\ &= \int_{\mathbb{R}^k} h\left(u - \eta\widehat{L}_{t-1}(v)\right)_i \zeta\left(u + \alpha\eta\widehat{\ell}_t(v)\right)du \\ &\le \exp\left(\left\|\alpha\eta\widehat{\ell}_t(v)\right\|_1\right)\int_{\mathbb{R}^k} h\left(u - \eta\widehat{L}_{t-1}(v)\right)_i \zeta(u)du \\ &\le \exp\left(\alpha\eta\sum_{i=1}^k B_t(i,v)\beta\right)\overline{x}_t(i,v) \\ &\le \exp\left(\alpha\eta k\beta\right)\overline{x}_t(i,v) \\ &\le e\,\overline{x}_t(i,v)\end{aligned}$$

where the last inequality follows by $\alpha \le 1$ and $\beta \le 1/(\eta k)$. Plugging this estimate in Eq. (3.15) yields

$$\begin{aligned}\frac{\eta^2}{2}\left\|\widehat{\ell}_t(v)\right\|^2_{\nabla^2 F^*(\xi(v))} &= \frac{\eta^2}{2}\sum_{i=1}^k\sum_{j=1}^k \nabla^2 F^*\left(\xi(v)\right)_{i,j}\widehat{\ell}_t(i,v)\widehat{\ell}_t(j,v) \\ &\le \frac{\eta^2 e}{2}\sum_{i=1}^k\sum_{j=1}^k \overline{x}_t(i,v)\widehat{\ell}_t(i,v)\widehat{\ell}_t(j,v) \\ &\le \frac{\eta^2 e}{2}\sum_{i=1}^k\sum_{j=1}^k \overline{x}_t(i,v)B_t(i,v)\min_{w\in\mathcal{N}(v)}\left\{G_t(i,w)\right\}B_t(j,v)\min_{w\in\mathcal{N}(v)}\left\{G_t(j,w)\right\},\end{aligned}$$

where the last inequality follows by neglecting the truncation with β. Hence multiplying (II) by $q(v)$ and summing over $v \in \mathcal{V}$ yields

$$\sum_{v\in\mathcal{V}} q(v)\mathbb{E}\left[\frac{\eta}{2}\sum_{t=1}^{T}\left\|\widehat{\ell}_t(v)\right\|^2_{\nabla^2 F^*(\xi(v))}\right] = \sum_{v\in\mathcal{V}} q(v)\frac{\eta}{2}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\left\|\widehat{\ell}_t(v)\right\|^2_{\nabla^2 F^*(\xi(v))}\right]\right]$$
$$\leq \sum_{v\in\mathcal{V}} q(v)\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\sum_{i,j=1}^{k}\overline{x}_t(i,v)B_t(i,v)\min_{w\in\mathcal{N}(v)}\{G_t(i,w)\}\,B_t(j,v)\min_{w\in\mathcal{N}(v)}\{G_t(j,w)\}\right]\right],$$

making use of Lemmas 4–6, gives

$$\sum_{v\in\mathcal{V}} q(v)\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\sum_{i,j=1}^{k}\overline{x}_t(i,v)B_t(i,v)\min_{w\in\mathcal{N}(v)}\{G_t(i,w)\}\,B_t(j,v)\min_{w\in\mathcal{N}(v)}\{G_t(j,w)\}\right]\right]$$
$$= \sum_{v\in\mathcal{V}} q(v)\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\overline{x}_t(i,v)B_t(i,v)\tilde{G}_t(i,v)B_t(j,v)\tilde{G}_t(j,v)\right]\right]$$
$$= \sum_{v\in\mathcal{V}} q(v)\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{j=1}^{k}\overline{x}_t(i,v)\mathbb{E}_t\left[B_t(i,v)B_t(j,v)\right]\mathbb{E}_t\left[\tilde{G}_t(i,v)\right]\mathbb{E}_t\left[\tilde{G}_t(j,v)\right]\right] =: (\star),$$

where in the first equality we defined $\tilde{G}_t(i,v) = \min_{w\in\mathcal{N}(v)}\{G_t(i,w)\}$ and, analogously, $\tilde{G}_t(j,v) = \min_{w\in\mathcal{N}(v)}\{G_t(j,w)\}$, while the second follows by the conditional independence of the three terms $\big(B_t(i,v), B_t(j,v)\big)$, $\tilde{G}_t(i,v)$, and $\tilde{G}_t(j,v)$ given the history up to time $t-1$. Further upper bounding, we get

$$\begin{aligned}
(\star) &= \sum_{v\in\mathcal{V}} q(v)\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\overline{x}_t(i,v)}{\overline{B}_t(i,v)}B_t(i,v)\frac{B_t(j,v)}{\overline{B}_t(j,v)}\right]\right] \\
&\leq \sum_{v\in\mathcal{V}} q(v)\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=1}^{T}\mathbb{E}_t\left[\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\overline{x}_t(i,v)}{\overline{B}_t(i,v)}\frac{B_t(j,v)}{\overline{B}_t(j,v)}\right]\right] \\
&= \sum_{v\in\mathcal{V}} q(v)\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{j=1}^{k}\frac{\overline{x}_t(i,v)}{\overline{B}_t(i,v)}\frac{\cancel{\overline{B}_t(j,v)}}{\cancel{\overline{B}_t(j,v)}}\right] \\
&= \frac{\eta e k}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{v\in\mathcal{V}}\frac{\overline{x}_t(i,v)q(v)}{\overline{B}_t(i,v)}\right] \\
&\leq \frac{\eta e k}{2}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\left(\frac{1}{1-e^{-1}}\left(\alpha_1 + \sum_{v\in\mathcal{V}}\overline{x}_t(i,v)q(v)\right)\right)\right] \\
&= \frac{\eta e k}{2}\mathbb{E}\left[\sum_{t=1}^{T}\left(\frac{1}{1-e^{-1}}\left(k\alpha_1 + \sum_{v\in\mathcal{V}}\sum_{i=1}^{k}\overline{x}_t(i,v)q(v)\right)\right)\right] \\
&= \frac{\eta e k T}{2\left(1-e^{-1}\right)}\left(k\alpha_1 + mQ\right),
\end{aligned}$$

where the first equality uses the expected value of the geometric random variables $\tilde{G}$, the first inequality is obtained neglecting the indicator function $B_t(i,v)$, the following equality uses the expected value of the geometric random variables B_t, the second

inequality follows by Lemma 9. We now consider the last term (III). Since $\ell_t \geq \mathbb{E}[\widehat{\ell}_t]$, from Lemma 7, we have

$$\begin{aligned}(\text{III}) &= \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\left[\left\langle \overline{x}_t(v) - a, \ell_t - \widehat{\ell}_t \right\rangle\right]\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\left[\left\langle \overline{x}_t(v), \ell_t - \widehat{\ell}_t(v) \right\rangle\right]\right] \\ &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k} \ell_t(i)\overline{x}_t(i,v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right].\end{aligned}$$

Multiplying (III) by $q(v)$ and summing over the agents, we can now upper bound $\ell_t(i)$ with 1, then we use facts that $1 - x \leq e^{-x}$ for $x \in [0,1]$ and that $e^{-y} \leq 1/y$ for all $y > 0$, to obtain

$$\begin{aligned}&\sum_{v\in\mathcal{V}} q(v)\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k} \ell_t(i)\overline{x}_t(i,v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right] \\ &\qquad\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{v\in\mathcal{V}} \overline{x}_t(i,v)\,q(v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right] \\ &\qquad= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{\substack{v\in\mathcal{V}\\ \overline{x}_t(i,v)\,q(v)>0}} \overline{x}_t(i,v)\,q(v)\left(\prod_{w\in\mathcal{N}(v)}\left(1-\overline{x}_t(i,w)\,q(w)\right)\right)^{\beta}\right] \\ &\qquad\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{\substack{v\in\mathcal{V}\\ \overline{x}_t(i,v)\,q(v)>0}} \overline{x}_t(i,v)q(v)\exp\left(-\beta\sum_{w\in\mathcal{N}(v)}\overline{x}_t(i,w)\,q(w)\right)\right] \\ &\qquad\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\sum_{\substack{v\in\mathcal{V}\\ \overline{x}_t(i,v)\,q(v)>0}} \frac{\overline{x}_t(i,v)\,q(v)}{\beta\sum_{w\in\mathcal{N}(v)}\overline{x}_t(i,w)\,q(w)}\right] \\ &\qquad\leq \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\alpha_1}{\beta}\right] = \frac{\alpha_1\,k\,T}{\beta}\end{aligned}$$

where in the last inequality follows by Lemma 8. Putting all together and recalling that

$\beta = \left\lfloor \frac{1}{k\eta} \right\rfloor \geq \frac{1}{2k\eta}$, we can conclude that for every $a \in \mathcal{A}$, thanks to Eq. (3.13), we have

$$\begin{aligned}
R_T &\leq \sum_{v \in \mathcal{V}} R_T(a, v)\, q(v) \\
&\leq Q\frac{m\,(1+\log(k))}{\eta} + Q\frac{\eta e k T}{2\,(1-e^{-1})}\left(\frac{k}{Q}\alpha_1 + m\right) + \frac{\alpha_1\, k\, T}{\beta} \\
&\leq Q\frac{m\,(1+\log(k))}{\eta} + Q\frac{\eta e k T}{2\,(1-e^{-1})}\left(\frac{k}{Q}\alpha_1 + m\right) + 2\eta\alpha_1 k^2 T \\
&= Q\frac{m\,(1+\log(k))}{\eta} + \eta Q k T\left(\frac{e}{2\,(1-e^{-1})}\left(\frac{k}{Q}\alpha_1 + m\right) + 2\alpha_1\frac{k}{Q}\right) \\
&\leq Q\frac{m\,(1+\log(k))}{\eta} + 5\eta Q k T\left(\frac{k}{Q}\alpha_1 + m\right) \\
&\leq 2Q\sqrt{10 m k T \log(k)\left(\frac{k}{Q}\alpha_1 + m\right)}\,.
\end{aligned}$$

□

3.8.5 Bounds on independence numbers

The two following lemmas provide upper bounds of sums of weights over nodes of a graph expressed in terms of its independence number.

Lemma 8. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with indedence number α_1, $q(v) \geq 0$, and $Q(v) = \sum_{w \in \mathcal{N}(v)} q(v) > 0$ for all $v \in \mathcal{V}$. Then*

$$\sum_{v \in \mathcal{V}} \frac{q(v)}{Q(v)} \leq \alpha_1$$

Proof. Initialize $V_1 = \mathcal{V}$, fix $w_1 \in \operatorname{argmin}_{w \in V_1} Q(w)$, and denote $V_2 = \mathcal{V} \setminus \mathcal{N}(w_1)$. For $k \geq 2$ fix $w_k \in \operatorname{argmin}_{w \in V_k} Q(w)$ and shrink $V_{k+1} = V_k \setminus \mathcal{N}(w_k)$ until $V_{k+1} = \varnothing$. Since $\mathcal{G}$ is undirected $w_k \notin \bigcup_{s=1}^{k-1} \mathcal{N}(w_s)$, therefore the number m of times that an action can be picked this way is upper bounded by α_1. Denoting $\mathcal{N}'(w_k) = V_k \cap \mathcal{N}(w_k)$ this implies

$$\begin{aligned}
\sum_{v \in \mathcal{V}} \frac{q(v)}{Q(v)} &= \sum_{k=1}^{m} \sum_{v \in \mathcal{N}'(w_k)} \frac{q(v)}{Q(v)} \leq \sum_{k=1}^{m} \sum_{v \in \mathcal{N}'(w_k)} \frac{q(v)}{Q(w_k)} \\
&\leq \sum_{k=1}^{m} \frac{\sum_{v \in \mathcal{N}(w_k)} q(v)}{Q(w_k)} = m \leq \alpha_1
\end{aligned}$$

concluding the proof. □

Lemma 9. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with independence number α_1. For each $v \in \mathcal{V}$, let $\mathcal{N}(v)$ be the neighborhood of node v (including v itself), and $\big(p(1,v), \ldots, p(k,v)\big)$ be a probability distribution over $\{1, \ldots, k\}$. Then, for all $i \in \{1, \ldots, k\}$,*

$$\sum_{v \in \mathcal{V}} \frac{p(i,v)}{q(i,v)} \leq \frac{1}{1-e^{-1}}\left(\alpha_1 + \sum_{v \in \mathcal{V}} p(i,v)\right) \quad \textit{where} \quad q(i,v) = 1 - \prod_{w \in \mathcal{N}(v)} \big(1 - p(i,w)\big)\,.$$

Proof. Fix $i \in \{1, \dots, k\}$ and set for brevity $P(i,v) = \sum_{w \in \mathcal{N}(v)} p(i,w)$. We can write

$$\sum_{v \in \mathcal{V}} \frac{p(i,v)}{q(i,v)} = \underbrace{\sum_{v \in \mathcal{V} \,:\, P(i,v) \geq 1} \frac{p(i,v)}{q(i,v)}}_{\text{(I)}} + \underbrace{\sum_{v \in \mathcal{V} \,:\, P(i,v) < 1} \frac{p(i,v)}{q(i,v)}}_{\text{(II)}} ,$$

and proceed by upper bounding the two terms (I) and (II) separately. Let $r(v)$ be the cardinality of $\mathcal{N}(v)$. We have, for any given $v \in \mathcal{V}$,

$$\min \left\{ q(i,v) \,:\, \sum_{w \in \mathcal{N}(v)} p(i,w) \geq 1 \right\} = 1 - \left(1 - \frac{1}{r(v)}\right)^{r(v)} \geq 1 - e^{-1} .$$

The equality is due to the fact that the minimum is achieved when $p(i,w) = \frac{1}{r(v)}$ for all $w \in \mathcal{N}(v)$, and the inequality comes from $r(v) \geq 1$ (for, $v \in \mathcal{N}(v)$). Hence

$$\text{(I)} \leq \sum_{v \in \mathcal{V} \,:\, P(i,v) \geq 1} \frac{p(i,v)}{1 - e^{-1}} \leq \sum_{v \in \mathcal{V}} \frac{p(i,v)}{1 - e^{-1}} .$$

As for (II), using the inequality $1 - x \leq e^{-x}, x \in [0,1]$, with $x = p(i,w)$, we can write

$$q(i,v) \geq 1 - \exp\left(- \sum_{w \in \mathcal{N}(v)} p(i,w) \right) = 1 - \exp\left(-P(i,v)\right) .$$

In turn, because $P(i,v) < 1$ in terms (II), we can use the inequality $1 - e^{-x} \geq (1 - e^{-1})\, x$, holding when $x \in [0,1]$, with $x = P(i,v)$, thereby concluding that

$$q(i,v) \geq (1 - e^{-1}) P(i,v)$$

Thus

$$\text{(II)} \leq \sum_{v \in \mathcal{V} \,:\, P(i,v) < 1} \frac{p(i,v)}{(1 - e^{-1}) P(i,v)} \leq \frac{1}{1 - e^{-1}} \sum_{v \in \mathcal{V}} \frac{p(i,v)}{P(i,v)} \leq \frac{\alpha_1}{1 - e^{-1}} ,$$

where in the last step we used Lemma 8. □

Chapter 4

Cooperative Online Learning with Delays

4.1 Introduction

Distributed online learning settings with communication constraints arise naturally in several applications. Consider a network of geographically distributed ad servers using real-time bidding to sell their inventory. Each server sequentially learns how to set the auction parameters (e.g., reserve price) to maximize the network's overall revenue, and shares feedback information with other servers to speed up learning. However, the rate at which data is exchanged through the communication network is slower than the typical rate at which ads are served. This causes each learner to acquire feedback information from other servers with a delay that depends on the network's structure.

Motivated by this, we introduce and analyze an online learning setting in which a network of agents solves a common online convex optimization problem, in the full and partial feedback setting, by sharing feedback with their network neighbours. We also study the impact of delay on the global performance of these agents, which do not have to be synchronized. At each time step, only some of them are requested to make a prediction and pay the corresponding loss: we call these agents "active" and the set of active agents at time t is denoted with $\mathcal{S}_t$. If $v \in \mathcal{S}_t$, it predicts with $x_t(v) \in \mathcal{X}$ and the network incurs the loss $\ell_t\big(x_t(v)\big)$. Besides observing their own feedback, each agent obtains some information previously broadcast by other agents with a delay equal to the shortest-path distance between the agents. Namely, at time t an agent learns what the active agents at shortest-path distance s did at time $t-s$ for each $s = 1, ..., d$, where d is a delay parameter. The goal is to minimize the total network regret after T time steps

$$R_T = \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \ell_t\big(x_t(v)\big) - \inf_{x \in \mathcal{X}} \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \ell_t(x)\,. \tag{4.1}$$

In words, this is the difference between the cumulative loss of the "active" agents and the loss that they would have incurred had they consistently made the best prediction in hindsight. The lack of global synchronization implies that agents who are not requested to make a prediction can get some "free feedback" with a certain delay. Since in online convex optimization the sequence of loss functions is fully arbitrary, it is not clear

whether this free feedback can improve the system's performance. Following some previous work by Cesa-Bianchi et al. [2019a] and Cesa-Bianchi et al. [2019b], we will show to which extent such improvements are possible, and we will see that we can characterize the improvement of the cooperative learning of the system in terms of the *d-th independence number* α_d of $\mathcal{G}$ is the cardinality of the biggest subset of agents, no two of which have shortest-path distance d or less.

We study the problem under two types of feedback, under the full-information feedback we study the family of algorithm of Online Mirror Descent (OMD), but, since this doesn't directly give the interesting case of Hedge we resort to a specific analysis for it that makes use of the update of Follow The Regularized Leader (FTRL). The other important case is partial information feedback. We study the case of a network of agents that cooperate to solve the same nonstochastic bandit problem, and we extend the analysis also to the case of semi-bandits on m-sets. The delays arise naturally in this distributed setting on a networks and depend on the topology of the graph itself. For this reason, the first step in our analysis is to obtain the regret for single agents that play with generic delays under the different types of feedback. This part of the analysis, presented in Section 4.3, relies heavily on previous work by Joulani et al. [2016] and Zimmert and Seldin [2019] with some minor adaptation to recover the particular case of Hedge (with adaptive learning rates) from FTRL and the case of semi-bandits on m-sets. In Section 4.4 we present the main novelty of our paper, which is an algorithm and an analysis that lets one transform a general algorithm that plays with delays into an algorithm on the communication network and retains a neat study for the total regret. We differentiate between two types of activation. In the first one, in Section 4.4.1, we treat the case of single agent activation whose analysis is more straightforward than the multiple agent activation of Section 4.4.2. We anticipate that the difference is that in the single agent activation setting, we obtain results for both full and partial information feedback. In contrast, multiple agent activation is studied just in the full information setting with the techniques developed up to there. To fill this gap, we propose an ad-hoc analysis for multiple agents activation and semi-bandits which generalizes the work of Cesa-Bianchi et al. [2019b].

4.2 Related work

The study of cooperative nonstochastic online learning on networks was pioneered by Awerbuch and Kleinberg [2008], who investigated a bandit setting in which the communication graph is a clique, agents belong to clusters characterized by the same loss, and some agents may be non-cooperative. In our multi-agent setting, the end goal is to control the total network regret (4.1). This objective was already studied by Cesa-Bianchi et al. [2019a] in the full-information case. A similar line of work was pursued by Cesa-Bianchi et al. [2019b], where the authors consider networks of learning agents that cooperate to solve the same nonstochastic bandit problem. In their setting, all agents are simultaneously active at all time steps, and the feedback propagates throughout the network with a maximum delay of d time steps, where d is a parameter of the proposed algorithm. The authors introduce a cooperative version of Exp3 that they call Exp3-COOP with regret of order $\sqrt{(d+1+K\alpha_d/N)(T\log K)}$ where K is the number of arms in the nonstochastic bandit problem, N is the total number of agents in the network,

and α_d is the independence number of the d-th power of the communication network. The case $d = 1$ corresponds to information that arrives with one round of delay and communication limited to first neighbours. In this setting Exp3-COOP has regret of order $\sqrt{(1 + K\alpha_1/N)(T \log K)}$. Cesa-Bianchi et al. [2019a] present a full information scenario where agents play instances of OMD and exchange information just with first neighbours. In a stochastic activation setting, at each time step t each agent $v \in \mathcal{G}$ is independently active with probability q_v, where q_v is a fixed and unknown number in $[0, 1]$. Under this assumption, they show that when each agent runs OMD, the network regret is $\mathcal{O}(\sqrt{\alpha_1 T})$, where $\alpha_1 \leq N$ is the independence number of the communication graph. The bound smoothly interpolates the two extreme cases of no communication ($\alpha_1 = N$) and full communication ($\alpha_1 = 1$). They also find a matching lower bound for their algorithm. More recently, Della Vecchia and Cesari [2020] considered the case of asynchronous online combinatorial semi-bandits on a network of communicating agents and stochastic activation of agents. They introduce Coop-FTPL, the first algorithm that is computationally efficient in this cooperative setting.

Our work can be seen as an extension of these settings along with two directions. On one side, we are interested in the case in which information is broadcast through the network up to a particular delay d and is successively dropped. On the other hand, we take two types of stochastic activations for the agents: a setting in which a single agent is activated per time-step, and another one, where multiple agents are activated together. From an algorithmic point of view, we extend the analysis of Cesa-Bianchi et al. [2019b] to the case of semi-bandits on m-sets where the study follows a very general proof strategy in the single agent activation, while follows more closely [Cesa-Bianchi et al., 2019b] for the multiple agents one. We point out that if the network consists of a single node, our cooperative setting always collapses into a single-agent setting. In particular, for combinatorial bandits, when the number of arms is k and $m = 1$, this becomes the well-known adversarial multiarmed bandit problem (see [Auer et al., 2002]). Hence, ours is a proper generalization of all the settings mentioned above. The main result of our work is stated in Theorem 11 for a general algorithm (for experts or bandits) in a delayed setting, with regret guarantee of the following form

$$R_T^{\text{delay}} \leq a + \sqrt{b_1 T} + \sqrt{b_2 \sum_{t=1}^{T} d_t} + \sqrt{c_1 T + c_2 \sum_{t=1}^{T} d_t}$$

where a, b_1, b_2, c_2 are constants. Our theorem states that such an algorithm has a correspondent cooperative counterpart that, in the case of single agent activation, satisfies

$$R_T^{\text{coop}} \leq a\,\alpha_d + \sqrt{b_1\,\alpha_d\,T} + \sqrt{b_2\,d\,T} + \sqrt{c_1\,\alpha_d\,T + c_2\,d\,T}\,.$$

The regret bound of the cooperative algorithm with multiple agents activation for the full-information setting is instead

$$\mathbb{E}\left[R_T^{\text{coop}}\right] \leq a\,\frac{\alpha_d + Q}{1 - e^{-1}} + Q\sqrt{\frac{b_1}{1 - e^{-1}}\left(\frac{\alpha_d}{Q} + 1\right)T} + Q\sqrt{b_2\,d\,T} + Q\sqrt{\frac{c_1}{1 - e^{-1}}\left(\frac{\alpha_d}{Q} + 1\right)T + c_2\,d\,T}\,,$$

where $Q = \sum_{v \in \mathcal{V}} q(v)$. These results are very general and explain how to pass from a delayed setting to a cooperative one, also showing how the two are deeply related. From this general formulation, it is then possible to recover the bounds for specific algorithms through a general and much more straightforward analysis. Through a unique framework, it is possible to treat the broadcasting of information through the network and the stochastic activation of agents in an elegant way.

4.3 Single agent with delay

In recent years, learning with delays has received a significant amount of attention in both stochastic and nonstochastic settings, under full and partial information feedback assumptions. In many practical applications, the learner does not have instant access to the feedback. For example, the time between clicking on a link and buying a product could be minutes, days, weeks, or longer. Similarly, the response to a drug does not come immediately. In most cases, the learner does not have the choice to wait before making the next decision because the arrival of new buyers and patients is beyond their control.

To the best of our knowledge, Weinberger and Ordentlich [2002] were the first to study online learning with delays in the full-information setting. Following their work, several extensions and variations have emerged in both stochastic and nonstochastic bandits, [Joulani et al., 2016, Cesa-Bianchi et al., 2019b, Pike-Burke et al., 2018, Desautels et al., 2014].

In this section, we will consider delayed online optimization under both full-information and partial feedback.

4.3.1 Full-information feedback with delay and linear losses

Let $\mathcal{X} \neq \varnothing$ be a convex and closed subset of $\mathbb{R}^k$, that we call *decision set*. We consider the following online protocol.

For all $t = 1, 2, \ldots$

1. a *linear loss function* $\ell_t(\cdot) = \langle \ell_t, \cdot \rangle : \mathcal{X} \to [0,1]$ and *delay* $d_t \in \{0, 1, 2, \ldots\}$ are chosen by the environment, independently of the learner's past actions
2. the learner makes a prediction $x_t \in \mathcal{X}$
3. the learner suffers a loss $\ell_t(x_t)$
4. the learner receives as *feedback* the set of pairs

$$H_t = \left\{(s, \ell_s) : s \in \{1, \ldots, t\}, s + d_s = t\right\}.$$

The goal is to minimize the *regret*, defined for any *time horizon* T by

$$R_T = \sup_{x \in \mathcal{X}} R_T(x) \qquad \text{where} \qquad R_T(x) = \sum_{t=1}^{T} \ell_t(x_t) - \sum_{t=1}^{T} \ell_t(x).$$

Note that if we know beforehand that $d_t = 0$ for all t, this setting collapses into a standard online convex optimization with full feedback. We also assume that delays

are bounded from above by a constant d. Note that this is without loss of generality. Indeed, for our application to learning on a communication network, d can be taken as the diameter of the network.

OMD with delays

We now present SOLID (Algorithm 4, Single-Instance Online Learning wIth Delays), originally introduced by Joulani et al. [2016] for delayed full-information settings. The idea behind SOLID is simple: it takes as input any algorithm BASE for non-delayed online convex optimization with full-information feedback and uses it to make updates whenever it receives a new loss function as feedback. If multiple losses are received at the same time step, they are all processed at the same time step, from the oldest to the newest.

Algorithm 4: SOLID (Single-Instance Online Learning wIth Delays)

Input: an algorithm BASE for the non-delayed setting, and its input
Initialization: let x_1 be the first prediction of BASE
for $t = 1, 2, \ldots$ **do**
  predict x_t and incur loss $\ell_t(x_t)$
  receive the feedback set H_t
  if $H_t = \varnothing$ **then**
    let $x_{t+1} \leftarrow x_t$
  else
    for *each* s *such that* $(s, \ell_s) \in H_t$, *in increasing order of* s, **do**
      update BASE with ℓ_s.
    let x_{t+1} be the next prediction of BASE

We choose as BASE the well-known Online Mirror Descent algorithm (Algorithm 5), where we denote by $\mathcal{B}_F$ the *Bregman divergence* $\mathcal{B}_F \colon \mathcal{X}' \times \operatorname{int}(\mathcal{X}') \to \mathbb{R}$ with respect to F:

$$\mathcal{B}_F(x, y) = F(x) - F(y) - \langle \nabla F(y), x - y \rangle \qquad \forall (x, y) \in \mathcal{X}' \times \operatorname{int}(\mathcal{X}') \,.$$

Algorithm 5: Online Mirror Descent (OMD)

Input: a set $\mathcal{X}' \subseteq \mathbb{R}^k$, a regularizer $F \colon \mathcal{X}' \to \mathbb{R}$ that is 1-strongly convex with respect to a norm $\|\cdot\|$ on $\mathcal{X}'$ and continuously differentiable on $\operatorname{int}(\mathcal{X}')$, a decision set $\mathcal{X} \subseteq \operatorname{int}(\mathcal{X}')$, a nonincreasing sequence of strictly positive learning rates $\eta_1, \eta_2, \ldots$, and an initial prediction $x_1 \in \mathcal{X}$
for $t = 1, 2, \ldots$ **do**
  predict x_t
  receive $\ell_t \colon \mathbb{R}^k \to \mathbb{R}$ and incur loss $\ell_t(x_t)$
  let $g_t \in \nabla \ell_t(x_t)$
  let $x_{t+1} \leftarrow \operatorname{argmin}_{x \in \mathcal{X}} \left\{ \langle g_t, x \rangle + \frac{1}{\eta_t} \mathcal{B}_F(x, x_t) \right\}$

The next result is an upper bound on the regret of SOLID run with OMD as BASE algorithm. The theorem is proven in Appendix 4.6.1 (Theorem 7) and is a direct adaptation from Joulani et al. [2016].

Theorem 7. *There is a choice of learning rates for which the regret of SOLID run with OMD as BASE algorithm against linear losses satisfies*

$$R_T \leq 2LR\sqrt{2\sum_{t=1}^{T}(1+2d_t)} + LR\sqrt{2d(2d-1)}\,,$$

where $d = \max_{t\in\{1,\dots,T\}}\{d_t\}$, *$R$ is a positive constant such that* $\max_{s\in\{1,\dots,T\}} \mathcal{B}_F(u, \widetilde{x}_s) \leq 2R^2$, *$\widetilde{x}_s$ is the prediction that OMD makes after receiving the s-th loss as feedback, and* $L = \max_{t\in\{1,\dots,T\}} \|\ell_t\|$ *is the Lipschitz constant of the linear losses.*

Hedge with delays

In this section, we focus on the delayed version (Algorithm 6) of the well-known Hedge algorithm. Unfortunately, running Online Mirror Descent with negative entropy regularizer, prediction set equal to the k-dimensional simplex of probabilities $\mathcal{X} = \Delta^{k-1}$, and adaptive learning rates does not yield to the classic anytime version of Hedge. Moreover, its regret guarantees become unbounded when the predictions get arbitrary close to the edges of the simplex.

To circumvent these problems, we use Follow The Regularized Leader (FTRL) with linear losses, negative entropy regularization, and adaptive learning rates. It is not clear if the techniques introduced by Joulani et al. [2016] that we presented above could apply to this instance of FTRL with changing learning rates because of the form of the prediction drift. For this reason, we adopt a different approach, following the work done in Zimmert and Seldin [2019] for bandits. Simplifying their analysis (which is, in turn, inspired by Joulani et al. [2016]) to the full-information scenario gives an upper bound to the regret of Hedge with delays that we state in Theorem 8. The proof of this result is deferred to Appendix 4.6.2 (Theorem 8).

Before stating the theorem, we define the *number of outstanding observations* at round t as

$$\mathfrak{o}_t = \sum_{s=1}^{t-1} \mathbb{I}\{s + d_s \geq t\}\,. \tag{4.2}$$

The quantity $\mathfrak{o}_t$ counts how many observations for the previous actions we are missing at the beginning of round t. Notably, $\mathfrak{o}_t$ is an *observable* quantity, unlike the delays d_t. As such, $\mathfrak{o}_t$ can be used for online tuning of the learning rates in the following theorem, through the quantity $\mathfrak{O}_t = \sum_{s=1}^{t} \mathfrak{o}_s$. We also note that, straightforwardly from the definitions, the following important equality holds:

$$\sum_{t=1}^{T} \mathfrak{o}_t = \sum_{t=1}^{T} d_t\,. \tag{4.3}$$

Furthermore, we use a negative entropy regularizer for the regret of Hedge with delays in the next theorem:

$$F_t(x) = \eta_t^{-1} F(x) = \eta_t^{-1} \sum_{i=1}^{k} x_i \log(x_i)\,. \tag{4.4}$$

Algorithm 6: Hedge with delays

Input: a sequence of regularizers $F_1, F_2, \dots$
Initialization: let $L_1^{obs} = 0$ and $\mathfrak{D}_0 = 0$
for $t = 1, 2, \dots$ **do**
  let $\mathfrak{D}_t \leftarrow \mathfrak{D}_{t-1} + \mathfrak{d}_t$
  let $\overline{x}_t \leftarrow \operatorname{argmin}_{x \in \Delta^{k-1}} \{\langle x, L_t^{obs} \rangle + F_t(x)\}$
  choose distribution P_t on $\{1, \dots, k\}$ with probabilities given by the components of $\overline{x}_t$
  sample an arm $x_t \in \{1, \dots, k\}$ according to the distribution P_t
  for ***each*** $s \in \{1, \dots, t\}$ *such that* $s + d_s = t$ **do**
    observe (s, ℓ_s)
    update $L_{t+1}^{obs} \leftarrow \sum_{s: s+d_s \le t} \ell_s$

Theorem 8. *The regret of Algorithm 6 run with decreasing learning rates* $(\eta_t)_{t \in \mathbb{N}}$ *and the regularizer F equal to the negative entropy in Eq. (4.4) satisfies*

$$R_T \le \frac{\ln k}{\eta_T} + \frac{1}{2} \sum_{t=1}^{T} \eta_t + \sum_{t=1}^{T} \eta_t \mathfrak{d}_t \,.$$

Furthermore if learning rates are chosen for all t, as $\eta_t = \sqrt{\frac{\ln k}{\sum_{s=1}^{t} (1 + 2\mathfrak{d}_s)}}$, *then*

$$R_T \le 2 \sqrt{(\ln k) \left(T + \sum_{t=1}^{T} d_t \right)}.$$

4.3.2 Partial information feedback with delay and linear losses

In this section we investigate the case of partial feedback that arrives with delay. We consider the following online protocol.

For all $t = 1, 2, \dots$

1. a linear *loss function* $\ell_t \colon \mathcal{X} \to [0, 1]$ and a *delay* $d_t \in \{0, 1, 2, \dots\}$ are chosen by the environment, independently of the learner's past actions
2. the learner makes a prediction $x_t \in \mathcal{X}$
3. the learner suffers a loss $\ell_t(x_t)$
4. the learner receives as *feedback* the set of pairs

$$f_t = \{(s, f_s) : s \in \{1, \dots, t\}, s + d_s = t\},$$

   where f_s is some feedback relative to the loss ℓ_t .

Note that, in the previous full-information scenario with linear losses, the feedback f_s coincides exactly with g_s, which fully determines the linear loss function $\ell_s(\cdot)$ through the identity $\ell_s(\cdot) = \langle g_s, \cdot \rangle$. The *regret* is defined as before and analogously, for $d_t = 0$ this setting collapses into a standard online convex optimization with partial information. We examine the two cases of bandits and semi-bandits for which we will define the corresponding feedback f_s.

FTRL for bandits with delays

In the bandit case, the feedback f_s is the loss $\ell_s(x_s)$ of the prediction x_s made by the learner at time s. Zimmert and Seldin [2019] studied the Follow The Regularized Leader algorithm for bandits with delay (Algorithm 7) for which they proved theoretical regret guarantees.

We recall their main result [Zimmert and Seldin, 2019, Theorem 1] below (Theorem 9). The loss estimators $\widehat{\ell}_s(i)$ used by the algorithm are defined, for all $i \in \{1, \dots, k\}$, by

$$\widehat{\ell}_s(i) = \frac{\ell_s(i)}{\overline{x}_s(i)} \mathbb{I}\{x_s = i\} \,,$$

where $\overline{x}_s$ is the algorithm's probability of selecting action x_s at round s. The cumulative observed loss estimator at time t is defined by

$$\widehat{L}_t^{obs} = \sum_{s : s + d_s < t} \widehat{\ell}_s \,.$$

The number of outstanding observations $\mathfrak{d}_t$ at round t is defined as in the previous Eq. (4.2) and similarly $\mathfrak{D}_t = \sum_{s=1}^t \mathfrak{d}_s$. The regularizer is of the form $F_t = F_{t,1} + F_{t,2}$ with the following choices for $F_{t,1}$ and $F_{t,2}$:

$$F_t(x) = \underbrace{-\sum_{i=1}^k 2\sqrt{t} x_i^{1/2}}_{F_{t,1}(x)} + \underbrace{\eta_t^{-1} \sum_{i=1}^k x_i \log(x_i)}_{F_{t,2}(x)} \,. \tag{4.5}$$

The first part of the regularizer $F_{t,1}(x) = \sqrt{t} F_1(x)$ is the $\frac{1}{2}$-Tsallis entropy $F_1(x) = -2\sum_{i=1}^k \sqrt{x_i}$ with learning rate $\frac{1}{\sqrt{t}}$, which is non-adaptive to the problem. The second part of the regularizer $F_{t,2}(x) = \eta_t^{-1} F_2(x)$ is the negative entropy $F_2(x) = \sum_{i=1}^k x_i \log(x_i)$ with adaptive learning rate η_t.

Algorithm 7: FTRL for bandits with delay

Input: a sequence of regularizers $F_1, F_2, \dots$
Initialization: let $\widehat{L}_1^{obs} = 0$ and $\mathfrak{D}_0 = 0$
for $t = 1, 2, \dots$ **do**
  let $\mathfrak{D}_t = \mathfrak{D}_{t-1} + \mathfrak{d}_t$
  let $\overline{x}_t = \operatorname{argmin}_{x \in \Delta^{k-1}} \langle x, \widehat{L}_t^{obs} \rangle + F_t(x)$
  sample $x_t \sim \overline{x}_t$
  for ***each*** $s \in \{1, \dots, t\}$ *such that* $s + d_s = t$ **do**
    observe $\big(s, \ell_s(x_s)\big)$
    construct $\widehat{\ell}_s$ and update $\widehat{L}_t^{obs}$

Theorem 9. *The regret of Algorithm 7 run with the regularizer in Eq. (4.5) and decreasing learning rates $(\eta_t)_{t \in \mathbb{N}}$ satisfies*

$$R_T \leq 4\sqrt{kT} + \eta_T^{-1} \ln k + \sum_{t=1}^T \eta_t \mathfrak{d}_t \,.$$

Furthermore if learning rates are chosen for all t, as $\eta_t^{-1} = \sqrt{\frac{2\mathfrak{D}_t}{\ln k}} = \sqrt{\frac{2\sum_{s=1}^{t}\mathfrak{d}_s}{\ln k}}$, then

$$R_T \leq 4\sqrt{kT} + \sqrt{8\sum_{t=1}^{T} d_t \ln k}\,.$$

FTRL for semi-bandits on m-sets with delays

Another important problem is finite optimization where at each round the player has to choose m alternatives out of the k possible choices. This corresponds to the set $\mathcal{A} \subseteq \{0,1\}^k$ of all vectors with exactly m ones. Note that there are $\binom{k}{m}$ such vectors and the set $\mathcal{A}$ is called m-set.

Also in this setting, we define $\widehat{L}_t^{obs}$, $\mathfrak{d}_t$ and $\mathfrak{D}_t$ as in the case of learning with bandit feedback (see previous section) and we use an instance of FTRL (see Algorithm 8) with appropriate loss estimators for learning with semi-bandits:

$$\widehat{\ell}_s(i) = \frac{\ell_s(i)\, x_s(i)}{\overline{x}_s(i)},$$

where $x_s \in \mathcal{A}$ and $\overline{x}_s \in \mathrm{co}(\mathcal{A})$ for each $s \in \{1,\dots,T\}$.

Algorithm 8: FTRL for semi-bandits on m-sets with delays

Input: Regularizer F, $\widehat{L}_1^{obs} = 0$ and $\mathfrak{D}_0 = 0$.
Initialization:
for $t = 1,\dots,T$ **do**
  set $\mathfrak{D}_t = \mathfrak{D}_{t-1} + \mathfrak{d}_t$
  set $\overline{x}_t = \mathrm{argmin}_{x\in\mathrm{co}(\mathcal{A})} \left\{ \left\langle x, \widehat{L}_t^{obs} \right\rangle + F_t(x) \right\}$
  choose distribution P_t on $\mathcal{A}$ such that $\sum_{a\in\mathcal{A}} P_t(a)a = \overline{x}_t$
  sample $x_t \sim P_t$
  for $s : s + d_s = t$ **do**
    observe $(s\,;\, x_s(1)\ell_s(1),\dots,x_s(d)\ell_s(d))$
    construct $\widehat{\ell}_s(i) = \frac{\ell_s(i)\,x_s(i)}{\overline{x}_s(i)}$ for all $i \in [k]$
    update $\widehat{L}_{t+1}^{obs} = \sum_{s:s+d_s<t+1} \widehat{\ell}_s$

What remains is to choose an appropriate regularizer. The choice of the unnormalized negentropy

$$F_t(x) = \eta_t^{-1} F(x) = \eta_t^{-1} \left(\sum_{i=1}^{k} x_i \log(x_i) - x_i \right), \tag{4.6}$$

leads to a sub-optimal regret bound

$$R_T \leq 2\sqrt{2km(1+\log(k/m))\left(T + \sum_{t=1}^{T} d_t\right)}, \tag{4.7}$$

for an appropriate choice of learning rates $\eta_t = \sqrt{\frac{m(1+\log(k/m))}{2k\sum_{s=1}^{t}(\mathfrak{d}_s+1)}}$. The proof of this bound is anyway presented in Appendix 4.6.4 and is again an adaptation of the proof by Zimmert and Seldin [2019] for bandits.

A different choice of regularizer leads to a better regret bound like for the case of bandits with delays in Zimmert and Seldin [2019]. Therefore, also in this case we use the hybrid regularizer of Eq. (4.5) and run Algorithm 8 with this new choice. In Appendix 4.6.5 (Theorem 10) we prove the following bound for such algorithm.

Theorem 10. *Algorithm 8 with proper learning rates $(\eta_t)_{t=1,\dots,n}$ and a choice of regularizer as in Eq. (4.5) satisfies*

$$\begin{aligned} R_T &\leq 2\sqrt{Tkm} + \frac{m\log\left(\frac{k}{m}\right)}{\eta_T} + \sqrt{Tkm} + \sum_{t=1}^{T}\eta_t k\mathfrak{d}_t \\ &\leq \frac{m\log\left(\frac{k}{m}\right)}{\eta_T} + 3\sqrt{Tkm} + k\sum_{t=1}^{T}\eta_t\mathfrak{d}_t \end{aligned}$$

Furthermore if one chooses $\eta_t = \sqrt{\frac{m(1+\log(k/m))}{2k\sum_{s=1}^{t}\mathfrak{d}_s}}$ then

$$R_T \leq 3\sqrt{Tkm} + 2\sqrt{2km\log(k/m)\left(\sum_{t=1}^{T}d_t\right)}.$$

4.4 From delayed single-agent to cooperative multi-agent

Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph. We say that $\mathcal{G}$ is a *communication* network and $\mathcal{V}$ is the set of *agents*. For any agent $v \in \mathcal{V}$ and all $d \in \mathbb{N}$, we denote by $\mathcal{N}_d(v)$ the set of nodes containing agent v and all agents w with $\delta_{\mathcal{G}}(v,w) \leq d$, where $\delta_{\mathcal{G}}(v,w)$ is the shortest-path distance on the graph. The *d-th independence number* α_d of $\mathcal{G}$ is the cardinality of the biggest subset of agents, no two of which have shortest-path distance d or less.

We study the following cooperative online convex optimization protocol: initially, hidden from the agents, the environment picks a sequence of random subsets $\mathcal{S}_1, \mathcal{S}_2, \dots \subseteq \mathcal{V}$ of *active* agents and a sequence of differentiable convex real loss functions $\ell_1, \ell_2, \dots$ defined on a convex decision set $\mathcal{X} \subset \mathbb{R}^k$.

Then, for each time step $t = 1, 2, \dots$:

1. for each active agent $v \in \mathcal{S}_t$
   (a) v predicts with $x_t(v) \in \mathcal{X}$ and the network incurs the loss $\ell_t(x_t(v))$;
   (b) v receives some feedback $f_t(v)$ and sends the message $m_t(v) = \langle t, v, f_t(v)\rangle$;
2. for each agent $v \in \mathcal{V}$
   (a) v receives from its neighbours all past messages $m_{t-s} = \langle t-s, v', f_{t-s}\rangle$ such that $v' \in \mathcal{S}_{t-s}$ and $\delta_G(v,v') = s \in \{1,\dots,d\}$;
   (b) v drops the messages that are older than $t-d$ and forwards the remaining ones;

(c) v (possibly) updates its local model and, depending on the setting, sends to its neighbors a message $m'_t(v) = \langle t, v, i_t(v) \rangle$ containing some local information $i_t(v)$.

The goal is to minimize the *network regret* as a function of the number T of time steps:

$$R_T = \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \ell_t\big(x_t(v)\big) - \inf_{x \in \mathcal{X}} \sum_{t=1}^{T} \sum_{v \in \mathcal{S}_t} \ell_t(x)\,. \tag{4.8}$$

Note that only the losses of active agents contribute to the network regret. We will study the problem under different feedback types and under two types of activation mechanisms.

Feedback type. We consider the bandit, semi-bandit, and full-info feedback types that we introduced in Section 4.3. In the bandit case, the feedback $f_t(v)$ received by v at time t (line 1 of the protocol) is the loss $\ell_t\big(x_t(v)\big)$ of the prediction $x_t(v)$. In the semi-bandit case, $\mathcal{X} = \mathrm{co}(\mathcal{A})$, where $\mathcal{A} = \big\{a \in \{0,1\}^k : \sum_{i=1}^k a_i = m\big\}$ for some $m \in \mathbb{N}$, the agent v predicts with $x_t(v) \in \mathcal{A}$, losses ℓ_t are linear, and the feedback $f_t(v)$ is the vector $\big(\ell_{t,1}x_{t,1}(v), \dots, \ell_{t,k}x_{t,k}(v)\big)$ where, with a slight abuse of notation, we write $\ell_t(x) = \langle \ell_t, x \rangle$ (i.e., we think of ℓ_t as a vector in $\mathbb{R}^k$). In the full-information case, the feedback $f_t(v)$ is the whole loss function ℓ_t.

Activations. We consider the two distinct settings in which a single agent is activated at each time step or multiple agents are. In the single-activation setting, we assume that there exists a distribution q on $\mathcal{V}$ and, for each time step t, the set $\mathcal{S}_t$ contains only one agent that is drawn i.i.d. from q. In the multiple-activation setting, we assume that there exists an activation probability $q(v) \in [0,1]$ for each agent $v \in \mathcal{V}$ and, at each time step t, each agent $v \in \mathcal{V}$ is activated i.i.d. with probability $q(v)$. In the case of full-information we are able to treat multiple activations, while the single activation setting is the one adopted for learning with partial feedback. This difference is due to technical reasons related to the difficulty of building loss estimators for partial feedback when the pieces of information on the loss at a specific time can arrive at possibly different later rounds. Despite these challenges, we present in Section 4.5 a different analysis for dealing with multiple agent activations and semi-bandit feedback.

Now we show how the regret guarantees of an algorithm for a delayed single-agent v (Theorems 7, 8, 9, 10) translates to cooperative multi-agent setting (for different choices of feedback type and activations). To this end, we define the random variable $D_t(v)$ by

$$D_t(v) = \sum_{s=t+1}^{t+\delta_t(v)} \mathbb{I}\big\{\exists v' \in \mathcal{S}_s \cap \mathcal{N}_d(v)\big\}\,, \qquad \text{where} \qquad \delta_t(v) = \min_{v' \in \mathcal{S}_t} \delta_{\mathcal{G}}(v', v)\,,$$

with the understanding that $D_t(v) = 0$ if $\delta_t(v) = 0$. At a high-level, the random variable $D_t(v)$ represents the delay of loss ℓ_t from the perspective of v.

We define the total delay of loss ℓ_t from the perspective of v by

$$\sum_{t=1}^{T} D_t(v)\, \mathbb{I}\big\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\big\}.$$

Algorithm 1 Delay Into Cooperation (DIC)

input: maximum delay d, single-agent non-delayed algorithm BASE for each time step $t = 1, 2, \ldots$

1. for each active agent $v \in \mathcal{S}_t$
   (a) v outputs the prediction $x_t(v) \in \mathcal{X}$ generated by SOLID(BASE)
   (b) v receives some feedback $f_t(v)$ and sends the message $m_t(v) = \langle t, v, f_t(v) \rangle$
2. for each agent $v \in \mathcal{V}$
   (a) v receives from its neighbours all past messages $m_{t-s}(w)$ and $m'_{t-s}(w)$ (see last item) such that $w \in \mathcal{S}_{t-s}$ and $\delta_G(v, w) = s \in \{1, \ldots, d\}$;
   (b) v drops the messages that are older than $t - d$ and forwards the remaining ones
   (c) v makes a step of the SOLID(BASE) algorithm for each newly received message
   (d) depending on the setting, v sends to its neighbors a message $m'_t(v) = \langle t, v, i_t(v) \rangle$ containing some local information $i_t(v)$

The following lemma controls the total (expected) delay of the losses delay of from the perspective of each node v.

Lemma 10. *For all agents $v \in \mathcal{V}$, for both single and multiple agents activation, we have*

$$\mathbb{E}\left[\sum_{t=1}^{T} D_t(v)\, \mathbb{I}\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\}\right] \leq T\, d\, Q_d(v)^2\,,$$

where $Q_d(v) = \mathbb{P}\big[\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\big]$.

Proof. For all agents $v \in \mathcal{V}$, we have

$$
\begin{aligned}
&\mathbb{E}\left[\sum_{t=1}^{T} D_t\, \mathbb{I}\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\}\right] \\
&\qquad = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\Big[D_t\, \mathbb{I}\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\}\Big]\right] \\
&\qquad = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\Big[D_t\, \mathbb{I}\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\} \mid \delta_t(v)\Big]\right] \\
&\qquad = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\big[D_t \mid \delta_t(v)\big]\, \mathbb{E}_t\Big[\mathbb{I}\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\} \mid \delta_t(v)\Big]\right] \\
&\qquad = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\big[D_t \mid \delta_t(v)\big]\, \mathbb{I}\{\delta_t(v) \le d\}\right] \\
&\qquad = \mathbb{E}\left[\sum_{t=1}^{T} \sum_{b=0}^{\operatorname{diam}(\mathcal{G})} \mathbb{E}_t\big[D_t \mid \delta_t(v) = b\big]\, \mathbb{P}_t\big[\delta_t(v) = b\big]\, \mathbb{I}\{\delta_t(v) \le d\}\right] \\
&\qquad = \mathbb{E}\left[\sum_{t=1}^{T} \sum_{b=0}^{d} \mathbb{E}_t\big[D_t \mid \delta_t(v) = b\big]\, \mathbb{P}\big[\delta_t(v) = b\big]\right] \\
&\qquad = \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\big[D_t \mid \delta_t(v) = 0\big]\, q(v)\right] \\
&\qquad\quad + \mathbb{E}\left[\sum_{t=1}^{T} \sum_{b=1}^{d} \mathbb{E}_t\big[D_t \mid \delta_t(v) = b\big]\big(Q_b(v) - Q_{b-1}(v)\big)\big(1 - Q_{b-1}(v)\big)\right] \\
&\qquad \le \mathbb{E}\left[\sum_{t=1}^{T} \mathbb{E}_t\big[D_t \mid \delta_t(v) = 0\big]\, q(v)\right] \\
&\qquad\quad + \mathbb{E}\left[\sum_{t=1}^{T} \sum_{b=1}^{d} \mathbb{E}_t\big[D_t \mid \delta_t(v) = b\big]\big(Q_b(v) - Q_{b-1}(v)\big)\right] \\
&\qquad = \sum_{t=1}^{T} \sum_{b=1}^{d} b\, Q_d(v)\big(Q_b(v) - Q_{b-1}(v)\big) \\
&\qquad \le \sum_{t=1}^{T} d\, Q_d(v) \sum_{b=1}^{d} \big(Q_b(v) - Q_{b-1}(v)\big) \\
&\qquad \le \sum_{t=1}^{T} d\, Q_d(v)^2 \\
&\qquad = T\, d\, Q_d(v)^2,
\end{aligned}
$$

where we used the fact that $\mathbb{P}\big[\delta_t(v) = b\big] = \big(Q_b(v) - Q_{b-1}(v)\big)\big(1 - Q_{b-1}(v)\big)$. □

Theorem 11. *Fix any algorithm (for experts or bandits) for the delayed setting having regret*

guarantees

$$R_T^{\text{delay}} \leq a + \sqrt{b_1 T} + \sqrt{b_2 \sum_{t=1}^{T} d_t} + \sqrt{c_1 T + c_2 \sum_{t=1}^{T} d_t}\,,$$

where the quantities a, b_1, b_2, c_2 are positive and possibly depend in arbitrary ways on all the relevant parameters of the problem and $d_1, \ldots, d_T$ are the delays and they are upper bounded by d. Then, the regret of the correspondent cooperative algorithm with single agent activation satisfies

$$R_T^{\text{coop}} \leq a\,\alpha_d + \sqrt{b_1\,\alpha_d\,T} + \sqrt{b_2\,d\,T} + \sqrt{c_1\,\alpha_d\,T + c_2\,d\,T}\,,$$

where α_d is the d-th independence number of graph $\mathcal{G}$. The regret bound of the cooperative algorithm with multiple agents activation for the full-information setting satisfies insted

$$\mathbb{E}\left[R_T^{\text{coop}}\right] \leq a\,\frac{\alpha_d + Q}{1 - e^{-1}} + Q\sqrt{\frac{b_1}{1 - e^{-1}}\left(\frac{\alpha_d}{Q} + 1\right)T} + Q\sqrt{b_2\,d\,T} + Q\sqrt{\frac{c_1}{1 - e^{-1}}\left(\frac{\alpha_d}{Q} + 1\right)T + c_2\,d\,T}\,.$$

where $Q = \sum_{v \in \mathcal{V}} q(v)$.

Proof. Fix any $x \in \mathcal{X}$ and let $r_t(v) = \ell_t\left(x_t(v)\right) - \ell_t(x)$. The regret of agent v on a communication network is

$$\sum_{t=1}^{T} r_t(v)\mathbb{I}\left\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\right\} \leq a + \sqrt{b_1 \sum_{t=1}^{T} \mathbb{I}\left\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\right\}} + \sqrt{b_2 \sum_{t=1}^{T} D_t(v)\,\mathbb{I}\left\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\right\}} + \sqrt{c_1 \sum_{t=1}^{T} \mathbb{I}\left\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\right\} + c_2 \sum_{t=1}^{T} D_t(v)\,\mathbb{I}\left\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\right\}}\,.$$

We now take expectations to both sides with respect to the activations of the nodes. The expectation of the left-hand side is

$$\mathbb{E}\left[\sum_{t=1}^{T} r_t(v)\,\mathbb{I}\left\{\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\right\}\right] = \mathbb{E}\left[\sum_{t=1}^{T} r_t\left(v\right) Q_d(v)\right].$$

By Jensen's inequality and Lemma 10, the expectation of the right-hand side can be upper bounded by

$$a + \sqrt{b_1\,Q_d(v)\,T} + \sqrt{b_2\,d\,Q_d(v)^2\,T} + \sqrt{c_1\,Q_d(v)\,T + c_2\,d\,Q_d(v)^2\,T}\,,$$

Putting everything together and dividing both sides by $Q_d(v)$ yields

$$\mathbb{E}\left[\sum_{t=1}^{T} r_t\left(v\right)\right] \leq \frac{a}{Q_d(v)} + \sqrt{b_1\,\frac{1}{Q_d(v)}\,T} + \sqrt{b_2\,d\,T} + \sqrt{c_1\,\frac{1}{Q_d(v)}\,T + c_2\,d\,T}\,,$$

Hence, by Jensen's inequality, we get

$$\mathbb{E}\left[R_T^{\text{coop}}\right] = \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}} r_t(v)\mathbb{I}\{v\in\mathcal{S}_t\}\right] = \sum_{t=1}^{T}\sum_{v\in\mathcal{V}} q(v)\mathbb{E}\left[r_t(v)\right] = \sum_{v\in\mathcal{V}} q(v)\sum_{t=1}^{T}\mathbb{E}\left[r_t(v)\right]$$
$$\leq \sum_{v\in\mathcal{V}} q(v)\left(\frac{a}{Q_d(v)} + \sqrt{b_1\frac{1}{Q_d(v)}T} + \sqrt{b_2\, d\, T} + \sqrt{c_1\frac{1}{Q_d(v)}T + c_2\, d\, T}\right). \tag{4.9}$$

The probability $Q_d(v) = \mathbb{P}\left[\exists v' \in \mathcal{S}_t \cap \mathcal{N}_d(v)\right]$ has the following two expressions depending on the type of activation

$$Q_d(v) = \begin{cases} \sum_{v'\in\mathcal{N}_d(v)} q(v') & \text{for single agent activation,} \\ 1 - \prod_{v'\in\mathcal{N}_d(v)}(1-q(v')) & \text{for multiple agents activation.} \end{cases}$$

Therefore continuing from Eq. (4.9) we have

$$\sum_{v\in\mathcal{V}} q(v)\left(\frac{a}{Q_d(v)} + \sqrt{b_1\frac{1}{Q_d(v)}T} + \sqrt{b_2\, d\, T} + \sqrt{c_1\frac{1}{Q_d(v)}T + c_2\, d\, T}\right)$$
$$= a\sum_{v\in\mathcal{V}}\frac{q(v)}{Q_d(v)} + Q\sum_{v\in\mathcal{V}}\frac{q(v)}{Q}\sqrt{b_1\frac{1}{Q_d(v)}T} + Q\sqrt{b_2\, d\, T} + Q\sum_{v\in\mathcal{V}}\frac{q(v)}{Q}\sqrt{c_1\frac{1}{Q_d(v)}T + c_2\, d\, T}$$
$$\leq a\sum_{v\in\mathcal{V}}\frac{q(v)}{Q_d(v)} + Q\sqrt{b_1\frac{1}{Q}\sum_{v\in\mathcal{V}}\frac{q(v)}{Q_d(v)}T} + Q\sqrt{b_2\, d\, T} + Q\sqrt{c_1\frac{1}{Q}\sum_{v\in\mathcal{V}}\frac{q(v)}{Q_d(v)}T + c_2\, d\, T},$$

where we recall that $Q = 1$ for single agent activation.

The analysis of the quantity $\sum_{v\in\mathcal{V}} q(v)/Q_d(v)$ differs at this point for the two types of activations. For single agent activation, using Lemma 2 in Cesa-Bianchi et al. [2019a], we get

$$\mathbb{E}\left[R_T^{\text{coop}}\right] \leq a\,\alpha_d + \sqrt{b_1\,\alpha_d\, T} + \sqrt{b_2\, d\, T} + \sqrt{c_1\,\alpha_d\, T + c_2\, d\, T}.$$

In the case of multiple agents activation we use Lemma 14 (setting $p(i,v)$ equal to $q(v)$) and this leads to the following bound

$$\mathbb{E}\left[R_T^{\text{coop}}\right] \leq a\frac{\alpha_d + Q}{1-e^{-1}} + Q\sqrt{\frac{b_1}{1-e^{-1}}\left(\frac{\alpha_d}{Q}+1\right)T} + Q\sqrt{b_2\, d\, T}$$
$$+ Q\sqrt{\frac{c_1}{1-e^{-1}}\left(\frac{\alpha_d}{Q}+1\right)T + c_2\, d\, T}.$$

□

4.4.1 Cooperative learning with single agent activation

At this point we have all the tools to show how the regret guarantees of the algorithms in Theorems 7, 8, 9, 10, translate when algorithms are played in a cooperative multi-agent setting. In this section anyway we treat the case of single agent activation which is

the simplest and is available for all the algorithms that we have presented so far. We postpone to the next section a description of the technical difficulties encountered when trying to do the same thing in a partial information setting. Going in the same order of Section 4.3 we have the following corollaries.

Corollary 2. *The regret of DIC, when it is run with maximum delay d in a single agent activation setting and the BASE algorithm is OMD, has regret bound that satisfies*

$$R_T^{\text{coop}} \leq 2LR\sqrt{2T(\alpha_d + 2d)} + LR\alpha_d\sqrt{2d(d+1)}.$$

Proof. Exploiting the result of Theorem 7 we have the following regret

$$R_T \leq 2LR\sqrt{2\sum_{t=1}^{T}(1+2d_t)} + LR\sqrt{2d(d+1)},$$

and from Theorem 11 we obtain the following regret for the communication network

$$R_T^{\text{coop}} \leq 2LR\sqrt{2T(\alpha_d + 2d)} + LR\alpha_d\sqrt{2d(d+1)}.$$

□

Under some mild conditions OMD and FTRL are the same family of algorithms (see Orabona [2019]). Therefore, we are interested to give a regret bound for a special member of these families which is Hedge. The regret of Hedge on a communication network follows in the corollary below.

Corollary 3. *The regret of Hedge, when it is run with maximum delay d in a single agent activation setting is*

$$R_T^{\text{coop}} \leq 2\sqrt{\log(k)T\,(\alpha_d + d)}.$$

Proof. Exploiting the result of Theorem 8, we have the following regret

$$R_T \leq 2\sqrt{\log(k)\left(T + \sum_{t=1}^{T} d_t\right)}.$$

and from Theorem 11 we obtain the following regret for the communication network

$$R_T^{\text{coop}} \leq 2\sqrt{\log(k)T\,(\alpha_d + d)}.$$

□

For the case of bandit feedback on the simplex like in Section 4.3.2 we have the following corollary.

Corollary 4. *The regret of FTRL for bandits, when it is run with maximum delay d in a single agent activation setting is*

$$R_T^{\text{coop}} \leq 4\sqrt{\alpha_d kT} + \sqrt{8Td\log k}.$$

Proof. Exploiting the result of Theorem 9 we have the following regret

$$R_T \leq 4\sqrt{kT} + \sqrt{8 \sum_{t=1}^{T} d_t \log k}.$$

and from Theorem 11 we obtain the following regret for the communication network

$$R_T^{\text{coop}} \leq 4\sqrt{\alpha_d kT} + \sqrt{8Td \log k}.$$

□

The cooperative regret of the optimal algorithm for learning with semi-bandit feedback is given in the following corollary.

Corollary 5. *The regret of FTRL for semi-bandits on m-sets with the regularizer in Eq. (4.5), when it is run with maximum delay d in a single agent activation setting is*

$$R_T^{\text{coop}} \leq 3\sqrt{\alpha_d Tkm} + 2\sqrt{2km \log(k/m)\, dT}.$$

Proof. Exploiting the result of Theorem 10 we have the following regret

$$R_T \leq 3\sqrt{Tkm} + 2\sqrt{2km \log(k/m) \left(\sum_{t=1}^{T} d_t \right)}.$$

and from Theorem 11 we obtain the following regret for the communication network

$$R_T^{\text{coop}} \leq 3\sqrt{\alpha_d Tkm} + 2\sqrt{2km \log(k/m)\, dT}.$$

□

We note that with the regularizer in (4.6), we obtain the following regret bound instead (which is not optimal):

$$R_T^{\text{coop}} \leq 2\sqrt{2km\,(1 + \log(k/m))\, T\,(\alpha_d + d)}.$$

4.4.2 Cooperative learning with multiple agents activation

The case of partial information is more complicated to treat for multiple agents activation than the full-info. This stems from the fact that is less trivial to construct the estimators for the losses. In fact, given a node v and its neighbourhood $\mathcal{N}_d(v)$ let us assume is possible to take two different nodes $v', v'' \in \mathcal{N}_d(v)$ such that $\delta_{\mathcal{G}}(v, v') \neq \delta_{\mathcal{G}}(v, v'')$. Furthermore, assuming $v', v'' \in \mathcal{S}_t$ they received feedbacks $f_t(v')$ and $f_t(v'')$ and both of these feedbacks contain some, a priori different, information on the loss ℓ_t. Since v' and v'' have different distances from v the two feedbacks will get to node v with two different delays and this doesn't allow for a direct applications of the techniques proposed in Zimmert and Seldin [2019]. For this reason the corollaries of this section just refer to the full-information case and are the following two.

Corollary 6. *The regret of DIC, when it is run with maximum delay d in a multiple agents activation setting and the BASE algorithm is OMD, has regret bound that satisfies*

$$R_T^{\text{coop}} \leq 2LRQ\sqrt{\frac{2}{1-e^{-1}}\left(\frac{\alpha_d}{Q}+1\right)T+4dT}+LR\frac{\alpha_d+Q}{1-e^{-1}}\sqrt{2d(d+1)}.$$

Proof. Exploiting the result of Theorem 7 we have the following regret

$$R_T \leq 2LR\sqrt{2\sum_{t=1}^{T}(1+2d_t)}+LR\sqrt{2d(d+1)},$$

and from Theorem 11 we obtain the following regret for the communication network

$$R_T^{\text{coop}} \leq 2LRQ\sqrt{\frac{2}{1-e^{-1}}\left(\frac{\alpha_d}{Q}+1\right)T+4dT}+LR\frac{\alpha_d+Q}{1-e^{-1}}\sqrt{2d(d+1)}.$$

□

Corollary 7. *The regret of Hedge run in a cooperative multiple agents activation setting, with maximum delay d is*

$$R_T^{\text{coop}} \leq 2Q\sqrt{\log(k)\ T\left(\frac{1}{1-e^{-1}}\left(\frac{\alpha_d}{Q}+1\right)+d\right)}.$$

Proof. Exploiting the result of Theorem 8 we have the following regret

$$R_T \leq 2\sqrt{\log(k)\left(T+\sum_{t=1}^{T}d_t\right)}.$$

and from Theorem 11 we obtain the following regret for the communication network

$$R_T^{\text{coop}} \leq 2Q\sqrt{\log(k)\ T\left(\frac{1}{1-e^{-1}}\left(\frac{\alpha_d}{Q}+1\right)+d\right)}.$$

□

4.5 Cooperative multiple agents activation setting for semi-bandits

As we anticipated in Section 4.4 the case of partial information is more complicated to treat for multiple agents activation than the full-info. For this reason we propose here Algorithm 2 that uses the loss estimator in Eq. (4.10). Its analysis is similar in spirit to the analysis contained in Cesa-Bianchi et al. [2019b], the main difference here is that we

Algorithm 2 baditCoopMsets

Paramters: learning rate $\eta > 0$

Initialization: each agent $v \in \mathcal{V}$ sets weights $w_1(i,v) = 1/k$ and $\overline{x}_1(i,v) = m/k$ for all $i \in \{1,\dots,k\}$

For: $t = 1, 2, \dots$

1. for each active agent $v \in \mathcal{S}_t$

   (a_1) v computes a probability distribution $P_t(v) = \big(P_t(a,v)\big)_{a\in\mathcal{A}}$ on $\mathcal{A}$ such that

   $$\overline{x}_t(i,v) = \sum_{a\in\mathcal{A}} a_i\, P_t(a,v), \qquad \forall i \in \{1,\dots,k\}$$

   (a_2) v outputs the prediction $x_t(v) \in \mathcal{A}$ drawn according to $P_t(v)$

   (b) v receives the feedback $f_t(v) = (x_{t,1}\ell_{t,1}, \dots, x_{t,k}\ell_{t,k})$ and sends the message

   $$m_t(v) = \big\langle t, v, f_t(v)\big\rangle$$

2. for each agent $v \in \mathcal{V}$

   (a) v receives from its neighbours all past messages $m_{t-s}(w)$ and $m'_{t-s}(w)$ (see last item) such that $w \in \mathcal{S}_{t-s}$ and $\delta_G(v,w) = s \in \{1,\dots,d\}$

   (b) v drops the messages that are older than $t-d$ and forwards the remaining ones

   (c_1) v performs the update

   $$w_{t+1}(i,v) = \overline{x}_t(i,v)\, \exp\big(-\eta\, \widehat{\ell}_t(i,v)\big), \qquad \forall i \in \{1,\dots,k\} \tag{4.10}$$

   where

   $$\widehat{\ell}_t(i,v) = \begin{cases} \frac{\ell_{t-d}(i)}{\overline{B}_{d,t-d}(i,v)} B_{d,t-d}(i,v) & \text{if } t > d \\ 0 & \text{otherwise} \end{cases}$$

   with

   $$B_{d,t-d}(i,v) = \mathbb{I}\big\{\exists v' \in \mathcal{N}_d(v) \cap \mathcal{S}_{t-d} : x_{t-d}(i,v') = 1\big\}$$

   and

   $$\overline{B}_{d,t-d}(i,v) = 1 - \prod_{v'\in\mathcal{N}_d(v)} \Big(1 - \overline{x}_{t-d}(i,v')\, q(v')\Big)$$

   (c_2) each agent $v \in \mathcal{V}$ computes $\overline{x}_t(v) = \big(\overline{x}_t(1,v), \dots, \overline{x}_t(k,v)\big)$ as

   $$\overline{x}_{t+1}(i,v) = m\frac{w_{t+1}(i,v)}{W_{t+1}(v)}, \quad \forall i \in \{1,\dots,k\} \quad \text{where} \quad W_{t+1}(v) = \sum_{j=1}^{k} w_{t+1}(j,v)$$

   (d_1) each agent $v \in \mathcal{S}_t$ sends to its neighbors the message $m'_t(v) = \big\langle t, v, i_t(v)\big\rangle$, where $i_t(v) = \big\langle \overline{x}_t(v), x_t(v)\big\rangle$

   (d_2) if $t > 1$ each agent $v \in \mathcal{V} \setminus \mathcal{S}_t$ such that $\bar{x}_t(v) \neq \bar{x}_{t-1}(v)$ sends to its neighbors the message $m'_t(v) = \big\langle t, v, i_t(v)\big\rangle$, where $i_t(v) = \big\langle \overline{x}_t(v)\big\rangle$

extend the analysis therein to the case of semi-bandit feedback on m-sets and stochastic activation of the agents on the communication network. The following lemma provides a deterministic bound for a single agent v, its analysis mimics very closely the anlysis that is done for EXP3 and its proof can be found in Appendix 4.6.6, Lemma 11.

Lemma 11. *If agent $v \in \mathcal{V}$ runs the Algorithm 2 with learning rate $\eta > 0$, the following deterministic bound holds for all $i \in \{1, \dots, k\}$:*

$$\sum_{t=1}^{T} \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) - \sum_{t=1}^{T} \widehat{\ell}_t(i^\star, v) \leq \frac{\ln k}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v)^2 \,.$$

Next we have a lemma to bound the additive drift of the algorithm in a semibandit setting and it is the analogous of Lemma 1 in Cesa-Bianchi et al. [2019b] for bandits. We postpone its proof to Appendix 4.6.6, Lemma 12.

Lemma 12. *If agent $v \in \mathcal{V}$ runs baditCoopMsets with learning rate $\eta > 0$, the following deterministic bounds for the drift probabilities hold for all $i \in \{1, \dots, k\}$:*

$$-\eta \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) \leq \frac{\overline{x}_{t+1}(i,v)}{m} - \frac{\overline{x}_t(i,v)}{m} \leq \eta \frac{\overline{x}_{t+1}(i,v)}{m} \sum_{j=1}^{k} \frac{\overline{x}_t(j,v)}{m} \widehat{\ell}_t(j,v) \,.$$

A lemma to bound the drift in a multiplicative way follows. It is the analogous of Lemma 2 in Cesa-Bianchi et al. [2019b] for bandits. Its proof is in Appendix 4.6.6, Lemma 13.

Lemma 13. *If agent $v \in \mathcal{V}$ runs baditCoopMsets with learning rate $\eta \in \left(0, \frac{m}{ke(d+1)}\right)$, the following deterministic bound holds for all $i \in \{1, \dots, k\}$:*

$$\overline{x}_{t+1}(i,v) \leq \left(1 + \frac{1}{d}\right) \overline{x}_t(i,v) \,.$$

Finally, a last lemma is used in Theorem 12 to link the regret for the network, that is obtained summing over the agents, to the independence number of the orrespondig graph, which is characheristic of the network topology (see Cesa-Bianchi et al. [2019b] for a proof).

Lemma 14. *Let $\mathcal{G} = (\mathcal{V}, \mathcal{E})$ be an undirected graph with independence number α_1. For each $v \in V$, let $\mathcal{N}_1(v)$ be the neighborhood of node v (including v itself), and $p_t(v) = \big(p(1,v), \dots, p(k,v)\big)$ has positive entries. Then, for all $i \in [k]$,*

$$\sum_{v \in \mathcal{V}} \frac{p(i,v)}{q(i,v)} \leq \frac{1}{1 - e^{-1}} \left(\alpha_1 + \sum_{v \in \mathcal{V}} p(i,v) \right) \quad \text{where} \quad q(i,v) = 1 - \prod_{v' \in \mathcal{N}_1(v)} \big(1 - p(i,v')\big) \,.$$

Theorem 12. *If baditCoopMsets (Algorithm 2) is run with $\eta > 0$, its regret satisfies*

$$R_T \leq 2dQm + \frac{m \ln k}{\eta} Q + 4\eta T Q \left(\frac{k}{Q} \alpha_d + md \right) \,,$$

where $Q = \sum_{v\in\mathcal{V}} q(v)$. *Choosing, in particular,* $\eta = Q\sqrt{(m\ln k)/\left(4TQ\left((k/Q)\alpha_d + md\right)\right)}$, *yields*

$$R_T \le 2dQm + 2Q\sqrt{mT\ln(k)\left(\frac{k}{Q}\alpha_d + md\right)}\,.$$

Proof. The standard analysis of the exponentially-weighted algorithm in Lemma 11 with importance-sampling estimates gives for each agent v and and each $i^\star \in \{1,\dots,k\}$, the deterministic bound

$$\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{m}\widehat{\ell}_t(i,v) - \sum_{t=1}^{T}\widehat{\ell}_t(i^\star,v) \le \frac{\ln k}{\eta} + \frac{\eta}{2}\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{m}\widehat{\ell}_t(i,v)^2\,. \tag{4.11}$$

Iterative applications of the first inequality in Lemma 12 gives, for $t > d$,

$$\frac{\overline{x}_t(i,v)}{m} \ge \frac{\overline{x}_{t-d}(i,v)}{m} - \eta\sum_{s=1}^{k}\frac{\overline{x}_{t-s}(i,v)}{m}\widehat{\ell}_{t-s}(i,v)\,,$$

so that, setting for brevity $A_t(i,v) = \sum_{s=1}^{k}\frac{\overline{x}_{t-s}(i,v)}{m}\widehat{\ell}_{t-s}(i,v)$ we have

$$\begin{aligned}\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{m}\widehat{\ell}_t(i,v) &\ge \sum_{t=2d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{m}\widehat{\ell}_t(i,v)\\ &\ge \sum_{t=2d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_{t-d}(i,v)}{m}\widehat{\ell}_t(i,v) - \eta\sum_{t=2d+1}^{T}\sum_{i=1}^{k}A_t(i,v)\widehat{\ell}_t(i,v)\,.\end{aligned}$$

Hence

$$\begin{aligned}&\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{m}\widehat{\ell}_t(i,v)\right]\\ &\ge \mathbb{E}\left[\sum_{t=2d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_{t-d}(i,v)}{m}\widehat{\ell}_t(i,v)\right] - \eta\mathbb{E}\left[\sum_{t=2d+1}^{T}\sum_{i=1}^{k}A_t(i,v)\widehat{\ell}_t(i,v)\right]\\ &= \mathbb{E}\left[\sum_{t=2d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_{t-d}(i,v)}{m}\mathbb{E}_{t-d}\left[\widehat{\ell}_t(i,v)\right]\right] - \eta\mathbb{E}\left[\sum_{t=2d+1}^{T}\sum_{i=1}^{k}A_t(i,v)\mathbb{E}_{t-d}\left[\widehat{\ell}_t(i,v)\right]\right]\\ &= \mathbb{E}\left[\sum_{t=2d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_{t-d}(i,v)}{m}\ell_{t-d}(i,v)\right] - \eta\mathbb{E}\left[\sum_{t=2d+1}^{T}\sum_{i=1}^{k}A_t(i,v)\ell_{t-d}(i,v)\right]\\ &\ge \mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{m}\ell_t(i)\right] - 2d - \eta Td\end{aligned}$$

where the last step follows by

$$\begin{aligned}\mathbb{E}\left[\sum_{i=1}^{k}A_t(i,v)\ell_{t-d}(i)\right] &\le \mathbb{E}\left[\sum_{i=1}^{k}A_t(i,v)\right] = \mathbb{E}\left[\sum_{i=1}^{k}\sum_{s=1}^{k}\frac{\overline{x}_{t-s}(i,v)}{m}\widehat{\ell}_{t-s}(i,v)\right]\\ &= \mathbb{E}\left[\sum_{i=1}^{k}\sum_{s=1}^{k}\frac{\overline{x}_{t-s}(i,v)}{m}\ell_{t-s-d}(i)\right] \le \mathbb{E}\left[\sum_{i=1}^{k}\sum_{s=1}^{k}\frac{\overline{x}_{t-s}(i,v)}{m}\right] = d\,.\end{aligned}$$

Similarly, for the second sum in (4.11), we have

$$\mathbb{E}\left[\sum_{t=d+1}^{T}\widehat{\ell}_t\left(i^\star,v\right)\right]=\sum_{t=d+1}^{T}\ell_{t-d}(i^\star)\leq\sum_{t=1}^{T}\ell_t(i^\star).$$

Finally for the third sum in (4.11), an iterative application of Lemma 13 and the inequality $\left(1+\frac{1}{d}\right)^k\leq e.$ yields, for $t>d$,

$$\frac{\overline{x}_t(i,v)}{m}\leq\left(1+\frac{1}{d}\right)^k\frac{\overline{x}_{t-d}(i,v)}{m}\leq e\frac{\overline{x}_{t-d}(i,v)}{m}\,,$$

so that we can finally write

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{m}\widehat{\ell}_t(i,v)^2\right]&=\frac{1}{m}\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\mathbb{E}_{t-d}\left[\overline{x}_t(i,v)\widehat{\ell}_t(i,v)^2\right]\right]\\&\leq\frac{1}{m}\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_t(i,v)}{\overline{B}_{d,t-d}(i,v)}\right]\\&\leq\frac{e}{m}\,\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_{t-d}(i,v)}{\overline{B}_{d,t-d}(i,v)}\right]\,.\end{aligned}$$

Therefore, putting everything together and multiplying by m, we have, for any $i^\star\in\{1,\dots,k\}$,

$$\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\ell_t(i)\,\overline{x}_t(i,v)\right]-m\sum_{t=1}^{T}\ell_t(i^\star)\leq 2dm+\eta dmT+\frac{m\ln k}{\eta}+\frac{\eta e}{2}\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_{t-d}(i,v)}{\overline{B}_{d,t-d}(i,v)}\right]\,. \tag{4.12}$$

We will use this estimate to upper bound the regret. Let

$$\begin{aligned}a^\star&\in\operatorname*{argmin}_{a\in\mathcal{A}}\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{j=1}^{k}\ell_t(j)\,a(j)\,\mathbb{I}\{v\in\mathcal{S}_t\}\\i^\star&\in\operatorname*{argmin}_{j\in\{1,\dots,k\}}\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}m\,\ell_t(j)\,q(v)\end{aligned}$$

Then

$$
\begin{aligned}
R_T &= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{i=1}^{k}\ell_t(i)\,x_t(i,v)\,\mathbb{I}\{v\in\mathcal{S}_t\} - \min_{a\in\mathcal{A}}\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{j=1}^{k}\ell_t(j)\,a(j)\,\mathbb{I}\{v\in\mathcal{S}_t\}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{i=1}^{k}\ell_t(i)\,\mathbb{E}_t\big[x_t(i,v)\,\mathbb{I}\{v\in\mathcal{S}_t\}\big]\right] - \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{j=1}^{k}\ell_t(j)\,a^\star(j)\,\mathbb{I}\{v\in\mathcal{S}_t\}\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{i=1}^{k}\ell_t(i)\,\mathbb{E}_t\big[x_t(i,v)\big]\,\mathbb{E}_t\big[\mathbb{I}\{v\in\mathcal{S}_t\}\big]\right] - \sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{j=1}^{k}\ell_t(j)\,a^\star(j)\,\mathbb{E}\big[\mathbb{I}\{v\in\mathcal{S}_t\}\big] \\
&= \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{i=1}^{k}\ell_t(i)\,\overline{x}_t(i,v)\,q(v)\right] - \sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\left(\sum_{j=1}^{k}\ell_t(j)\,a^\star(j)\right)q(v) \\
&\le \mathbb{E}\left[\sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\sum_{i=1}^{k}\ell_t(i)\,\overline{x}_t(i,v)\,q(v)\right] - \sum_{t=1}^{T}\sum_{v\in\mathcal{V}}\big(m\,\ell_t(i^\star)\big)\,q(v) \\
&= \sum_{v\in\mathcal{V}}\left(\mathbb{E}\left[\sum_{t=1}^{T}\sum_{i=1}^{k}\ell_t(i)\,\overline{x}_t(i,v)\right] - m\sum_{t=1}^{T}\ell_t(i^\star)\right)q(v) \\
&\overset{(4.12)}{\le} \sum_{v\in\mathcal{V}}\left(2dm + \eta dmT + \frac{m\ln k}{\eta} + \frac{\eta e}{2}\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\frac{\overline{x}_{t-d}(i,v)}{\overline{B}_{d,t-d}(i,v)}\right]\right)q(v) \\
&= 2dmQ + \eta dmTQ + \frac{m\ln k}{\eta}Q + \frac{\eta e}{2}\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\sum_{v\in\mathcal{V}}\frac{\overline{x}_{t-d}(i,v)\,q(v)}{\overline{B}_{d,t-d}(i,v)}\right]
\end{aligned}
$$

For the last term, applying Lemma 14 to the d-th power of the graph $\mathcal{G}$, we get the following upper bound,

$$
\begin{aligned}
\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\sum_{v\in\mathcal{V}}\frac{\overline{x}_{t-d}(i,v)q(v)}{q_{d,t-d}(i,v)}\right] &\le \frac{1}{(1-e^{-1})}\mathbb{E}\left[\sum_{t=d+1}^{T}\sum_{i=1}^{k}\left(\alpha_d + \sum_{v\in\mathcal{V}}\overline{x}_{t-d}(i,v)q(v)\right)\right] \\
&\le \frac{1}{(1-e^{-1})}T\left(k\alpha_d + mQ\right) \\
&= \frac{1}{(1-e^{-1})}TQ\left(\frac{k}{Q}\alpha_d + m\right).
\end{aligned}
$$

Putting all together

$$
R_T \le 2dmQ + \eta dmTQ + \frac{m\ln k}{\eta}Q + \frac{\eta}{2}\frac{eTQ}{(1-e^{-1})}\left(\frac{k}{Q}\alpha_d + m\right).
$$

□

4.6 Appendix

We recall some notation that we will use extensively in the analysis of OMD. We denote the topological interior of $\mathcal{X}'$ by $\mathrm{int}(\mathcal{X}')$ and define the *Bregman divergence* $\mathcal{B}_F\colon \mathcal{X}'\times\mathrm{int}(\mathcal{X}')\to\mathbb{R}$ with respect to a differentiable function $F\colon \mathrm{int}(\mathcal{X}')\to\mathbb{R}$ as

$$
\mathcal{B}_F(x,y) = F(x) - F(y) - \langle\nabla F(y), x-y\rangle \qquad \forall (x,y)\in\mathcal{X}'\times\mathrm{int}(\mathcal{X}') \tag{4.13}
$$

We recall that for all $z \in \mathbb{R}^k$, the *dual norm* of z (or, equivalently, of the linear functional $\langle z, \cdot \rangle$) is given by

$$\|z\|_* = \max\{\langle z, x \rangle : x \in \mathbb{R}^k, \|x\| \leq 1\} \tag{4.14}$$

We also recall that a differentiable function f is called *σ-strongly convex* with respect to a norm $\|\cdot\|$, if

$$f(x) \geq f(y) + \langle \nabla f(y), x - y \rangle + \frac{\sigma}{2} \|x - y\|^2$$

for all x, y and for some $\sigma > 0$. Note that the previous expression is equivalent to

$$\mathcal{B}_F(x, y) \geq \frac{\sigma}{2} \|x - y\|^2 \tag{4.15}$$

by definition of Bregman divergence.

Finally, let F be a convex function and $\mathcal{X} = \mathrm{dom}(F)$ and $C = \mathrm{int}(\mathcal{X})$. Then F is *Legendre* if

1. C is nonempty;
2. F is differentiable and strictly convex on C
3. $\lim_{n\to\infty} \|\nabla F(x_n)\|_2 = \infty$ for any sequence $(x_n)_n$ with $x_n \in C$ for all n and $\lim_{n\to\infty} x_n = x$ and some $x \in \partial C$

4.6.1 Analysis of Online Mirror Descent with delays

Let us introduce the delayed environment following the work of Joulani et al. [2016]. For all s, we denote by $\rho(s)$ the time step in which the s-th feedback pair, that BASE receives from SOLID (line 8 of Algorithm 4), was generated. Next, for all s, we denote by $\widetilde{\tau}_s$ the number of feedbacks that BASE receives between the time $\rho(s)$ in which SOLID makes the prediction $x_{\rho(s)}$ (i.e., when the learner incurs loss $\ell_{\rho(s)}$) and that in which $\ell_{\rho(s)}$ is received by BASE (i.e., when the learner receives the loss $\ell_{\rho(s)}$ at round $\rho(s) + d_{\rho(s)}$). For all s, we set $\widetilde{\ell}_s = \ell_{\rho(s)}$. In words, $\widetilde{\ell}_1, \widetilde{\ell}_2, \ldots$ is the sequence of losses in the order received by BASE. In the same spirit, for all s we denote the prediction made by BASE after receiving $\widetilde{\ell}_s$ by $\widetilde{x}_{s+1}$. Note that $\widetilde{x}_{s-\widetilde{\tau}_s} = x_{\rho(s)}$. Furthermore, without loss of generality we will assume that for any $1 \leq t \leq T$, $t + d_t \leq T$, i.e., all feedbacks are received by the end of round T. This does not restrict generality because the feedbacks that arrive in round T are not used to make any predictions and hence do not influence the regret of SOLID. Note that under this assumption $\sum_{s=1}^{T} \widetilde{\tau}_s = \sum_{t=1}^{T} d_t$ (both count over time the total number of outstanding feedbacks), and $(\rho(s))_{1 \leq s \leq T}$ is a permutation of the integers $\{1, \ldots, T\}$.

We recall here an important identity that was proven in Joulani et al. [2016].

Theorem 13. *Let BASE be any deterministic algorithm for the non-delyed setting. For every $x \in \mathcal{X}$ and all time horizons T, the regret of SOLID with input BASE satisfies*

$$R_T(x) = \widetilde{R}_T(x) + \sum_{s=1}^{T} \widetilde{D}_{s,\widetilde{\tau}_s} \tag{4.16}$$

where

$$\widetilde{R}_T(x) = \sum_{s=1}^{T} \widetilde{\ell}_s(\widetilde{x}_s) - \sum_{s=1}^{T} \widetilde{\ell}_s(x)$$

is the regret of BASE relative to x for the sequence of losses $\widetilde{\ell}_1, \dots, \widetilde{\ell}_T$ and

$$\widetilde{D}_{s,\widetilde{\tau}_s} = \widetilde{\ell}_s(\widetilde{x}_{s-\widetilde{\tau}_s}) - \widetilde{\ell}_s(\widetilde{x}_s) = \ell_{\rho(s)}\left(x_{\rho(s)}\right) - \widetilde{\ell}_s(\widetilde{x}_s)$$

is the prediction drift of BASE while feedback $\widetilde{\ell}_s$ is outstanding.

In the following we will use Online Mirror Descent (OMD) as BASE and we study the regret of OMD in a delayed environment bounding separately the two contributions coming from the non-delayed regret of BASE and its prediction drift. Let us start with the following lemma that bounds the stability for OMD.

Lemma 15. *If OMD is run with inputs $\mathcal{X}, F, (\eta_t)_{t\in\mathbb{N}}, x_1$ and F is 1-strongly convex with respect to a norm $\|\cdot\|$, then, for all $t = 1, 2, \dots$*

$$\|x_t - x_{t+1}\| \leq \eta_t \|g_t\|_*$$

Proof. Fix any t. Without loss of generality, assume that $\|x_t - x_{t+1}\| > 0$. Recall that by definition of subgradient, a point $x^\star$ is the minimum of a convex function f if and only if $\langle \nabla f(x^\star), x - x^\star \rangle \geq 0$ for all x, where $\nabla f(x^\star)$ is any subgradient of f at $x^\star$. Since for all t, $x_{t+1} = \min_{x\in\mathcal{X}} \{f_t(x)\}$, where $f_t(x) = \langle g_t, x\rangle + \frac{1}{\eta_t}\mathcal{B}_F(x, x_t)$ is convex, we have that

$$\eta_t \langle \nabla f_t(x_{t+1}), x_t - x_{t+1}\rangle = \langle \eta_t g_t + \nabla F(x_{t+1}) - \nabla F(x_t), x_t - x_{t+1}\rangle \geq 0$$

or, equivalently,

$$\langle \eta_t g_t, x_t - x_{t+1}\rangle \geq \langle \nabla F(x_t) - \nabla F(x_{t+1}), x_t - x_{t+1}\rangle \tag{4.17}$$

By the 1-strong convexity of F, we get

$$\begin{aligned}\|x_t - x_{t+1}\|^2 &= \frac{1}{2}\|x_t - x_{t+1}\|^2 + \frac{1}{2}\|x_{t+1} - x_t\|^2 \\ &\leq \mathcal{B}_F(x_t, x_{t+1}) + \mathcal{B}_F(x_{t+1}, x_t) = \langle \nabla F(x_t) - \nabla F(x_{t+1}), x_t - x_{t+1}\rangle\end{aligned}$$

Further upper bounding with Eq. (4.17) and by definition of dual norm, we have

$$\|x_t - x_{t+1}\|^2 \leq \langle \eta_t g_t, x_t - x_{t+1}\rangle \leq \eta_t \|g_t\|_* \|x_t - x_{t+1}\| \ .$$

The result follows by dividing both the left and the right hand side by $\|x_t - x_{t+1}\|$. □

The following known result states an upper bound for the regret of OMD (run in a non-delayed setting).

Theorem 14. *The regret of OMD (Algorithm 5) for the sequence of losses $\widetilde{\ell}_1, \dots, \widetilde{\ell}_T$ with $\{\tilde{\eta}_s\}_{s=1,\dots,T}$ as the set of learning rates, $u \in \mathcal{X}$, F a 1-strongly convex regularizer w.r.t. norm $\|\cdot\|$ is:*

$$\widetilde{R}_T^{BASE}(u) \leq \frac{\max_{1\leq s\leq T} \mathcal{B}_F\left(u, \tilde{x}_s\right)}{\tilde{\eta}_T} + \frac{1}{2}\sum_{s=1}^{T} \tilde{\eta}_s \|\tilde{g}_s\|_*^2 \ .$$

If we assume that $\max_{1\leq s\leq T}\mathcal{B}_F\left(u,\tilde{x}_s\right)\leq 2R^2$ *for a positive constant* $R>0$*, then the regret is*

$$\widetilde{R}_T^{BASE}(u)\leq\frac{2R^2}{\tilde{\eta}_T}+\frac{1}{2}\sum_{s=1}^{T}\tilde{\eta}_s\left\|\tilde{g}_s\right\|_*^2\,.$$

From Lemma 15 the stability of OMD is bounded by $\left\|\tilde{x}_j-\tilde{x}_{j+1}\right\|\leq\tilde{\eta}_j\left\|\tilde{g}_j\right\|_*$, and the drift term for linear losses is

$$\begin{aligned}\tilde{D}_{s,\tilde{\tau}_s}&=\tilde{\ell}_s\left(\tilde{x}_{s-\tilde{\tau}_s}\right)-\tilde{\ell}_s\left(\tilde{x}_s\right)=\sum_{j=s-\widetilde{\tau}_s}^{s-1}\tilde{\ell}_s\left(\tilde{x}_j\right)-\tilde{\ell}_s\left(\tilde{x}_{j+1}\right)\\&\leq\sum_{j=s-\widetilde{\tau}_s}^{s-1}\left\langle\nabla\tilde{\ell}_s\left(\tilde{x}_j\right),\tilde{x}_j-\tilde{x}_{j+1}\right\rangle\\&\leq\sum_{j=s-\tilde{\tau}_s}^{s-1}\left\|\nabla\tilde{\ell}_s\left(\tilde{x}_j\right)\right\|_*\left\|\tilde{x}_j-\tilde{x}_{j+1}\right\|\end{aligned}$$

and now if losses are linear, from $\tilde{\ell}_s\left(\tilde{x}_j\right)=\left\langle\tilde{\ell}_s,\tilde{x}_j\right\rangle$ we get $\nabla\tilde{\ell}_s\left(\tilde{x}_j\right)=\tilde{\ell}_s$, and the drift term becomes

$$\tilde{D}_{s,\tilde{\tau}_s}\leq\sum_{j=s-\tilde{\tau}_s}^{s-1}\left\|\tilde{\ell}_s\right\|_*\left\|\tilde{x}_j-\tilde{x}_{j+1}\right\|\leq\sum_{j=s-\tilde{\tau}_s}^{s-1}\tilde{\eta}_j\left\|\tilde{\ell}_s\right\|_*\left\|\tilde{\ell}_j\right\|_*.$$

We need now a technical lemma to prove a regret bound for OMD in the delayed feedback environment.

Lemma 16. *For all* $j,s\in\{1,\ldots,T\}$*, let*

$$\widehat{G}_j^{fwd}=1+2\sum_{s=j+1}^{T}\mathbb{I}\left\{s-\tilde{\tau}_s\leq j\right\}\qquad\text{and}\qquad\widehat{G}_s^{bck}=1+2\tilde{\tau}_s\,,$$

with the understanding that $\widehat{G}_T^{fwd}=1$*. For all* $t\in\{1,\ldots,T\}$*, let* $\widehat{G}_{1:t}^{fwd}=\sum_{j=1}^{t}\widehat{G}_s^{fwd}$*,* $\widehat{G}_{1:t}^{bck}=\sum_{s=1}^{t}\widehat{G}_s^{bck}$ *and* $d\equiv\max_{s=1,\ldots,T}\{d_s\}$*. Then, for all* $t\in\{1,\ldots,T\}$*,*

$$\widehat{G}_{1:t}^{bck}\leq\widehat{G}_{1:t}^{fwd}\leq\widehat{G}_{1:t}^{bck}+d(2d-1)$$

and $\widehat{G}_{1:T}^{bck}=\widehat{G}_{1:T}^{fwd}$*.*

Proof. From the definitions, for all $t \in \{1, \dots, T\}$,

$$\begin{aligned}
\widehat{G}^{bck}_{1:t} &= \sum_{s=1}^{t} \widehat{G}^{bck}_{s} = \sum_{s=1}^{t} (1 + 2\tilde{\tau}_s) = \sum_{s=1}^{t} 1 + 2 \sum_{s=1}^{t} \sum_{j=s-\tilde{\tau}_s}^{s-1} 1 \\
&\leq \sum_{j=1}^{t} 1 + 2 \sum_{j=1}^{t} \sum_{s=j+1}^{t} \mathbb{I}\{s - \tilde{\tau}_s \leq j\} \\
&= \sum_{j=1}^{t} 1 + 2 \sum_{j=1}^{t} \sum_{s=j+1}^{T} \mathbb{I}\{s - \tilde{\tau}_s \leq j\} \\
&\quad - 2 \sum_{j=1}^{t} \sum_{s=t+1}^{T} \mathbb{I}\{s - \tilde{\tau}_s \leq j\} \\
&= \sum_{j=1}^{t} \widehat{G}^{fwd}_{j} - 2 \sum_{j=1}^{t} \sum_{s=t+1}^{T} \mathbb{I}\{s - \tilde{\tau}_s \leq j\} \\
&\leq \widehat{G}^{fwd}_{1:t} ,
\end{aligned}$$

furthermore for $t = T$ we have $-2\sum_{j=1}^{t}\sum_{s=t+1}^{T} \mathbb{I}\{s - \tilde{\tau}_s \leq j\} = 0$ and therefore we have $\widehat{G}^{bck}_{1:T} = \widehat{G}^{fwd}_{1:T}$. We want to lower bound the negative term now to conclude the proof. Let us define $\tau^* = \max_{s=1,\dots,T}\{\tilde{\tau}_s\}$. We notice that for $s > t$ and $j \leq s - \tau^*$ the indicator function $\mathbb{I}\{s - \tilde{\tau}_s \leq j\}$ is equal to zero. Also note that $\mathbb{I}\{s - \tilde{\tau}_s \leq j\} = 0$ for $s > j + \tau^*$. Hence

$$\begin{aligned}
\sum_{j=1}^{t} \sum_{s=t+1}^{T} \mathbb{I}\{s - \tilde{\tau}_s \leq j\} &= \sum_{j=t-\tau^*+1}^{t} \sum_{s=t+1}^{j+\tau^*} \mathbb{I}\{s - \tilde{\tau}_s \leq j\} \\
&\leq \sum_{j=t-\tau^*+1}^{t} (j + \tau^* - t) \\
&= \sum_{i=1}^{\tau^*} i = \frac{1}{2}\tau^*(\tau^* + 1).
\end{aligned}$$

If the maximum delay is $d = \max_{s=1,\dots,T}\{d_s\}$, from the definition of $\tilde{\tau}_s$ we have that $\tau^* \leq 2d - 1$. We conclude that

$$\widehat{G}^{fwd}_{1:t} \leq \widehat{G}^{bck}_{1:t} + \tau^*(\tau^* + 1) \leq \widehat{G}^{bck}_{1:t} + d(2d - 1) .$$

□

Lemma 17 (see McMahan and Streeter [2014], Lemma 9)**.** *For any sequence of real numbers $x_1, x_2, \dots, x_n$ such that $x_{1:t} = \sum_{s=1}^{t} x_s > 0$ for all $t = 1, 2, \dots, n$, we have*

$$\sum_{t=1}^{n} \frac{x_t}{\sqrt{x_{1:t}}} \leq 2\sqrt{x_{1:n}}.$$

We have the following theorem for the regret of OMD in the delayed environment.

Theorem 7. *Suppose losses are linear and we run SOLID in a delayed-environment. Let $\tilde{\eta}_j$ denote the learning rates that BASE uses in its simulated non-delayed run inside SOLID environment. If $\alpha = \sqrt{2}\frac{R}{L}$ and $\tilde{\eta}_j = \alpha / \sqrt{\widehat{G}^{bck}_{1:j} + d(2d-1)}$ then the regret of SOLID with OMD can be bounded as*

$$R_T \leq 2LR\sqrt{2\sum_{t=1}^{T}(1+2d_t)} + LR\sqrt{2d(2d-1)}.$$

Proof. The total regret is bounded in the following way, where from L-Lipschitzness of the losses we have $\left\|\tilde{\ell}_j\right\|_* \leq L$ for each $j = 1, \ldots, T$:

$$\begin{aligned}
R_T &\leq \frac{2R^2}{\tilde{\eta}_T} + \frac{1}{2}\sum_{s=1}^{T}\tilde{\eta}_s\left\|\tilde{\ell}_s\right\|_*^2 + \sum_{s=1}^{T}\sum_{j=s-\tilde{\tau}_s}^{s-1}\tilde{\eta}_j\left\|\tilde{\ell}_s\right\|_*\left\|\tilde{\ell}_j\right\|_* \\
&\leq \frac{2R^2}{\tilde{\eta}_T} + \frac{L^2}{2}\left(\sum_{s=1}^{T}\tilde{\eta}_s + 2\sum_{s=1}^{T}\sum_{j=s-\tilde{\tau}_s}^{s-1}\tilde{\eta}_j\right) \\
&= \frac{2R^2}{\tilde{\eta}_T} + \frac{L^2}{2}\left(\sum_{j=1}^{T}\tilde{\eta}_j + 2\sum_{j=1}^{T}\tilde{\eta}_j\sum_{s=j+1}^{T}\mathbb{I}\left\{s-\tilde{\tau}_s \leq j\right\}\right) \\
&= \frac{2R^2}{\tilde{\eta}_T} + \frac{L^2}{2}\left(\sum_{j=1}^{T}\tilde{\eta}_j\left(1 + 2\sum_{s=j+1}^{T}\mathbb{I}\left\{s-\tilde{\tau}_s \leq j\right\}\right)\right) \\
&= \frac{2R^2}{\tilde{\eta}_T} + \frac{L^2}{2}\left(\sum_{j=1}^{T}\tilde{\eta}_j\widehat{G}^{fwd}_j\right).
\end{aligned}$$

Let us define

$$\tilde{\eta}_j = \frac{\alpha}{\sqrt{\widehat{G}^{bck}_{1:j} + d(2d-1)}} = \frac{\alpha}{\sqrt{\sum_{i=1}^{j}(1+2\tilde{\tau}_i) + d(2d-1)}} \tag{4.18}$$

Then

$$\begin{aligned}
\frac{2R^2}{\tilde{\eta}_T} &= 2R^2\frac{\sqrt{\sum_{s=1}^{T}(1+2\tilde{\tau}_s) + d(2d-1)}}{\alpha} \\
&\leq 2R^2\frac{\sqrt{\sum_{s=1}^{T}(1+2\tilde{\tau}_s)} + \sqrt{d(2d-1)}}{\alpha} \\
&= 2R^2\frac{\sqrt{\sum_{t=1}^{T}(1+2d_t)} + \sqrt{d(2d-1)}}{\alpha},
\end{aligned}$$

where the last equality follows thanks to the identity $\sum_{s=1}^{T}\tilde{\tau}_s = \sum_{t=1}^{T} d_t$ and we conclude that

$$\begin{aligned}
\sum_{j=1}^{T}\tilde{\eta}_j\widehat{G}^{fwd}_j &= \alpha\sum_{j=1}^{T}\frac{\widehat{G}^{fwd}_j}{\sqrt{\widehat{G}^{bck}_{1:j} + d(2d-1)}} \leq \alpha\sum_{j=1}^{T}\frac{\widehat{G}^{fwd}_j}{\sqrt{\widehat{G}^{fwd}_{1:j}}} \\
&\leq 2\alpha\sqrt{\widehat{G}^{fwd}_{1:T}} = 2\alpha\sqrt{\widehat{G}^{bck}_{1:T}} = 2\alpha\sqrt{\sum_{t=1}^{T}(1+2d_t)},
\end{aligned}$$

where in the second inequality we use Lemma 17. Choosing $\alpha = \sqrt{2}\frac{R}{L}$ the upper bound on the regret becomes

$$R_T \leq 2LR\sqrt{2\sum_{t=1}^{T}(1+2d_t)} + LR\sqrt{2d(2d-1)}.$$

□

4.6.2 Analysis of Hedge with delays

In order to simplify the reasoning of Zimmert and Seldin [2019] to adapt it to the study of Hedge with delays and in a second moment to its cooperative version on the communication network we remind the following definitions taken exactly like in Zimmert and Seldin [2019]. For a convex function F we use F^* to denote its convex conjugate and $\overline{F}^*$ the constrained convex conjugate. They are defined as

$$F^*(y) = \max_{x\in\mathbb{R}^k} \langle x, y\rangle - F(x)\ ,$$
$$\overline{F}^*(y) = \max_{x\in\mathcal{X}} \langle x, y\rangle - F(x)\ ,$$

where here in Appendix 4.6.2 we consider the case of $\mathcal{X} = \Delta^{k-1}$ and the negative entropy regularizer

$$\underbrace{F_t(x)}_{=\sum_{i=1}^k f_t(x_i)} = \underbrace{\eta_t^{-1}\sum_{i=1}^{k} x_i\log(x_i)}_{F_t(x)=\sum_{i=1}^k f_t(x_i)}\ . \tag{4.19}$$

Standard properties of FTRL analysis

We introduce properties of FTRL that will be useful in the following.

Claim 15. *$f_t''(x) : \mathbb{R}_+ \to \mathbb{R}_+$ are monotonically decreasing functions and $f_t^{*\prime} : \mathbb{R} \to \mathbb{R}_+$ are convex and monotonically increasing.*

Proof. By definition $f_t''(x) = \eta_t^{-1}x^{-1}$, which concludes the first statement. Since f_t are Legendre functions, we have $f_t^{*\prime}(f_t'(x)) = x$ and, taking derivatives on both sides and applying the chain rule, we get the identity $f_t^{*\prime\prime}(f_t'(x))\, f_t''(x) = 1$. We set $y = f_t'(x)$, with inverse $f_t^{*\prime}(y) = x$. Therefore, substituting in the previous identity and inverting thanks to monotonicity, we obtain

$$f^{*\prime\prime}_t(y) = f_t''(f_t^{*\prime}(y))^{-1} > 0\ . \tag{4.20}$$

Therefore the function is monotonically increasing. Since both $f_t''(x)^{-1}$, as well as $f_t^{*\prime}(y)$ are increasing, the composition is as well and $f_t^{*\prime\prime\prime} \geq 0$ and this implies convexity of $f_t^{*\prime}$. □

Claim 16. *For any convex F, $L \in \mathbb{R}^k$ and $c \in \mathbb{R}$:*

$$\overline{F}^*(L + c\,\vec{1}) = \overline{F}^*(L) + c\,.$$

Proof. By definition $\overline{F}^*(L + c\vec{1}) = \max_{x \in \Delta^{k-1}} \left\langle x, L + c\vec{1} \right\rangle - F(x) = \max_{x \in \Delta^{k-1}} \langle x, L \rangle - F(x) + c = \overline{F}^*(L) + c$. □

Claim 17. *For any x_t there exists $\lambda \in \mathbb{R}$ such that:*

$$x_t = \nabla \overline{F}_t^*(-\widehat{L}_t^{obs}) = \nabla F_t^*(-\widehat{L}_t^{obs} + \lambda \vec{1}) = \nabla F_t^*(\nabla F_t(x_t))\,.$$

Proof. By the KKT conditions, there exists $\lambda \in \mathbb{R}$ such that

$$x_t = \operatorname*{argmax}_{x \in \Delta^{k-1}} \left\{ \left\langle x, -\widehat{L}_t^{obs} \right\rangle + F_t(x) \right\} = \operatorname*{argmax}_{x \in \mathrm{dom}(F_t)} \left\{ \left\langle x, -\widehat{L}_t^{obs} \right\rangle + F_t(x) + \lambda \left(\sum_{i=1}^{k} x_i - 1 \right) \right\}$$

satisfies $\nabla F_t(x_t) = -\widehat{L}_t^{obs} + \lambda \vec{1}$. The rest follows from the standard result of $\nabla F = (\nabla F^*)^{-1}$ for Legendre F. □

Claim 18. *For any Legrendre function F and $L \in \mathbb{R}^k$ it holds that*

$$\overline{F}^*(L) \leq F^*(L)\,,$$

with equality iff there exists $x \in \Delta^{k-1}$ such that $L = \nabla F(x)$.

Proof. The first statement follows from the definition since for any $A \subset B$: $\max_{x \in A} f(x) \leq \max_{x \in B} f(x)$. The second part follows because $L = \nabla F(x)$ for some $x \in \Delta^{k-1}$ holds if and only if $\nabla F^*(L) = x$ which is equivalent to $\operatorname{argmax}_{x' \in \mathbb{R}^k} \langle x', L \rangle - F(x') = \nabla F^*(L) = x \in \Delta^{k-1}$. Therefore, if the unrestricted maximum x is on the simplex then the maximum restricted to the simplex will be also at the same point x. This statement is equivalent to $\overline{F}^*(L) = F^*(L)$ from the properties of Legendre functions. □

Claim 19. *For any $x \in \Delta^{k-1}$, $L \in [0, \infty)^k$ and $i \in [k]$:*

$$\nabla \overline{F}_t^*(\nabla F_t(x) - L)_i \geq \nabla F_t^*(\nabla F_t(x) - L)_i\,.$$

Proof. As in the proof of Claim 17, there exists $\lambda \in \mathbb{R}$: $\nabla \overline{F}_t^*(\nabla F_t(x) - L) = \nabla F_t^*(\nabla F_t(x) - L + \lambda \vec{1})$. The statement is equivalent to λ being non-negative, since ${f_t^*}'$ are monotonically increasing. If $\lambda < 0$, then observing that $\nabla \overline{F}_t^*(y) \in \Delta^{k-1}$ for all y we have

$$\begin{aligned} 1 &= \sum_{i=1}^{k} \left(\nabla \overline{F}_t^*(\nabla F_t(x) - L) \right)_i = \sum_{i=1}^{k} (\nabla F_t^*(\nabla F_t(x) - L + \lambda \vec{1}))_i \\ &= \sum_{i=1}^{k} {f_t^*}'(f_t'(x_i) - L_i + \lambda) < \sum_{i=1}^{k} {f_t^*}'(f_t'(x_i)) = \sum_{i=1}^{k} x_i = 1\,, \end{aligned}$$

which is a contradiction and completes the proof. □

Claim 20. *For any Legendre function f with monotonically decreasing second derivative, $x \in \mathrm{dom}(f)$ and $\ell \in [0, \infty)$ such that $f'(x) - \ell \in \mathrm{dom}(f^*)$:*

$$\mathcal{B}_{f^*}(f'(x) - \ell, f'(x)) \leq \frac{\ell^2}{2f''(x)}\,.$$

Proof. Based on Taylor's theorem, there exists an $\tilde{x} \in \left[f^{*\prime}\left(f'(x)-\ell\right), x\right]$, such that

$$\mathcal{B}_{f^*}\left(f'(x)-\ell, f'(x)\right) = \frac{\ell^2}{2f''(\tilde{x})}.$$

$\tilde{x}$ is smaller than x, since $f^{*\prime}$ is monotonically increasing. Finally using the fact that the second derivative is decreasing allows us to bound $f''\left(\tilde{x}\right)^{-1} \leq f''\left(x\right)^{-1}$. □

Proof of Theorem 8

Let us define the cumulative loss vector L_t as follows, for every arm i we have

$$L_{t,i} = \sum_{s=1}^{t-1} \ell_{s,i}.$$

The arm with the best cumulative loss in hindsight is $i^* = \text{argmin}_{i\in[k]} \sum_{t=1}^{T} \ell_{t,i}$.

Lemma 18. *For any t it holds*

$$\overline{F}_t^*(-L_t^{obs} - \ell_t) - \overline{F}_t^*(-L_t^{obs}) + \langle x_t, \ell_t\rangle \leq \frac{\eta_t}{2}\,.$$

Proof. We have

$$\begin{aligned}
&\overline{F}_t^*(-L_t^{obs} - \ell_t) - \overline{F}_t^*(-L_t^{obs}) + \langle x_t, \ell_t\rangle \\
&= \overline{F}_t^*(\nabla F_t(x_t) - \ell_t + \lambda\vec{1}) - \overline{F}_t^*(\nabla F_t(x_t) + \lambda\vec{1}) + \langle x_t, \ell_t\rangle \\
&= \overline{F}_t^*(\nabla F_t(x_t) - \ell_t) - \overline{F}_t^*(\nabla F_t(x_t)) + \langle x_t, \ell_t\rangle \\
&\leq F_t^*(\nabla F_t(x_t) - \ell_t) - F_t^*(\nabla F_t(x_t)) + \langle x_t, \ell_t\rangle \\
&= \sum_{i=1}^{k} \mathcal{B}_{f_t^*}(f_t'(x_{t,i}) - \ell_{t,i}, f_t'(x_{t,i})) \\
&\leq \frac{1}{2}\sum_{i=1}^{k} \ell_{t,i}^2\, f_t''(x_{t,i})^{-1} \\
&= \frac{\eta_t}{2}\sum_{i=1}^{k} \ell_{t,i}^2\, x_{t,i} \\
&\leq \frac{\eta_t}{2}
\end{aligned}$$

where the first equality follows using Claim 17, the second equality is from Claim 16, the first inequality is from both parts of Claim 18, the second inequality follows from Claim 20 and finally the last equality is from the expression of $f_t''(x) = 1/x$ that holds for the negative etropy regularizer. □

Lemma 19. *For any non-increasing learning rate η_t, it holds that*

$$\sum_{t=1}^{T}\left(\overline{F}_t^*(-L_t) - \overline{F}_t^*(-L_{t+1}) - \langle \mathbf{e}_{i^*}, \ell_t\rangle\right) \leq \frac{\log(k)}{\eta_T}.$$

Proof. Let $\tilde{x}_t = \text{argmax}_{x\in\Delta^{k-1}} \{\langle x, -L_t\rangle - F_t(x)\}$, then

$$\overline{F}_t^*(-L_t) = \langle \tilde{x}_t, -L_t\rangle - F_t(\tilde{x}_t).$$

Furthermore, since $\overline{F}^*(-L_t) = \max_{x\in\Delta^{k-1}} \{\langle x, -L_t\rangle - F(x)\}$, we have

$$-\overline{F}_{t-1}^*(-L_t) \le \langle \tilde{x}_t, L_t\rangle + F_{t-1}(\tilde{x}_t) \tag{4.21}$$

$$-\overline{F}_T^*(-L_{T+1}) \le \langle \mathbf{e}_{i^*}, L_{T+1}\rangle + F_T(\mathbf{e}_{i^*}) = \sum_{t=1}^T \langle \mathbf{e}_{i^*}, \ell_t\rangle \ . \tag{4.22}$$

Plugging these inequalities into the LHS leads to

$$\begin{aligned}
&\sum_{t=1}^T \left(\overline{F}_t^*(-L_t) - \overline{F}_t^*(-L_{t+1}) - \langle \mathbf{e}_{i^*}, \ell_t\rangle \right) \\
&\qquad = \sum_{t=1}^T \overline{F}_t^*(-L_t) - \sum_{t=2}^T \overline{F}_{t-1}^*(-L_t) - \overline{F}_T^*(-L_{T+1}) - \sum_{t=1}^T \langle \mathbf{e}_{i^*}, \ell_t\rangle \\
&\qquad \le \sum_{t=1}^T \overline{F}_t^*(-L_t) - \sum_{t=2}^T \overline{F}_{t-1}^*(-L_t) \\
&\qquad \le \sum_{t=1}^T \langle \tilde{x}_t, -L_t\rangle - \sum_{t=1}^T F_t(\tilde{x}_t) + \sum_{t=2}^T \langle \tilde{x}_t, L_t\rangle + \sum_{t=2}^T F_{t-1}(\tilde{x}_t) \\
&\qquad = -F_1(\tilde{x}_1) + \sum_{t=2}^T F_{t-1}(\tilde{x}_t) - F_t(\tilde{x}_t) \\
&\qquad \le \sum_{t=1}^T F_{t-1}(\tilde{x}_t) - F_t(\tilde{x}_t) \\
&\qquad = \sum_{t=1}^T \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) (-F(\tilde{x}_t)) \\
&\qquad \le \sum_{t=1}^T \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) \max_{x\in\Delta^{k-1}} (-F(x)) \\
&\qquad = \frac{1}{\eta_T} \max_{x\in\Delta^{k-1}} \{-F(x)\} \ ,
\end{aligned}$$

where in the first inequality we used Eq. (4.22), in the second inequality we used Eq. (4.21), and with a slight abuse of notation we defined $\eta_0^{-1} = 0$ in the third inequality. We are left with computing the maximum, and using Jensen's inequality

$$\begin{aligned}
\max_{x\in\Delta^{k-1}} \{-F(x)\} &= \max_{x\in\Delta^{k-1}} \left\{ \sum_{i=1}^k x_i \log\left(\frac{1}{x_i}\right) \right\} \\
&\le \log(k) \, .
\end{aligned}$$

□

Lemma 20. *For any t it holds that*

$$\overline{F}_t^*(-L_t^{obs}) - \overline{F}_t^*(-L_t^{obs} - \ell_t) - \overline{F}_t^*(-L_t) + \overline{F}_t^*(-L_{t+1}) \le \eta_t \mathfrak{d}_t.$$

Proof. We define $L_t^{miss} = L_t - L_t^{obs}$. Then we have

$$\begin{aligned}-\overline{F}_t^*(-L_t) + \overline{F}_t^*(-L_{t+1}) &= -\int_0^1 \left\langle \ell_t, \nabla \overline{F}_t^*(-L_t - x\ell_t) \right\rangle dx \\ &= -\int_0^1 \left\langle \ell_t, \nabla \overline{F}_t^*(-L_t^{obs} - L_t^{miss} - x\ell_t) \right\rangle dx\end{aligned}$$

where the first equality uses the fundamental theorem of calculus. In the same way we have

$$\overline{F}_t^*(-L_t^{obs}) - \overline{F}_t^*(-L_t^{obs} - \ell_t) = \int_0^1 \left\langle \ell_t, \nabla \overline{F}_t^*(-L_t^{obs} - x\ell_t) \right\rangle dx$$

Now, putting the previous equations together and defining $\tilde{z}(x) = \nabla \overline{F}_t^*(-L_t^{obs} - x\ell_t)$ we have the following

$$\begin{aligned}&\overline{F}_t^*(-L_t^{obs}) - \overline{F}_t^*(-L_t^{obs} - \ell_t) - \overline{F}_t^*(-L_t) + \overline{F}_t^*(-L_{t+1}) \\ &\quad= \int_0^1 \left\langle \ell_t, \nabla \overline{F}_t^*(-L_t^{obs} - x\ell_t) \right\rangle dx - \int_0^1 \left\langle \ell_t, \nabla \overline{F}_t^*(-L_t^{obs} - L_t^{miss} - x\ell_t) \right\rangle dx \\ &\quad= \int_0^1 \left\langle \ell_t, \tilde{z}(x) - \nabla \overline{F}_t^*\left(\nabla F_t(\tilde{z}(x)) - L_t^{miss}\right) \right\rangle dx \\ &\quad\leq \int_0^1 \left\langle \ell_t, \tilde{z}(x) - \nabla F_t^*\left(\nabla F_t(\tilde{z}(x)) - L_t^{miss}\right) \right\rangle dx \\ &\quad= \sum_{i=1}^k \int_0^1 \ell_{t,i} \left(\tilde{z}(x) - \nabla F_t^*\left(\nabla F_t(\tilde{z}(x)) - L_t^{miss}\right)\right)_i dx \\ &\quad= \sum_{i=1}^k \int_0^1 \ell_{t,i} \left(\tilde{z}_i(x) - f_t^{*\prime}\left(f_t'(\tilde{z}_i(x)) - L_{t,i}^{miss}\right)\right) dx \\ &\quad\leq \sum_{i=1}^k \int_0^1 \ell_{t,i} f_t^{*\prime\prime}\left(f_t'(\tilde{z}_i(x))\right) L_{t,i}^{miss}\, dx \\ &\quad= \sum_{i=1}^k \int_0^1 \ell_{t,i} \left(f_t''(\tilde{z}_i(x))\right)^{-1} L_{t,i}^{miss}\, dx \\ &\quad= \eta_t \int_0^1 \left(\sum_{i=1}^k \ell_{t,i}\, L_{t,i}^{miss}\, \tilde{z}_i(x)\right) dx\,,\end{aligned}$$

where the second equality uses the definition of $\tilde{z}(x)$ and Claim 17, the first inequality applies Claim 19, the second inequality follows because $f_t^{*\prime}$ is convex, so $-f_t^{*\prime}\left(f_t{}'(\tilde{z}_i) - \ell\right) \leq -\tilde{z}_i + f_t^{*\prime\prime}\left(f_t{}'(\tilde{z}_i)\right)\ell$, the second to last equlity follows by Eq. (4.20), and the last follows

because $f_t''(x) = \frac{1}{x}$. From the previous series of inequalities we have

$$\begin{aligned}
&\overline{F}_t^*(-L_t^{obs}) - \overline{F}_t^*(-L_t^{obs} - \ell_t) - \overline{F}_t^*(-L_t) + \overline{F}_t^*(-L_{t+1}) \\
&\qquad \leq \eta_t \int_0^1 \left(\sum_{i=1}^k \ell_{t,i} L_{t,i}^{miss} \tilde{z}_i(x) \right) dx \\
&\qquad = \eta_t \int_0^1 \left(\sum_{i=1}^k \ell_{t,i} \left(\sum_{s:s<t} \mathbb{I}\left\{ s + d_s \geq t \right\} \ell_{s,i} \right) \tilde{z}_i(x) \right) dx \\
&\qquad \leq \eta_t \int_0^1 \left(\sum_{i=1}^k \left(\sum_{s:s<t} \mathbb{I}\left\{ s + d_s \geq t \right\} \right) \tilde{z}_i(x) \right) dx \\
&\qquad = \eta_t \int_0^1 \sum_{s:s<t} \mathbb{I}\left\{ s + d_s \geq t \right\} \sum_{i=1}^k \tilde{z}_i(x)\, dx \\
&\qquad = \eta_t \sum_{s:s<t} \mathbb{I}\left\{ s + d_s \geq t \right\} \\
&\qquad = \eta_t \mathfrak{d}_t \,,
\end{aligned}$$

where the first equality follows by expanding the definition of $L_{t,i}^{miss} = \sum_{s:s<t} \mathbb{I}\left\{ s + d_s \geq t \right\} \ell_{s,i}$, the second inequality bounds losses with one, finally the second to last equality follows by the fact that by definition $\tilde{z}(x) \in \Delta^{k-1}$. □

Theorem 8. *Algorithm 6 with decreasing learning rates $(\eta_t)_{t=1,\dots,n}$ satisfies*

$$R_T \leq \frac{\log(k)}{\eta_T} + \frac{1}{2} \sum_{t=1}^T \eta_t + \sum_{t=1}^T \eta_t \mathfrak{d}_t \,.$$

Furthermore if one chooses $\eta_t = \sqrt{\frac{\log k}{\sum_{s=1}^t 1 + 2\mathfrak{d}_s}}$ then

$$R_T \leq 2\sqrt{\log(k) \left(T + \sum_{t=1}^T \mathfrak{d}_t \right)}.$$

Proof. We have the following decomposition of R_T

$$\begin{aligned}
R_T &= \sum_{t=1}^T \langle x_t - \mathbf{e}_{i^*}, \ell_t \rangle = \sum_{t=1}^T \langle x_t, \ell_t \rangle - \langle \mathbf{e}_{i^*}, \ell_t \rangle \\
&= \sum_{t=1}^T \left(\overline{F}_t^*(-L_t) - \overline{F}_t^*(-L_{t+1}) - \langle \mathbf{e}_{i^*}, \ell_t \rangle \right) \\
&\quad + \sum_{t=1}^T \left(\overline{F}_t^*(-L_t^{obs} - \ell_t) - \overline{F}_t^*(-L_t^{obs}) + \langle x_t, \ell_t \rangle \right) \\
&\quad + \sum_{t=1}^T \left(\overline{F}_t^*(-L_t^{obs}) - \overline{F}_t^*(-L_t^{obs} - \ell_t) - \overline{F}_t^*(-L_t) + \overline{F}_t^*(-L_{t+1}) \right) \\
&\leq \frac{\log(k)}{\eta_T} + \frac{1}{2} \sum_{t=1}^T \eta_t + \sum_{t=1}^T \eta_t \mathfrak{d}_t \,.
\end{aligned}$$

□

4.6.3 Analysis of partial information settings

Standard properties of FTRL analysis for semi-bandits

In this section we present an adaptation of the reasoning of Zimmert and Seldin [2019] to the study of semibandits with delays.

We use the following hybrid regularizer $F_t = F_{t,1} + F_{t,2}$, where each of the two parts of the regularizer has its own learning rate.

$$\underbrace{F_t(x)}_{=\sum_{i=1}^k f_t(x_i)} = \underbrace{-\sum_{i=1}^k 2\sqrt{t}x_i^{1/2}}_{F_{t,1}(x)=\sum_{i=1}^k f_{t,1}(x_i)} + \underbrace{\eta_t^{-1}\sum_{i=1}^k x_i \log(x_i)}_{F_{t,2}(x)=\sum_{i=1}^k f_{t,2}(x_i)} . \tag{4.23}$$

The first part of the regularizer $F_{t,1}(x) = \sqrt{t}F_1(x)$ is the $\frac{1}{2}$-Tsallis entropy $F_1(x) = -2\sum_{i=1}^k \sqrt{x_i}$ with learning rate $\frac{1}{\sqrt{t}}$, which is non-adaptive to the problem. The second part of the regularizer $F_{t,2}(x) = \eta_t^{-1}F_2(x)$ is the negative entropy $F_2(x) = \sum_{i=1}^k x_i \log(x_i)$ with adaptive learning rate η_t. We define this regularizer on the domain $\mathrm{co}(\mathcal{A}) = \left\{x \in [0,1]^k : \sum_{i=1}^k x_i = m\right\}$ which corresponds to the probability simplex just in the case of $m = 1$.

Claim 21. *$f_t''(x) : \mathbb{R}_+ \to \mathbb{R}_+$ are monotonically decreasing functions and $f_t^{*\prime} : \mathbb{R} \to \mathbb{R}_+$ are convex and monotonically increasing. An analogous result holds for the function $f_{t,2}$.*

Proof. By definition $f_t''(x) = \sqrt{t}x^{-3/2} + \eta_t^{-1}x^{-1}$, which concludes the first statement. Since f_t are Legendre functions, we have $f_t^{*\prime\prime}(y) = f_t''(f_t^{*\prime}(y))^{-1} > 0$. Therefore the function is monotonically increasing. Since both $f_t''(x)^{-1}$, as well as $f_t^{*\prime}(y)$ are increasing, the composition is as well and $f_t^{*\prime\prime\prime} > 0$. The result for $f_{t,2}$ follows immediately from its definition in the same way as for f_t. □

Claim 22. *For any convex F, $L \in \mathbb{R}^k$ and $c \in \mathbb{R}$:*

$$\overline{F}^*(L + c\,\vec{1}) = \overline{F}^*(L) + mc\,.$$

Proof. By definition $\overline{F}^*(L + c\vec{1}) = \max_{x\in\mathrm{co}(\mathcal{A})} \left\langle x, L + c\vec{1}\right\rangle - F(x) = \max_{x\in\mathrm{co}(\mathcal{A})} \langle x, L\rangle - F(x) + mc = \overline{F}^*(L) + mc$. □

Claim 23. *For any x_t there exists $\lambda \in \mathbb{R}$ such that:*

$$x_t = \nabla\overline{F}_t^*(-\widehat{L}_t^{obs}) = \nabla F_t^*(-\widehat{L}_t^{obs} + \lambda\vec{1}) = \nabla F_t^*(\nabla F_t(x_t))\,.$$

An analogous result also holds if we use as a regularizer $F_{t,2}$ in place of the hybrid F_t.

Proof. By the KKT conditions, there exists $\lambda \in \mathbb{R}$ such that $x_t = \mathrm{argmax}_{x\in\mathrm{co}(\mathcal{A})} \left\langle x, -\widehat{L}_t^{obs}\right\rangle + F_t(x)$ satisfies $\nabla F_t(x_t) = -\widehat{L}_t^{obs} + \lambda\vec{1}$. The rest follows from the standard result of $\nabla F = (\nabla F^*)^{-1}$ for Legendre F. □

Claim 24. *For any Legrendre function F and $L \in \mathbb{R}^k$ it holds that*

$$\overline{F}^*(L) \leq F^*(L)\,,$$

with equality iff there exists $x \in co(\mathcal{A})$ such that $L = \nabla F(x)$.

Proof. The first statement follows from the definition since for any $A \subset B$: $\max_{x\in A} f(x) \leq \max_{x\in B} f(x)$. The second part follows because equality means that $\operatorname{argmax}_x \langle x, L\rangle - F(x) = \nabla F^*(L) \in \mathrm{co}(\mathcal{A})$, which is equivalent to the statement. □

Claim 25. *For any $x \in co\,(\mathcal{A})$, $L \in [0,\infty)^k$ and $i \in [k]$:*

$$\nabla \overline{F}_t^*(\nabla F_t(x) - L)_i \geq \nabla F_t^*(\nabla F_t(x) - L)_i\,.$$

Proof. By Claim 23, there exists $\lambda \in \mathbb{R}:\ \nabla \overline{F}_t^*(\nabla F_t(x) - L) = \nabla F_t^*(\nabla F_t(x) - L + \lambda\vec{1})$. The statement is equivalent to λ being non-negative, since ${f_t^*}'$ are monotonically increasing. If $\lambda < 0$, then

$$\begin{aligned} m &= \sum_{i=1}^k \left(\nabla \overline{F}_t^*(\nabla F_t(x) - L)\right)_i = \sum_{i=1}^k (\nabla F_t^*(\nabla F_t(x) - L + \lambda\vec{1}))_i \\ &= \sum_{i=1}^k {f_t^*}'(f_t'(x_i) - L_i + \lambda) < \sum_{i=1}^k {f_t^*}'(f_t'(x_i)) = \sum_{i=1}^k x_i = m\,, \end{aligned}$$

which is a contradiction and completes the proof. □

Claim 26. *For any Legendre function f with monotonically decreasing second derivative, $x \in \mathrm{dom}(f)$ and $\ell \in \mathbb{R}$ such that $f'(x) - \ell \in \mathrm{dom}(f^*)$:*

$$\mathcal{B}_{f^*}(f'(x) - \ell, f'(x)) \leq \frac{\ell^2}{2f''(x)}\,.$$

Proof. For any Legendre function f with monotonically decreasing second derivative, $x \in \mathrm{dom}(f)$ and $\ell \in [0,\infty)$ such that $f'(x) - \ell \in \mathrm{dom}(f^*)$:

$$\mathcal{B}_{f^*}(f'(x) - \ell, f'(x)) \leq \frac{\ell^2}{2f''(x)}\,.$$

□

Claim 27. *For each $j \neq i$ and $c > 0$ holds*

$$\nabla \overline{F}^*\,(-L)_i \geq \nabla \overline{F}^*\left(-L + c\boldsymbol{e}_j\right)_i$$

Proof. Let $x = \nabla \overline{F}^*\left(-L + c\mathbf{e}_j\right)$, this definition is equivalent to $\nabla F(x) - c\mathbf{e}_j = -L$ then

$$\begin{aligned} \nabla \overline{F}^*\,(-L)_i &= \nabla \overline{F}^*\left(\nabla F(x) - c\mathbf{e}_j\right)_i \geq \nabla F^*\left(\nabla F(x) - c\mathbf{e}_j\right)_i \\ &= f^{*\prime}\left(f'(x_i) - c\left(\mathbf{e}_j\right)_i\right) = f^{*\prime}\left(f'(x_i)\right) = x_i \\ &= \nabla \overline{F}^*\left(-L + c\mathbf{e}_j\right)_i \end{aligned}$$

□

4.6.4 Proof of sub-optimal bound in Eq. (4.7)

In this section we use the non-optimal negative entropy regularizer which corresponds to taking just the term $F_{t,2}$ in Eq. (4.23). For convenience of notation we call in this section F_t the term $F_{t,2}$ of the previous section.

Like we did before for L_t, let us define $\widehat{L}_t$ as the cumulative estimated loss vector with components $i \in [k]$ that are given by

$$\widehat{L}_{t,i} = \sum_{s=1}^{t-1} \widehat{\ell}_{s,i}.$$

Lemma 21. *For any t it holds*

$$\mathbb{E}\left[\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle\right] \leq \frac{k}{2}\eta_t\,.$$

Proof. We have

$$\begin{aligned}
&\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&= \overline{F}_t^*(\nabla F_t(x_t) - \widehat{\ell}_t + \lambda\vec{1}) - \overline{F}_t^*(\nabla F_t(x_t) + \lambda\vec{1}) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&= \overline{F}_t^*(\nabla F_t(x_t) - \widehat{\ell}_t) - \overline{F}_t^*(\nabla F_t(x_t)) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&\leq F_t^*(\nabla F_t(x_t) - \widehat{\ell}_t) - F_t^*(\nabla F_t(x_t)) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&= \sum_{i=1}^{k} \mathcal{B}_{f_t^*}(f_t'(x_{t,i}) - \widehat{\ell}_{t,i}, f_t'(x_{t,i})) \\
&= \sum_{i=1}^{k} \mathcal{B}_{f_t^*}(f_t'(x_{t,i}) - \frac{A_{t,i}\ell_{t,i}}{x_{t,i}}, f_t'(x_{t,i})) \\
&\leq \frac{1}{2}\sum_{i=1}^{k} A_{t,i}\frac{\ell_{t,i}^2}{x_{t,i}^2} f_t''(x_{t,i})^{-1} \\
&= \frac{\eta_t}{2}\sum_{i=1}^{k} A_{t,i}\frac{\ell_{t,i}^2}{x_{t,i}},
\end{aligned}$$

where the first equality follows using lemma 23, the second equality follows by Lemma 22, the first inequality is from Lemma 24 and finally the second inequality follows from Lemma 26. In expectation we get

$$\begin{aligned}
&\mathbb{E}\left[\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle\right] \\
&\leq \frac{\eta_t}{2}\sum_{i=1}^{k} \mathbb{E}\left[A_{t,i}\right]\frac{\ell_{t,i}^2}{x_{t,i}} = \frac{\eta_t}{2}\sum_{i=1}^{k}\ell_{t,i}^2 \leq \frac{k}{2}\eta_t.
\end{aligned}$$

□

Lemma 22. *For any non-increasing learning rate η_t, it holds that*

$$\sum_{t=1}^{T}\left(\overline{F}_t^*(-\widehat{L}_t) - \overline{F}_t^*(-\widehat{L}_{t+1}) - \left\langle a, \widehat{\ell}_t \right\rangle\right) \leq \frac{1}{\eta_T}\left(m + m\log\left(\frac{k}{m}\right)\right).$$

Proof. Let $\tilde{x}_t = \operatorname{argmax}_{x \in \mathrm{co}(\mathcal{A})} \left\{ \left\langle x, -\widehat{L}_t \right\rangle - F_t(x) \right\}$, then

$$\overline{F}_t^* \left(-\widehat{L}_t \right) = \left\langle \tilde{x}_t, \widehat{L}_t \right\rangle - F_t(\tilde{x}_t).$$

Furthermore, since $\overline{F}^* \left(-\widehat{L}_t \right) = \max_{x \in \mathrm{co}(\mathcal{A})} \left\{ \left\langle x, -\widehat{L}_t \right\rangle - F(x) \right\}$, we have

$$\begin{aligned}
&- \overline{F}_{t-1}^*(-\widehat{L}_t) \leq \left\langle \tilde{x}_t, \widehat{L}_t \right\rangle + F_{t-1}(\tilde{x}_t) \\
&- \overline{F}_T^* \left(-\widehat{L}_{T+1} \right) \leq \left\langle a, \widehat{L}_{T+1} \right\rangle + F_T(a) \leq \sum_{t=1}^{T} \left\langle a, \widehat{\ell}_t \right\rangle ,
\end{aligned}$$

where we observe that $F_T(a) \leq 0$ for all $a \in [0,1]^k$. Plugging these inequalities into the LHS leads to

$$\begin{aligned}
&\sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t) - \overline{F}_t^*(-\widehat{L}_{t+1}) - \left\langle a, \widehat{\ell}_t \right\rangle \right) \\
&= \sum_{t=1}^{T} \overline{F}_t^*(-\widehat{L}_t) - \sum_{t=2}^{T} \overline{F}_{t-1}^*(-\widehat{L}_t) - \overline{F}_T^*(-\widehat{L}_{T+1}) - \sum_{t=1}^{T} \left\langle a, \widehat{\ell}_t \right\rangle \\
&\leq \sum_{t=1}^{T} \overline{F}_t^*(-\widehat{L}_t) - \sum_{t=2}^{T} \overline{F}_{t-1}^*(-\widehat{L}_t) \\
&= \sum_{t=1}^{T} \left\langle \tilde{x}_t, -\widehat{L}_t \right\rangle - \sum_{t=1}^{T} F_t(\tilde{x}_t) + \sum_{t=2}^{T} \left\langle \tilde{x}_t, \widehat{L}_t \right\rangle + \sum_{t=2}^{T} F_{t-1}(\tilde{x}_t) \\
&\leq \sum_{t=1}^{T} F_{t-1}(\tilde{x}_t) - F_t(\tilde{x}_t) \\
&= \sum_{t=1}^{T} \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) (-F(\tilde{x}_t)) \\
&\leq \sum_{t=1}^{T} \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) \max_{x \in \mathrm{co}(\mathcal{A})} (-F(x)) \\
&= \frac{1}{\eta_T} \max_{x \in \mathrm{co}(\mathcal{A})} \{-F(x)\} ,
\end{aligned}$$

where with a slight abuse of notation we defined $\eta_0^{-1} = 0$. We are left with computing the maximum:

$$\begin{aligned}
\max_{x \in \mathrm{co}(\mathcal{A})} \{-F(x)\} &= \max_{x \in \mathrm{co}(\mathcal{A})} \left\{ \sum_{i=1}^{k} x_i + \sum_{i=1}^{k} x_i \log \left(\frac{1}{x_i} \right) \right\} \\
&= \max_{x \in \mathrm{co}(\mathcal{A})} \left\{ m + m \frac{\sum_{i=1}^{k} x_i}{m} \log \left(\frac{1}{x_i} \right) \right\} \\
&\leq m + m \log \left(\frac{k}{m} \right) .
\end{aligned}$$

Follows that

$$\sum_{t=1}^{T}\left(\overline{F}_t^*(-\widehat{L}_t)-\overline{F}_t^*(-\widehat{L}_{t+1})-\left\langle a,\widehat{\ell}_t\right\rangle\right)\leq\frac{1}{\eta_T}\left(m+m\log\left(\frac{k}{m}\right)\right).$$

□

Lemma 23. *For any t it holds that*

$$\overline{F}_t^*(-\widehat{L}_t^{obs})-\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{\ell}_t)-\overline{F}_t^*(-\widehat{L}_t)+\overline{F}_t^*(-\widehat{L}_{t+1})\leq\eta_t k\mathfrak{d}_t.$$

Proof. We define $\widehat{L}_t^{miss}=\widehat{L}_t-\widehat{L}_t^{obs}$. Then we have

$$\begin{aligned}
&-\overline{F}_t^*(-\widehat{L}_t)+\overline{F}_t^*(-\widehat{L}_{t+1})\\
&=-\int_0^1\left\langle\widehat{\ell}_t,\nabla\overline{F}_t^*(-\widehat{L}_t-x\widehat{\ell}_t)\right\rangle dx\\
&=-\int_0^1\left\langle\widehat{\ell}_t,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}-x\widehat{\ell}_t)\right\rangle dx\\
&=-\int_0^1\sum_{i=1}^{k}\frac{A_{t,i}\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}-x\widehat{\ell}_t)_i\,dx\\
&=-\sum_{i:A_{t,i}=1}\int_0^1\frac{\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}-x\widehat{\ell}_t)_i\,dx\\
&\leq-\sum_{i:A_{t,i}=1}\int_0^1\frac{\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}+\sum_{j:A_{t,j}=0}^{k}\widehat{L}_{t,j}^{miss}\mathbf{e}_j-x\widehat{\ell}_t)_i\,dx\\
&=-\sum_{i:A_{t,i}=1}\int_0^1\frac{\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\sum_{j=1}^{k}A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j-x\widehat{\ell}_t)_i\,dx\\
&=-\int_0^1\left\langle\widehat{\ell}_t,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\sum_{j=1}^{k}A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j-x\widehat{\ell}_t)\right\rangle dx\,,
\end{aligned}$$

where the first equality uses the fundamental theorem of calculus and the inequality follows from Claim 27. Now let us define $\tilde{z}(x)=\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-x\widehat{\ell}_t)$. We have the following

$$
\begin{aligned}
&\overline{F}_t^*(-\widehat{L}_t^{obs}) - \overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t) + \overline{F}_t^*(-\widehat{L}_{t+1}) \\
&\leq \int_0^1 \left\langle \widehat{\ell}_t, \nabla \overline{F}_t^*(-\widehat{L}_t^{obs} - x\widehat{\ell}_t) \right\rangle dx - \int_0^1 \left\langle \widehat{\ell}_t, \nabla \overline{F}_t^*(-\widehat{L}_t^{obs} - \sum_{j=1}^k A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j - x\widehat{\ell}_t) \right\rangle dx \\
&= \int_0^1 \left\langle \widehat{\ell}_t, \tilde{z}(x) - \nabla \overline{F}_t^* \left(\nabla F_t(\tilde{z}(x)) - \sum_{j=1}^k A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j \right) \right\rangle dx \\
&\leq \int_0^1 \left\langle \widehat{\ell}_t, \tilde{z}(x) - \nabla F_t^* \left(\nabla F_t(\tilde{z}(x)) - \sum_{j=1}^k A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j \right) \right\rangle dx \\
&= \sum_{i=1}^k \int_0^1 \widehat{\ell}_{t,i} \left(\tilde{z}(x) - \nabla F_t^* \left(\nabla F_t(\tilde{z}(x)) - \sum_{j=1}^k A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j \right) \right)_i dx \\
&= \sum_{i=1}^k \int_0^1 \widehat{\ell}_{t,i} \left(\tilde{z}_i(x) - f_t^{*\prime} \left(f_t'(\tilde{z}_i(x)) - \widehat{L}_{t,i}^{miss} \right) \right) dx \\
&\leq \sum_{i=1}^k \int_0^1 \widehat{\ell}_{t,i} f_t^{*\prime\prime}(f_t'(\tilde{z}_i(x)))\widehat{L}_{t,i}^{miss}\, dx \\
&= \sum_{i=1}^k \int_0^1 \widehat{\ell}_{t,i} \left(f_t{}''(\tilde{z}_i(x)) \right)^{-1} \widehat{L}_{t,i}^{miss}\, dx \\
&= \eta_t \int_0^1 \left(\sum_{i=1}^k \widehat{\ell}_{t,i}\widehat{L}_{t,i}^{miss}\tilde{z}_i(x) \right) dx\,,
\end{aligned}
$$

where the first inequality uses the fundamental theorem of calculus together with the inequality above, the first equality substitutes $\widetilde{z}(x) = \nabla \bar{F}_t^* \left(-\hat{L}_t^{obs} - x\hat{\ell}_t \right)$ and applies Claim 23. The second inequality applies Claim 25 and the third uses the fact that $f^{*\prime}(t)$ is convex, so $-f^{*\prime}\left(f'\left(\tilde{z}_{A_t}\right) - \ell\right) \leq -\tilde{z}_{A_t} + f^{*\prime\prime}\left(f'\left(\tilde{z}_{A_t}\right)\right)$. The second to last equality follows by Eq. (4.20), and the last follows because $f_t{}''(x) = \frac{1}{x}$. Taking the expected

value we have

$$
\begin{aligned}
&\mathbb{E}\left[\overline{F}_t^*(-\widehat{L}_t^{obs}) - \overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t) + \overline{F}_t^*(-\widehat{L}_{t+1})\right] \\
&\leq \eta_t \mathbb{E}\left[\int_0^1 \left(\sum_{i=1}^k \widehat{\ell}_{t,i}\widehat{L}_{t,i}^{miss} z_i(x)\right) dx\right] \\
&\leq \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{\ell}_{t,i}\widehat{L}_{t,i}^{miss}\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{L}_{t,i}^{miss}\mathbb{E}_t\left[\widehat{\ell}_{t,i}\right]\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{L}_{t,i}^{miss}\ell_{t,i}\right] \\
&\leq \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{L}_{t,i}^{miss}\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\}\widehat{\ell}_{s,i}\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\}\mathbb{E}_s\left[\widehat{\ell}_{s,i}\right]\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\}\ell_{s,i}\right] \\
&\leq \eta_t k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\} \\
&\leq \eta_t k \mathfrak{d}_t\,.
\end{aligned}
$$

□

Theorem 28 (Proof of Eq. (4.7))**.** *Algorithm 8 with proper learning rates* $(\eta_t)_{t=1,\ldots,n}$ *satisfies*

$$R_T \leq \frac{m}{\eta_T}\left(1 + \log\left(\frac{k}{m}\right)\right) + \frac{k}{2}\sum_{t=1}^T \eta_t + k\sum_{t=1}^T \eta_t \mathfrak{d}_t.$$

Furthermore if one chooses $\eta_t = \sqrt{\frac{m(1+\log(k/m))}{2k\sum_{s=1}^t(\mathfrak{d}_s+1)}}$ *then*

$$R_T \leq 2\sqrt{2km(1+\log(k/m))\left(T + \sum_{t=1}^T \mathfrak{d}_t\right)}.$$

Proof. We have the following decomposition of R_T

$$
\begin{aligned}
R_T &= \mathbb{E}\left[\sum_{t=1}^{T} \langle x_t - a, \ell_t \rangle\right] = \mathbb{E}\left[\sum_{t=1}^{T} \left\langle x_t, \widehat{\ell}_t \right\rangle - \left\langle a, \widehat{\ell}_t \right\rangle\right] \\
&= \mathbb{E}\left[\sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t) - \overline{F}_t^*(-\widehat{L}_{t+1}) - \left\langle a, \widehat{\ell}_t \right\rangle\right)\right. \\
&\quad \sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle\right) \\
&\quad \left.\sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t^{obs}) - \overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t) + \overline{F}_t^*(-\widehat{L}_{t+1})\right)\right] \\
&\leq \frac{m}{\eta_T}\left(1 + \log\left(\frac{k}{m}\right)\right) + \frac{k}{2}\sum_{t=1}^{T}\eta_t + k\sum_{t=1}^{T}\eta_t \mathfrak{d}_t.
\end{aligned}
$$

□

4.6.5 Proof of Theorem 10

Lemma 24. *For any t it holds*

$$
\sum_{t=1}^{T} \mathbb{E}\left[\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle\right] \leq \sqrt{Tkm}\,.
$$

Proof. We have

$$\begin{aligned}
&\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&= \overline{F}_t^*(\nabla F_t(x_t) - \widehat{\ell}_t + \lambda\vec{1}) - \overline{F}_t^*(\nabla F_t(x_t) + \lambda\vec{1}) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&= \overline{F}_t^*(\nabla F_t(x_t) - \widehat{\ell}_t) - \overline{F}_t^*(\nabla F_t(x_t)) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&\leq F_t^*(\nabla F_t(x_t) - \widehat{\ell}_t) - F_t^*(\nabla F_t(x_t)) + \left\langle x_t, \widehat{\ell}_t \right\rangle \\
&= \sum_{i=1}^{k} \mathcal{B}_{f_t^*}(f_t'(x_{t,i}) - \widehat{\ell}_{t,i}, f_t'(x_{t,i})) \\
&= \sum_{i=1}^{k} \mathcal{B}_{f_t^*}(f_t'(x_{t,i}) - \frac{A_{t,i}\ell_{t,i}}{x_{t,i}}, f_t'(x_{t,i})) \\
&\leq \frac{1}{2}\sum_{i=1}^{k} A_{t,i}\frac{\ell_{t,i}^2}{x_{t,i}^2} f_t''(x_{t,i})^{-1} \\
&\leq \frac{1}{2}\sum_{i=1}^{k} A_{t,i}\frac{\ell_{t,i}^2}{x_{t,i}^2} f_{t,1}''(x_{t,i})^{-1} \\
&= \sum_{i=1}^{k} A_{t,i}\frac{\ell_{t,i}^2}{x_{t,i}^2}\frac{x_{t,i}^{\frac{3}{2}}}{\sqrt{t}} \\
&= \sum_{i=1}^{k} A_{t,i}\ell_{t,i}^2\frac{x_{t,i}^{-\frac{1}{2}}}{\sqrt{t}},
\end{aligned}$$

where the first equality follows using lemma 23, the second equality follows by Lemma 22, the first inequality is from Lemma 24 and finally the second inequality follows from Lemma 26. In expectation we get

$$\begin{aligned}
&\mathbb{E}\left[\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle\right] \\
&\leq \frac{1}{2}\sum_{i=1}^{k} \mathbb{E}\left[A_{t,i}\right]\ell_{t,i}^2\frac{x_{t,i}^{-\frac{1}{2}}}{\sqrt{t}} = \frac{1}{2}\sum_{i=1}^{k}\frac{x_{t,i}^{\frac{1}{2}}}{\sqrt{t}}\ell_{t,i}^2 \leq \frac{1}{2}\sum_{i=1}^{k}\frac{x_{t,i}^{\frac{1}{2}}}{\sqrt{t}}.
\end{aligned}$$

Summing over t and using Cauchy-Schwarz gives

$$\frac{1}{2}\sum_{t=1}^{T}\sum_{i=1}^{k}\frac{x_{t,i}^{\frac{1}{2}}}{\sqrt{t}} \leq \sqrt{Tkm}.$$

□

Lemma 25. *For any non-increasing learning rate η_t, it holds that for every $a \in \mathcal{S}$*

$$\sum_{t=1}^{T}\left(\overline{F}_t^*(-\widehat{L}_t) - \overline{F}_t^*(-\widehat{L}_{t+1}) - \left\langle a, \widehat{\ell}_t \right\rangle\right) \leq 2\sqrt{Tkm} + \frac{m\log\left(\frac{k}{m}\right)}{\eta_T}.$$

Proof. Let $\tilde{x}_t = \operatorname{argmax}_{x \in \mathrm{co}(\mathcal{A})} \left\{ \left\langle x, -\widehat{L}_t \right\rangle - F_t(x) \right\}$, then

$$\overline{F}_t^* \left(-\widehat{L}_t \right) = \left\langle \tilde{x}_t, \widehat{L}_t \right\rangle - F_t(\tilde{x}_t).$$

Furthermore, since $\overline{F}^* \left(-\widehat{L}_t \right) = \max_{x \in \mathrm{co}(\mathcal{A})} \left\{ \left\langle x, -\widehat{L}_t \right\rangle - F(x) \right\}$, we have

$$\begin{aligned} -\overline{F}_{t-1}^*(-\widehat{L}_t) &\leq \left\langle \tilde{x}_t, \widehat{L}_t \right\rangle + F_{t-1}(\tilde{x}_t) \\ -\overline{F}_T^* \left(-\widehat{L}_{T+1} \right) &\leq \left\langle a, \widehat{L}_{T+1} \right\rangle + F_T(a) = \sum_{t=1}^{T} \left\langle a, \widehat{\ell}_t \right\rangle . \end{aligned}$$

Plugging these inequalities into the LHS leads to

$$\begin{aligned} &\sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t) - \overline{F}_t^*(-\widehat{L}_{t+1}) - \left\langle a, \widehat{\ell}_t \right\rangle \right) \\ &= \sum_{t=1}^{T} \overline{F}_t^*(-\widehat{L}_t) - \sum_{t=2}^{T} \overline{F}_{t-1}^*(-\widehat{L}_t) - \overline{F}_T^*(-\widehat{L}_{T+1}) - \left\langle a, \widehat{\ell}_t \right\rangle \\ &\leq \sum_{t=1}^{T} \overline{F}_t^*(-\widehat{L}_t) - \sum_{t=2}^{T} \overline{F}_{t-1}^*(-\widehat{L}_t) \\ &= \sum_{t=1}^{T} \left\langle \tilde{x}_t, -\widehat{L}_t \right\rangle - \sum_{t=1}^{T} F_t(\tilde{x}_t) + \sum_{t=2}^{T} \left\langle \tilde{x}_t, \widehat{L}_t \right\rangle + \sum_{t=2}^{T} F_{t-1}(\tilde{x}_t) \\ &\leq \sum_{t=1}^{n} F_{t-1}(\tilde{x}_t) - F_t(\tilde{x}_t) \\ &= \sum_{t=1}^{n} F_{1,t-1}(\tilde{x}_t) + F_{2,t-1}(\tilde{x}_t) - F_{1,t}(\tilde{x}_t) - F_{2,t}(\tilde{x}_t) \\ &= \sum_{t=1}^{n} \left(\sqrt{t} - \sqrt{t-1} \right) (-F_1(\tilde{x}_t)) + \sum_{t=1}^{n} \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) (-F_2(\tilde{x}_t)) \\ &= \sum_{t=1}^{n} \left(\sqrt{t} - \sqrt{t-1} \right) \max_{x \in \mathrm{co}(\mathcal{A})} (-F_1(x)) + \sum_{t=1}^{n} \left(\frac{1}{\eta_t} - \frac{1}{\eta_{t-1}} \right) \max_{x \in \mathrm{co}(\mathcal{A})} (-F_2(x)) \\ &= \sqrt{T} \max_{x \in \mathrm{co}(\mathcal{A})} (-F_1(x)) + \frac{1}{\eta_T} \max_{x \in \mathrm{co}(\mathcal{A})} (-F_2(x)) \end{aligned}$$

We are left with computing the maximum and using Hölder we get:

$$\begin{aligned} \max_{x \in \mathrm{co}(\mathcal{A})} \{-F_1(x)\} &= \max_{x \in \mathrm{co}(\mathcal{A})} \left\{ 2 \sum_{i=1}^{k} x_i^{\frac{1}{2}} \right\} \\ &\leq 2\sqrt{km} \,, \end{aligned}$$

where the inequality is due to Cauchy-Schwarz; also

$$\begin{aligned} \max_{x \in \mathrm{co}(\mathcal{A})} \{-F_2(x)\} &= \max_{x \in \mathrm{co}(\mathcal{A})} \left\{ \sum_{i=1}^{k} x_i \log \left(\frac{1}{x_i} \right) \right\} \\ &\leq m \log \left(\frac{k}{m} \right) , \end{aligned}$$

where we used Jensen's inequality. Follows that

$$\sum_{t=1}^{T}\left(\overline{F}_t^*(-\widehat{L}_t)-\overline{F}_t^*(-\widehat{L}_{t+1})-\left\langle a,\widehat{\ell}_t\right\rangle\right)\leq 2\sqrt{Tkm}+\frac{m\log\left(\frac{k}{m}\right)}{\eta_T}.$$

□

Lemma 26. *For any t it holds that*

$$\overline{F}_t^*(-\widehat{L}_t^{obs})-\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{\ell}_t)-\overline{F}_t^*(-\widehat{L}_t)+\overline{F}_t^*(-\widehat{L}_{t+1})\leq\eta_t k\mathfrak{d}_t.$$

Proof. We define $\widehat{L}_t^{miss}=\widehat{L}_t-\widehat{L}_t^{obs}$. Then we have

$$\begin{aligned}
&-\overline{F}_t^*(-\widehat{L}_t)+\overline{F}_t^*(-\widehat{L}_{t+1})\\
&=-\int_0^1\left\langle\widehat{\ell}_t,\nabla\overline{F}_t^*(-\widehat{L}_t-x\widehat{\ell}_t)\right\rangle dx\\
&=-\int_0^1\left\langle\widehat{\ell}_t,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}-x\widehat{\ell}_t)\right\rangle dx\\
&=-\int_0^1\sum_{i=1}^{k}\frac{A_{t,i}\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}-x\widehat{\ell}_t)_i\,dx\\
&=-\sum_{i:A_{t,i}=1}\int_0^1\frac{\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}-x\widehat{\ell}_t)_i\,dx\\
&\leq-\sum_{i:A_{t,i}=1}\int_0^1\frac{\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\widehat{L}_t^{miss}+\sum_{j:A_{t,j}=0}^{k}\widehat{L}_{t,j}^{miss}\mathbf{e}_j-x\widehat{\ell}_t)_i\,dx\\
&=-\sum_{i:A_{t,i}=1}\int_0^1\frac{\ell_{t,i}}{x_{t,i}}\,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\sum_{j=1}^{k}A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j-x\widehat{\ell}_t)_i\,dx\\
&=-\int_0^1\left\langle\widehat{\ell}_t,\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-\sum_{j=1}^{k}A_{t,j}\widehat{L}_{t,j}^{miss}\mathbf{e}_j-x\widehat{\ell}_t)\right\rangle dx\,,
\end{aligned}$$

where the first equality uses the fundamental theorem of calculus and the inequality follows from Claim 27. Now let us define $\tilde{z}(x)=\nabla\overline{F}_t^*(-\widehat{L}_t^{obs}-x\widehat{\ell}_t)$. We have the following

$$
\begin{aligned}
&\overline{F}_t^*(-\widehat{L}_t^{obs}) - \overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t) + \overline{F}_t^*(-\widehat{L}_{t+1}) \\
&\leq \int_0^1 \left\langle \widehat{\ell}_t, \nabla \overline{F}_t^*(-\widehat{L}_t^{obs} - x\widehat{\ell}_t) \right\rangle dx - \int_0^1 \left\langle \widehat{\ell}_t, \nabla \overline{F}_t^*(-\widehat{L}_t^{obs} - \sum_{j=1}^k A_{t,j} \widehat{L}_{t,j}^{miss} \mathbf{e}_j - x\widehat{\ell}_t) \right\rangle dx \\
&= \int_0^1 \left\langle \widehat{\ell}_t, \tilde{z}(x) - \nabla \overline{F}_t^*(\nabla F_t(\tilde{z}(x)) - \sum_{j=1}^k A_{t,j} \widehat{L}_{t,j}^{miss} \mathbf{e}_j) \right\rangle dx \\
&\leq \int_0^1 \left\langle \widehat{\ell}_t, \tilde{z}(x) - \nabla F_t^*(\nabla F_t(\tilde{z}(x)) - \sum_{j=1}^k A_{t,j} \widehat{L}_{t,j}^{miss} \mathbf{e}_j) \right\rangle dx \\
&= \sum_{i=1}^k \int_0^1 \widehat{\ell}_{t,i} \left(\tilde{z}(x) - \nabla F_t^* \left(\nabla F_t(\tilde{z}(x)) - \sum_{j=1}^k A_{t,j} \widehat{L}_{t,j}^{miss} \mathbf{e}_j \right) \right)_i dx \\
&= \sum_{i=1}^k \int_0^1 \widehat{\ell}_{t,i} (\tilde{z}_i(x) - f_t^{*\prime}(f_t'(\tilde{z}_i(x)) - \widehat{L}_{t,i}^{miss}) \, dx \\
&\leq \sum_{i=1}^k \int_0^1 \widehat{\ell}_{t,i} f_t^{*\prime\prime}(f_t'(\tilde{z}_i(x))) \widehat{L}_{t,i}^{miss} \, dx \\
&= \sum_{i=1}^k \int_0^1 \frac{\widehat{\ell}_{t,i}}{f_t''(\tilde{z}_i(x))} \widehat{L}_{t,i}^{miss} \, dx \\
&\leq \sum_{i=1}^k \int_0^1 \frac{\widehat{\ell}_{t,i}}{f_{t,2}''(\tilde{z}_i(x))} \widehat{L}_{t,i}^{miss} \, dx \\
&= \eta_t \int_0^1 \left(\sum_{i=1}^k \widehat{\ell}_{t,i} \widehat{L}_{t,i}^{miss} \tilde{z}_i(x) \right) dx
\end{aligned}
$$

where the first inequality uses the fundamental theorem of calculus together with the inequality above, the first equality substitutes $\tilde{z}(x) = \nabla \bar{F}_t^* \left(-\hat{L}_t^{obs} - x\hat{\ell}_t \right)$ and applies Claim 23. The second inequality applies Claim 25 and the third uses the fact that $f^{*\prime}(t)$ is convex, so $-f^{*\prime}\left(f'\left(\tilde{z}_{A_t}\right) - \ell\right) \leq -\tilde{z}_{A_t} + f^{*\prime\prime}\left(f'\left(\tilde{z}_{A_t}\right)\right)$. Taking the expected value we have

$$
\begin{aligned}
&\mathbb{E}\left[\overline{F}_t^*(-\widehat{L}_t^{obs}) - \overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t) + \overline{F}_t^*(-\widehat{L}_{t+1})\right] \\
&\leq \eta_t \mathbb{E}\left[\int_0^1 \left(\sum_{i=1}^k \widehat{\ell}_{t,i}\widehat{L}_{t,i}^{miss} z_i(x)\right) dx\right] \\
&\leq \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{\ell}_{t,i}\widehat{L}_{t,i}^{miss}\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{L}_{t,i}^{miss}\mathbb{E}_t\left[\widehat{\ell}_{t,i}\right]\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{L}_{t,i}^{miss}\ell_{t,i}\right] \\
&\leq \eta_t \mathbb{E}\left[\sum_{i=1}^k \widehat{L}_{t,i}^{miss}\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\}\widehat{\ell}_{s,i}\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\}\mathbb{E}_s\left[\widehat{\ell}_{s,i}\right]\right] \\
&= \eta_t \mathbb{E}\left[\sum_{i=1}^k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\}\ell_{s,i}\right] \\
&\leq \eta_t k \sum_{s:s<t} \mathbb{I}\left\{s + d_s \geq t\right\} \\
&\leq \eta_t k \mathfrak{d}_t \,.
\end{aligned}
$$

□

Theorem 10. *Algorithm 8 with proper learning rates* $(\eta_t)_{t=1,\dots,n}$ *satisfies*

$$
R_T \leq \frac{m\log\left(\frac{k}{m}\right)}{\eta_T} + 3\sqrt{Tkm} + k\sum_{t=1}^T \eta_t \mathfrak{d}_t
$$

Furthermore if one chooses $\eta_t = \sqrt{\frac{m(1+\log(k/m))}{2k\sum_{s=1}^t \mathfrak{d}_s}}$ *then*

$$
R_T \leq 3\sqrt{Tkm} + 2\sqrt{2km\log\left(k/m\right)\left(\sum_{t=1}^T \mathfrak{d}_t\right)}.
$$

Proof. We have the following decomposition of R_T

$$\begin{aligned} R_T = \mathbb{E}\Bigg[& \sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t) - \overline{F}_t^*(-\widehat{L}_{t+1}) - \left\langle a, \widehat{\ell}_t \right\rangle \right) \\ & \sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t^{obs}) + \left\langle x_t, \widehat{\ell}_t \right\rangle \right) \\ & \sum_{t=1}^{T} \left(\overline{F}_t^*(-\widehat{L}_t^{obs}) - \overline{F}_t^*(-\widehat{L}_t^{obs} - \widehat{\ell}_t) - \overline{F}_t^*(-\widehat{L}_t) + \overline{F}_t^*(-\widehat{L}_{t+1}) \right) \Bigg] \\ & = 2\sqrt{Tkm} + \frac{m \log\left(\frac{k}{m}\right)}{\eta_T} + \sqrt{Tkm} + \sum_{t=1}^{T} \eta_t k \mathfrak{d}_t \end{aligned}$$

□

4.6.6 Proof of lemmas from Section 4.5

Lemma 11. *If agent $v \in \mathcal{V}$ runs the Algorithm 2 with learning rate $\eta > 0$, the following deterministic bound holds for all $i \in \{1, \dots, k\}$:*

$$\sum_{t=1}^{T} \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) - \sum_{t=1}^{T} \widehat{\ell}_t(i^\star, v) \le \frac{\ln k}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v)^2 \,.$$

Proof. For all $t \in \{2, \dots, T\}$, $i \in \{1, \dots, K\}$, and $v \in \mathcal{V}$, let

$$w'_t(i,v) = w'_{t-1}(i,v) \, \exp\left(-\eta \widehat{\ell}_{t-1}(i,v)\right) \qquad \text{and} \qquad w'_1(i,v) = \frac{1}{k} = w_1(i,v) \,.$$

Note that, for all t, i, v, we have $\frac{w_t(i,v)}{W_t(v)} = \frac{w'_t(i,v)}{W'_t(v)}$, where $W'_t(v) = \sum_{i=1}^{k} w'_t(i,v)$. Then,

$$\begin{aligned} \frac{W'_{t+1}(v)}{W'_t(v)} &= \sum_{i=1}^{k} \frac{w'_{t+1}(i,v)}{W'_t(v)} \\ &= \sum_{i=1}^{k} \frac{w'_t(i,v) e^{-\eta \widehat{\ell}_t(i,v)}}{W'_t(v)} \\ &= \sum_{i=1}^{k} \frac{w_t(i,v) e^{-\eta \widehat{\ell}_t(i,v)}}{W_t(v)} \\ &= \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} e^{-\eta \widehat{\ell}_t(i,v)} \\ &\le \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \left(1 - \eta \widehat{\ell}_t(i,v) + \frac{1}{2}\eta^2 \left(\widehat{\ell}_t(i,v)\right)^2\right) \\ &= 1 - \eta \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) + \frac{\eta^2}{2} \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \left(\widehat{\ell}_t(i,v)\right)^2 \end{aligned}$$

Taking the logarithm and using the inequality $\ln(1+x) \le x$ for all $x > -1$, and summing over $t = 1, \dots, T$ yields

$$\ln \frac{W'_{T+1}(v)}{W'_1(v)} \leq -\eta \sum_{t=1}^{T} \sum_{i=1}^{K} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) + \frac{\eta^2}{2} \sum_{t=1}^{T} \sum_{i=1}^{K} \frac{\overline{x}_t(i,v)}{m} \left(\widehat{\ell}_t(i,v)\right)^2 .$$

Moreover, for any fixed comparison action $i^\star$, we also have

$$\ln \frac{W'_{T+1}(v)}{W'_1(v)} \geq \ln \frac{w'_t(i^\star,v)}{W'_1(v)} = -\eta \sum_{t=1}^{T} \widehat{\ell}_t(i^\star,v) - \ln k$$

Putting this together and rearranging gives

$$\sum_{t=1}^{T} \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) - \sum_{t=1}^{T} \widehat{\ell}_t(i^\star,v) \leq \frac{\ln k}{\eta} + \frac{\eta}{2} \sum_{t=1}^{T} \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v)^2 .$$

□

Lemma 12. *If agent $v \in \mathcal{V}$ runs baditCoopMsets with learning rate $\eta > 0$, the following deterministic bounds for the drift probabilities hold for all $i \in \{1,\ldots,k\}$:*

$$-\eta \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) \leq \frac{\overline{x}_{t+1}(i,v)}{m} - \frac{\overline{x}_t(i,v)}{m} \leq \eta \frac{\overline{x}_{t+1}(i,v)}{m} \sum_{j=1}^{k} \frac{\overline{x}_t(j,v)}{m} \widehat{\ell}_t(j,v) .$$

Proof. Directly from the definition of the update $w_{t+1}(i,v) \leq \overline{x}_t(i,v)$ for all $i \in \{1,\ldots,k\}$, so that $W_{t+1}(v) \leq m$ which in turn implies

$$w_{t+1}(i,v) \leq \frac{w_{t+1}(i,v)}{W_{t+1}(v)/m} = \overline{x}_{t+1}(i,v).$$

Therefore,

$$\frac{\overline{x}_{t+1}(i,v)}{m} - \frac{\overline{x}_t(i,v)}{m} \geq \frac{w_{t+1}(i,v)}{m} - \frac{\overline{x}_t(i,v)}{m} = -\frac{\overline{x}_t(i,v)}{m}\left(1 - e^{-\eta \widehat{\ell}_t(i,v)}\right) \geq -\eta \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v),$$

the last inequality using $1 - e^{-x} \leq x$ for $x \geq 0$. Similarly,

$$\begin{aligned}
\frac{\overline{x}_{t+1}(i,v)}{m} - \frac{\overline{x}_t(i,v)}{m} &\leq \frac{\overline{x}_{t+1}(i,v)}{m} - \frac{w_{t+1}(i,v)}{m} \\
&= \frac{\overline{x}_{t+1}(i,v)}{m} - \frac{\overline{x}_{t+1}(i,v)}{m^2} W_{t+1}(v) \\
&= \frac{\overline{x}_{t+1}(i,v)}{m}\left(1 - \frac{W_{t+1}(v)}{m}\right) \\
&= \frac{\overline{x}_{t+1}(i,v)}{m} \sum_{j=1}^{k} \left(\frac{\overline{x}_t(j,v)}{m} - \frac{w_{t+1}(j,v)}{m}\right) \\
&= \frac{\overline{x}_{t+1}(i,v)}{m} \left(\sum_{j=1}^{k} \frac{\overline{x}_t(j,v)}{m}\left(1 - e^{-\eta \widehat{\ell}_t(j,v)}\right)\right) \\
&\leq \eta \frac{\overline{x}_{t+1}(i,v)}{m} \sum_{j=1}^{k} \left(\frac{\overline{x}_t(j,v)}{m} \widehat{\ell}_t(j,v)\right).
\end{aligned}$$

□

Lemma 13. *If agent $v \in \mathcal{V}$ runs baditCoopMsets with learning rate $\eta \in \left(0, \frac{m}{ke(d+1)}\right)$, the following deterministic bound holds for all $i \in \{1, \ldots, k\}$:*

$$\overline{x}_{t+1}(i,v) \leq \left(1 + \frac{1}{d}\right) \overline{x}_t(i,v) \; .$$

Proof. We proceed by induction over t. For all $t \leq d$, $\widehat{\ell}_t(\cdot) = 0$. Hence $\overline{x}_t(\cdot) = \frac{m}{K}$ and this lemma trivially holds. For $t > d$ we can write

$$\begin{aligned}
\sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \widehat{\ell}_t(i,v) &= \sum_{i=1}^{k} \frac{\overline{x}_t(i,v)}{m} \frac{\ell_{t-d}(i)}{\overline{B}_{d,t-d}(i,v)} B_{d,t-d}(i,v) \\
&\leq \sum_{i=1}^{k} \frac{x_t(i,v)}{m\overline{B}_{d,t-d}(i,v)} \\
&\leq \sum_{i=1}^{k} \left(1 + \frac{1}{d}\right)^k \frac{\overline{x}_{t-d}(i,v)}{m\overline{B}_{d,t-d}(i,v)} \\
&\leq \left(1 + \frac{1}{d}\right)^k \frac{K}{m} \\
&\leq \frac{K}{m} e
\end{aligned}$$

where the second inequality follows by the inductive hypothesis. Hence, using Lemma 12 we have

$$\frac{\overline{x}_{t+1}(i,v)}{m}\left(1 - \eta\frac{Ke}{m}\right) \leq \frac{\overline{x}_{t+1}(i,v)}{m}\left(1 - \eta\sum_{j=1}^{k}\frac{\overline{x}_t(j,v)}{m}\widehat{\ell}_t(j,v)\right) \leq \frac{\overline{x}_t(i,v)}{m}$$

□

Bibliography

Madhu Advani, Subhaneil Lahiri, and Surya Ganguli. Statistical mechanics of complex neural systems and high dimensional data. *Journal of Statistical Mechanics: Theory and Experiment*, 2013(03):P03014, 2013.

Noga Alon, Nicolo Cesa-Bianchi, Claudio Gentile, Shie Mannor, Yishay Mansour, and Ohad Shamir. Nonstochastic multi-armed bandits with graph-structured feedback. *SIAM Journal on Computing*, 46(6):1785–1826, 2017.

Benjamin Aubin, Will Perkins, and Lenka Zdeborová. Storage capacity in symmetric binary perceptrons. *Journal of Physics A: Mathematical and Theoretical*, 52(29):294003, jun 2019. doi: 10.1088/1751-8121/ab227a.

Jean-Yves Audibert, Sébastien Bubeck, and Gábor Lugosi. Regret in online combinatorial optimization. *Mathematics of Operations Research*, 39(1):31–45, 2014.

Peter Auer, Nicolo Cesa-Bianchi, and Paul Fischer. Finite-time analysis of the multi-armed bandit problem. *Machine learning*, 47(2-3):235–256, 2002.

Peter Auer, Thomas Jaksch, and Ronald Ortner. Near-optimal regret bounds for reinforcement learning. In *Advances in neural information processing systems*, pages 89–96, 2009.

Baruch Awerbuch and Robert Kleinberg. Competitive collaborative learning. *Journal of Computer and System Sciences*, 74(8):1271–1288, 2008.

Carlo Baldassi. Generalization learning in a perceptron with binary synapses. *Journal of Statistical Physics*, 136(5):902–916, 2009. doi: 10.1007/s10955-009-9822-1.

Carlo Baldassi and Alfredo Braunstein. A max-sum algorithm for training discrete neural networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2015(8):P08008, 2015. doi: 10.1088/1742-5468/2015/08/P08008.

Carlo Baldassi, Alfredo Braunstein, Nicolas Brunel, and Riccardo Zecchina. Efficient supervised learning in networks with binary synapses. *Proceedings of the National Academy of Sciences of the United States of America*, 104(26):11079–1084, 2007. doi: 10.1073/pnas.0700324104.

Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Subdominant dense clusters allow for simple learning and high computational performance in neural networks with discrete synapses. *Phys. Rev. Lett.*, 115: 128101, Sep 2015. doi: 10.1103/PhysRevLett.115.128101.

Carlo Baldassi, Christian Borgs, Jennifer T. Chayes, Alessandro Ingrosso, Carlo Lucibello,

Luca Saglietti, and Riccardo Zecchina. Unreasonable effectiveness of learning neural networks: From accessible states and robust ensembles to basic algorithmic schemes. *Proceedings of the National Academy of Sciences*, 113(48):E7655–E7662, November 2016a. ISSN 0027-8424, 1091-6490. doi: 10.1073/pnas.1608103113.

Carlo Baldassi, Alessandro Ingrosso, Carlo Lucibello, Luca Saglietti, and Riccardo Zecchina. Local entropy as a measure for sampling solutions in constraint satisfaction problems. *Journal of Statistical Mechanics: Theory and Experiment*, 2016(2):P023301, February 2016b. ISSN 1742-5468. doi: 10.1088/1742-5468/2016/02/023301.

Carlo Baldassi, Enrico M Malatesta, and Riccardo Zecchina. Properties of the geometry of solutions and capacity of multilayer neural networks with rectified linear unit activations. *Physical Review Letters*, 123(17):170602, 2019. doi: 10.1103/PhysRevLett.123.170602.

Carlo Baldassi, Riccardo Della Vecchia, Carlo Lucibello, and Riccardo Zecchina. Clustering of solutions in the symmetric binary perceptron. *Journal of Statistical Mechanics: Theory and Experiment*, 2020(7):073303, 2020a.

Carlo Baldassi, Fabrizio Pittorino, and Riccardo Zecchina. Shaping the learning landscape in neural networks around wide flat minima. *Proceedings of the National Academy of Sciences*, 117(1):161–170, 2020b. doi: 10.1073/pnas.1908636117.

Gábor Bartók. A near-optimal algorithm for finite partial-monitoring games against adversarial opponents. In *Conference on Learning Theory*, pages 696–710, 2013.

Alfredo Braunstein and Riccardo Zecchina. Learning by message passing in networks of discrete synapses. *Physical Review Letters*, 96:030201, Jan 2006. doi: 10.1103/PhysRevLett.96.030201.

Alfredo Braunstein, Marc Mézard, and Riccardo Zecchina. Survey propagation: An algorithm for satisfiability. *Random Structures & Algorithms*, 27(2):201–226, 2005.

Sébastien Bubeck, Nicolo Cesa-Bianchi, et al. Regret analysis of stochastic and nonstochastic multi-armed bandit problems. *Foundations and Trends® in Machine Learning*, 5(1):1–122, 2012.

Nicolo Cesa-Bianchi and Gábor Lugosi. Combinatorial bandits. *Journal of Computer and System Sciences*, 78(5):1404–1422, 2012.

Nicolò Cesa-Bianchi, Tommaso R Cesari, and Claire Monteleoni. Cooperative online learning: Keeping your neighbors updated. *arXiv preprint arXiv:1901.08082*, 2019a.

Nicolo Cesa-Bianchi, Claudio Gentile, and Yishay Mansour. Delay and cooperation in nonstochastic bandits. *The Journal of Machine Learning Research*, 20(1):613–650, 2019b.

Nicolò Cesa-Bianchi, Tommaso Cesari, and Riccardo Della Vecchia. Online cooperative learning with broadcasting and delays. *(in preparation)*.

Olivier Chapelle, Eren Manavoglu, and Romer Rosales. Simple and scalable response prediction for display advertising. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 5(4):1–34, 2014.

Pratik Chaudhari, Anna Choromanska, Stefano Soatto, Yann LeCun, Carlo Baldassi, Christian Borgs, Jennifer Chayes, Levent Sagun, and Riccardo Zecchina. Entropy-sgd:

Biasing gradient descent into wide valleys. *Journal of Statistical Mechanics: Theory and Experiment*, 2019(12):124018, 2019.

Hervé Daudé, Marc Mézard, Thierry Mora, and Riccardo Zecchina. Pairs of sat-assignments in random boolean formulæ. *Theoretical Computer Science*, 393(1):260–279, 2008. ISSN 0304-3975. doi: 10.1016/j.tcs.2008.01.005.

Riccardo Della Vecchia and Tommaso Cesari. An efficient algorithm for cooperative semi-bandits. *arXiv preprint arXiv:2010.01818*, 2020.

Thomas Desautels, Andreas Krause, and Joel W Burdick. Parallelizing exploration-exploitation tradeoffs in gaussian process bandit optimization. *Journal of Machine Learning Research*, 15:3873–3923, 2014.

Jian Ding and Nike Sun. Capacity lower bound for the ising perceptron. In *Proceedings of the 51st Annual ACM SIGACT Symposium on Theory of Computing*, pages 816–827. ACM, 2019. doi: 10.1145/3313276.3316383.

Andreas Engel and Christian Van den Broeck. *Statistical mechanics of learning*. Cambridge University Press, 2001.

Elizabeth Gardner and Bernard Derrida. Optimal storage properties of neural network models. *Journal of Physics A: Mathematical and General*, 21(1):271–284, jan 1988. doi: 10.1088/0305-4470/21/1/031.

Martin Grötschel, László Lovász, and Alexander Schrijver. *Geometric algorithms and combinatorial optimization*, volume 2. Springer Science & Business Media, 2012.

James Hannan. Approximation to bayes risk in repeated play. *Contributions to the Theory of Games*, 3:97–139, 1957.

Elad Hazan. Introduction to online convex optimization. *arXiv preprint arXiv:1909.05207*, 2019.

Heinz Horner. Dynamics of learning for the binary perceptron problem. *Zeitschrift für Physik B Condensed Matter*, 86(2):291–308, 1992. doi: 10.1007/BF01313839.

Haiping Huang and Yoshiyuki Kabashima. Origin of the computational hardness for learning with binary synapses. *Physical Review E*, 90(5):052813, 2014. doi: 10.1103/PhysRevE.90.052813.

Haiping Huang, K. Y. Michael Wong, and Yoshiyuki Kabashima. Entropy landscape of solutions in the binary perceptron problem. *Journal of Physics A: Mathematical and Theoretical*, 46(37):375002, aug 2013. doi: 10.1088/1751-8113/46/37/375002.

Pooria Joulani, Andras Gyorgy, and Csaba Szepesvári. Delay-tolerant online convex optimization: Unified analysis and adaptive-gradient algorithms. In *Thirtieth AAAI Conference on Artificial Intelligence*, 2016.

Adam Kalai and Santosh Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

Wouter M Koolen, Manfred K Warmuth, Jyrki Kivinen, et al. Hedging structured concepts. In *COLT*, pages 93–105. Citeseer, 2010.

Werner Krauth and Marc Mézard. Storage capacity of memory networks with binary couplings. *Journal de Physique*, 50(20):3057–3066, 1989. doi: 10.1051/jphys:0198900500200305700.

Florent Krzaka…a, Andrea Montanari, Federico Ricci-Tersenghi, Guilhem Semerjian, and Lenka Zdeborová. Gibbs states and the set of solutions of random constraint satisfaction problems. *Proceedings of the National Academy of Sciences*, 104(25):10318–10323, 2007.

Frank R Kschischang, Brendan J Frey, and H-A Loeliger. Factor graphs and the sum-product algorithm. *IEEE Transactions on information theory*, 47(2):498–519, 2001.

Branislav Kveton, Zheng Wen, Azin Ashkan, and Csaba Szepesvari. Combinatorial cascading bandits. In *Advances in Neural Information Processing Systems*, pages 1450–1458, 2015.

Tor Lattimore and Csaba Szepesvári. Bandit algorithms. *preprint*, page 28, 2018.

Tor Lattimore and Csaba Szepesvári. *Bandit algorithms*. Cambridge University Press, 2020.

Tor Lattimore, Branislav Kveton, Shuai Li, and Csaba Szepesvari. Toprank: A practical algorithm for online stochastic ranking. In *Advances in Neural Information Processing Systems*, pages 3945–3954, 2018.

Yann LeCun, Yoshua Bengio, and Geoffrey Hinton. Deep learning. *nature*, 521(7553):436, 2015.

Yin Tat Lee, Aaron Sidford, and Santosh S Vempala. Efficient convex optimization with membership oracles. In *Conference On Learning Theory*, pages 1292–1294. PMLR, 2018.

Lihong Li, Wei Chu, John Langford, and Robert E Schapire. A contextual-bandit approach to personalized news article recommendation. In *Proceedings of the 19th international conference on World wide web*, pages 661–670, 2010.

Brendan McMahan and Matthew Streeter. Delay-tolerant algorithms for asynchronous distributed online learning. In *Advances in Neural Information Processing Systems*, pages 2915–2923, 2014.

Marc Mezard and Andrea Montanari. *Information, physics, and computation*. Oxford University Press, 2009.

Marc Mézard, Giorgio Parisi, and Miguel Virasoro. *Spin glass theory and beyond: An Introduction to the Replica Method and Its Applications*, volume 9. World Scientific Publishing Company, 1987.

Marc Mézard, Giorgio Parisi, and Riccardo Zecchina. Analytic and algorithmic solution of random satisfiability problems. *Science*, 297(5582):812–815, 2002.

Marc Mézard, Thierry Mora, and Riccardo Zecchina. Clustering of solutions in the random satisfiability problem. *Physical Review Letters*, 94:197205, May 2005. doi: 10.1103/PhysRevLett.94.197205.

Gergely Neu and Gábor Bartók. An efficient algorithm for learning with semi-bandit

feedback. In *International Conference on Algorithmic Learning Theory*, pages 234–248. Springer, 2013.

Francesco Orabona. A modern introduction to online learning. *arXiv preprint arXiv:1912.13213*, 2019.

Ciara Pike-Burke, Shipra Agrawal, Csaba Szepesvari, and Steffen Grunewalder. Bandits with delayed, aggregated anonymous feedback. In *International Conference on Machine Learning*, pages 4105–4113, 2018.

Fabrizio Pittorino, Carlo Lucibello, Christoph Feinauer, Enrico M Malatesta, Gabriele Perugini, Carlo Baldassi, Matteo Negri, Elizaveta Demyanenko, and Riccardo Zecchina. Entropic gradient descent algorithms and wide flat minima. *arXiv preprint arXiv:2006.07897*, 2020.

Levent Sagun, Leon Bottou, and Yann LeCun. Eigenvalues of the hessian in deep learning: Singularity and beyond. *arXiv preprint arXiv:1611.07476*, 2016.

Hyunjune Sebastian Seung, Haim Sompolinsky, and Naftali Tishby. Statistical mechanics of learning from examples. *Physical Review A*, 45:6056–6091, Apr 1992. doi: 10.1103/PhysRevA.45.6056.

Shai Shalev-Shwartz and Shai Ben-David. *Understanding machine learning: From theory to algorithms*. Cambridge university press, 2014.

David Silver, Aja Huang, Chris J Maddison, Arthur Guez, Laurent Sifre, George Van Den Driessche, Julian Schrittwieser, Ioannis Antonoglou, Veda Panneershelvam, Marc Lanctot, et al. Mastering the game of go with deep neural networks and tree search. *nature*, 529(7587):484–489, 2016.

Daiki Suehiro, Kohei Hatano, Shuji Kijima, Eiji Takimoto, and Kiyohito Nagano. Online prediction under submodular constraints. In *International Conference on Algorithmic Learning Theory*, pages 260–274. Springer, 2012.

William R Thompson. On the likelihood that one unknown probability exceeds another in view of the evidence of two samples. *Biometrika*, 25(3/4):285–294, 1933.

Timothy L. H. Watkin, Albrecht Rau, and Michael Biehl. The statistical mechanics of learning a rule. *Reviews of Modern Physics*, 65:499–556, Apr 1993. doi: 10.1103/RevModPhys.65.499.

Marcelo J Weinberger and Erik Ordentlich. On delayed prediction of individual sequences. *IEEE Transactions on Information Theory*, 48(7):1959–1976, 2002.

Sixin Zhang, Anna E Choromanska, and Yann LeCun. Deep learning with elastic averaging sgd. In *Advances in Neural Information Processing Systems*, pages 685–693, 2015.

Julian Zimmert and Yevgeny Seldin. An optimal algorithm for adversarial bandits with arbitrary delays. *arXiv preprint arXiv:1910.06054*, 2019.