Analysis of Perturbation Techniques in Online Learning

by

Chansoo Lee

A dissertation submitted in partial fulfillment
of the requirements for the degree of
Doctor of Philosophy
(Computer Science and Engineering)
in the University of Michigan
2018

Doctoral Committee:

Assistant Professor Jacob Abernethy, Co-Chair
Associate Professor Ambuj Tewari, Co-Chair
Professor Satinder Singh
Professor Pascal Van Hentenryck

Chansoo Lee
chansool@umich.edu
ORCID iD: 0000-0003-0802-5301

For my wife, Jiwon

ACKNOWLEDGMENTS

I was incredibly fortunate to have two great advisors, Jake Abernethy and Ambuj Tewari. They are not only amazingly smart but also great people who are always caring and cheerful. This dissertation owes to their intellectual and emotional support. I also am grateful to Satinder Singh and Pascal Van Hentenryck for serving on my committee.

I met a lot of brilliant people during my Ph.D. and had an opportunity to work with some of them. This dissertation includes my joint work with Jake Abernethy, Audra McMillan and Abhinav Sinha, and Ambuj Tewari. I would like to thank all my co-authors on other papers: Ferdinando Fioretto, Pascal Van Hentenryck, Satyen Kale, and David Pál.

I would not have survived harsh winters in Ann Arbor without my friends and family. No matter how far apart we are and how often we meet, I am always grateful to you. I would like to also thank Lulu for always letting me pet her soft curly fur. Most importantly, Jiwon, this thesis is the result of our joint endeavor.

TABLE OF CONTENTS

LIST OF APPENDICES

Appendix

LIST OF SYMBOLS

Symbol	Meaning
$[N]$	Set $\{1, 2, \ldots, N\}$
$x_{1:t}$	Sequence $x_1, \ldots, x_t$
Δ^N	Probability simplex with N vertices
$(\|\cdot\|, \|\cdot\|_\star)$	Arbitrary pair of dual norms
$\|\cdot\|_p$	L_p-norm, defined as $\|x\|_p = (\sum_{i=1}^N x_i^p)^{\frac{1}{p}}$
$\mathcal{B}_r(\mathcal{S}, \|\cdot\|)$	$\{x \in \mathcal{S} : \|x\| \leq r\}$, Ball of radius r centered at zero.
$\mathcal{B}_r^+(\mathcal{S}, \|\cdot\|)$	$\{x \in \mathcal{B}_r(\mathcal{S}, \|\cdot\|) : x_1, \ldots, x_N \geq 0\}$
$\mathcal{B}_r^-(\mathcal{S}, \|\cdot\|)$	$\{x \in \mathcal{B}_r(\mathcal{S}, \|\cdot\|) : x_1, \ldots, x_N \leq 0\}$
$\mathrm{dom}(f)$	Domain of function f
$f^\star$	Convex conjugate of f
f'	Derivative of f (if f is a scalar-valued function)
∇f	Gradient of f
$\nabla^2 f$	Hessian of f
$D_f(x, y)$	Bregman divergence of f for points x, y
$\tilde{f}(x; \mathcal{D})$	Stochastic smoothing of a function f with distribution $\mathcal{D}$
$\mathcal{N}(\mu, \Sigma)$	Multivariate normal with mean μ and covariance matrix Σ
$\mathrm{Lap}(\lambda)$	Zero-mean Laplace distribution with scaling parameter λ
$\mathrm{Exponential}(\lambda)$	Exponential distribution with scaling parameter λ
$\mathcal{D}^N$	Distribution over $\mathbb{R}^N$ with i.i.d. samples from $\mathcal{D}$ over $\mathbb{R}$
$\mu_\mathcal{D}$	The probability density function of $\mathcal{D}$
$\varphi_\mathcal{D}$	The cumulative density function of $\mathcal{D}$
$\mathrm{haz}_\mathcal{D}(x)$	The hazard rate of a distribution $\mathcal{D}$ evaluated at x

Symbol	Meaning
$\mathrm{Tr}(M)$	The trace of a matrix M.
$\mathrm{diag}(M)$	The diagonal elements of a matrix M as a vector
$\lambda(M)$	Vector of eigenvalues of M in decreasing order
S^{N}	Set of all N-by-N symmetric matrices
$\mathrm{S}_{+}^{\mathrm{N}}$	Set of all N-by-N positive semidefinite matrices
$\mathrm{S}_{+}^{\mathrm{N}}(k)$	k-fantope, the subset of $\mathrm{S}_{+}^{\mathrm{N}}$ comprised of matrices whose eigenvalues sum up to k.
Γ	The Gamma function

ABSTRACT

The most commonly used regularization technique in machine learning is to directly add a penalty function to the optimization objective. For example, L_2 regularization is universally applied to a wide range of models including linear regression and neural networks. The alternative regularization technique, which has become essential in modern applications of machine learning, is implicit regularization by injecting random noise into the training data.

In fact, this idea of using random perturbations as regularizer has been one of the first algorithms for online learning, where a learner chooses actions iteratively on a data sequence that may be designed adversarially to thwart learning process. One such classical algorithm is known as Follow The Perturbed Leader (FTPL).

This dissertation presents new interpretations of FTPL. In the first part, we show that FTPL is equivalent to playing the gradients of a stochastically smoothed potential function in the dual space. In the second part, we show that FTPL is the extension of a differentially private mechanism that has inherent stability guarantees. These perspectives lead to novel frameworks for FTPL regret analysis, which not only prove strong performance guarantees but also help characterize the optimal choice of noise distributions. Furthermore, they extend to the partial information setting where the learner observes only part of the input data.

CHAPTER I

Introduction

In this thesis, we study the problem of *online learning*, where a learner iteratively plays a sequence of actions based on the data received up to the previous iteration. The learner's goal is to *minimize the regret*, defined as the difference between the learner's loss and the loss of the best fixed action in hindsight. In developing online learning algorithms, we consider an *adversarial* environment where we do not make any stochastic assumptions about the sequence of data. The learners goal in this case is to minimize the *worst-case regret*.

The simplest heuristic for this setting is playing the optimal action on the observed sequence of the data up to the previous iteration. This algorithm, called Follow the Leader (FTL) and formally defined in Section 2.2.1, is identical to Empirical Risk Minimization (ERM), which has nice generalization guarantees when data is always an i.i.d. sample from an unknown distribution. In the non-stochastic setting that we study, however, FTL algorithm suffers a constant worst-case regret on the new data point no matter how much data it has received. The problem is that FTL algorithm overfits to the observed data, which may be adversarially designed to have large fluctuations. The key element in developing optimal algorithms is to avoid overfitting and to induce stability by *regularization*.

A standard regularization technique in machine learning is *regularization via penalty*, which is to directly add a penalty function (called *regularizer*) to the optimization objective. A popular method in this category is the L_2 regularization, which is universally applied to a wide range of models including linear regression and multi-layer convolutional neural networks.

Follow the Regularized Leader (FTRL) algorithm is the application of this technique to the online setting. This algorithm is very well understood thanks to the powerful convex

analysis tools; the regret analysis reduces to the analysis of the second-order behavior of the regularizer (Shalev-Shwartz 2012). Srebro et al. (2011) proved that Mirror Descent, which is equivalent to FTRL under some assumptions (McMahan 2011) achieves a nearly optimal regret guarantee for a general class of online learning problems,.

The alternative regularization technique, which has become essential in modern applications of machine learning, is implicit *regularization via perturbations* by injecting random noise into the training data. This method has been the main driving force behind successful applications of complex deep learning architectures, with the most notable example of dropout by Hinton et al. (2012).

Interestingly, one of the earliest online learning algorithm by Hannan (1957) uses a regularization via perturbations, and its extension was named Follow the Perturbed Leader (FTPL) by Kalai and Vempala (2005). Due to the stochastic nature of these techniques, however, it is difficult to analyze their behavior. The FTPL analysis relies substantially on clever algebra tricks and heavy probabilistic analysis (Devroye et al. 2013; Erven et al. 2014; Kalai and Vempala 2005). These results are unsatisfying because these techniques do not generalize among different distributions and thus fail to provide intuitions on what are the core properties of the noise distributions that lead to optimal regret guarantees.

This thesis presents two new interpretations of FTPL, each of which leads to a new framework for analyzing FTPL regret. The first interpretation in Chapter III is based on convex duality. The key observation is that FTL, FTPL, and FTRL all belong to the same family of algorithms that play the *gradient* of a potential function. This connection is not a surprising fact; the gradient is the direction of the *maximal* rate of increase, and FTL-family algorithms all plays an action that *maximizes* an objective function. In this framework, FTPL naturally arises as a *smoothing operation* of a non-smooth potential function. The FTPL regret analysis now boils down to understanding the second-order properties of the smoothed function, in the same way that FTRL is analyzed. Indeed, we show that FTPL implicitly defines a strongly convex regularizer via convex duality (Section 3.2.4).

This interpretation leads to a generic analysis framework for FTPL, especially for the case where the noise distribution belongs to the exponential family. We obtain powerful general-case regret guarantees effortlessly by directly applying the results from the optimization literature (Section 3.3). By a more careful analysis, we prove that FTPL with the Gaussian distribution is minimax-optimal for canonical online learning problems (Section 3.4 and 3.5). Our analysis technique extends to the multi-armed adversarial bandit setting (Section 3.7), proving that the *hazard rate* of the noise distribution plays a key role

in the regret analysis.

The second interpretation in Chapter IV is that FTPL achieves *differential privacy* (DP) (Dwork and Roth 2014). DP requires that the output distribution of a randomized algorithm remains mostly identical given a small change in the input. This implies that a DP algorithm, when applied to the online learning problem, naturally avoids overfitting to any individual data point. Furthermore, DP is a *multiplicative* guarantee on the stability, which natural leads to first-order regret bounds that can take advantage of easy problem instances.

Given a DP algorithm, the regret analysis reduces to controlling the accuracy lost in order to achieve privacy. In fact, differentially private algorithms are already developed to obtain a favorable tradeoff between privacy and accuracy. We can directly use the existing tools in DP literature, such as Gaussian mechanism, to obtain strong general-case regret guarantees (Section 4.2). We establish that DP is also closely connected to the hazard rate of a distribution. Based on this observation, we prove that the hazard rate is the key to not only low-regret algorithms for the multi-armed bandits (Section 4.5), but also minimax-optimal algorithms for the experts setting (Section 4.3).

CHAPTER II

Online Linear Optimization

2.1 Problem Definition

The online linear optimization (OLO) is defined as a repeated game between two entities that we call the *learner* and the *adversary*. An instance of OLO is specified by two convex and closed subsets of $\mathbb{R}^N$ that define possible actions for learner and adversary, respectively. We assume an *oblivious* adversary that chooses the whole sequence of moves ahead of time. The learner is allowed access to its private source of randomness in making its moves.

We present two equivalent formulations of OLO, one as a reward maximization problem and the other as a loss minimization problem.

2.1.1 Reward Formulation

On round $t = 1, \ldots, T$,

- the learner plays an *action* $x_t \in \mathcal{X}$;
- the adversary reveals a *reward* vector $g_t \in \mathcal{Y}$;
- the learner receives a *linear reward* $\langle x_t, g_t \rangle$.

We say $\mathcal{X}$ is the *decision set* and $\mathcal{Y}$ is the *reward set*. Let $G_t = \sum_{s=1}^{t} g_s$ be the cumulative reward. The learner's goal is to minimize the *expected* regret, defined as:

$$\mathbb{E}\mathsf{Regret}_T = \max_{x \in \mathcal{X}} \langle x, g_T \rangle - \mathbb{E}\left[\sum_{t=1}^{T} \langle x_t, g_t \rangle\right] \tag{2.1}$$

where the expectations are taken over the learner's randomness. In the *gain-only setting*, rewards are always non-negative: $\min_{x \in \mathcal{X}} \langle x, g_t \rangle \geq 0$. In the *loss-only setting*, rewards are always non-positive: $\min_{x \in \mathcal{X}} \langle x, g_t \rangle \leq 0$. In the *loss/gain setting*, rewards can be positive or negative.

2.1.2 Loss Formulation

We can equivalently define OLO in terms of losses as follows. On round $t = 1, \ldots, T$,

- the learner plays an *action* $x_t \in \mathcal{X}$;
- the adversary reveals a *loss* vector $\ell_t \in \mathcal{Y}$;
- the learner receives a *linear loss* $\langle x_t, \ell_t \rangle$.

Let $L_t = \sum_{s=1}^{t} \ell_s$ be the cumulative loss. The learner's goal is to minimize the *expected* regret, defined as:

$$\mathbb{E}\mathsf{Regret}_T = \mathbb{E}\left[\sum_{t=1}^{T} \langle x_t, \ell_t \rangle\right] - \min_{x \in \mathcal{X}} \langle x, L_T \rangle \tag{2.2}$$

where the expectations are taken over the learner's randomness. We use L_T^* to denote the comparator term in the regret definition, i.e., the best loss in hindsight. In the *loss-only setting*, losses are always positive: $\min_{x \in \mathcal{X}} \langle x, \ell_t \rangle \geq 0$ for all t. In the *loss/gain setting*, losses can be positive or negative.

2.1.3 Comparison

Note that it is easy to switch between the reward and loss formulations by simply flipping the signs of the adversary's moves. We use the reward formulation by default up to Chapter III , so that we can directly analyze the convex function $\max(\cdot)$ without cumbersome sign changes.

We switch to the loss formulation in Chapter IV where we focus on proving *first-order regret bounds* that grow in L_T^* instead of T for the loss-only setting; such bounds give a very strong guarantee that the algorithm is able to benefit from the existence of one "good" action that suffers little loss. Note that its counterpart for the gain-only setting, namely a regret bound that grows in the maximum gain in hindsight, does not have the same implication and is not commonly studied in the literature. For the full comparison of the *gain-only* and *loss-only* settings, please see (Kwon and Perchet 2016).

2.2 Algorithms

This dissertation focuses on a family of online linear optimization algorithms that select their actions by solving an optimization problem. The only interface these algorithms have to the data is the cumulative reward up to the previous time step; that is, we can write $x_t = \arg\max_{x \in \mathcal{X}} f(x; G_{t-1})$ for some objective function f.

2.2.1 Follow the Leader

The most simple algorithm, *Follow the Leader* (FTL), does not incorporate any perturbation or regularization into the optimization, and uses the objective $f(x; G) = \langle x, G \rangle$. Unfortunately FTL does not enjoy non-trivial regret guarantees due to the inherent instability of linear optimization. Even small changes in the input can lead to large fluctuations in the optimal solution.

2.2.2 Follow the Regularized Leader

Follow the Regularized Leader (FTRL) uses the regularized objective function

$$f(w; G) = \langle x, G \rangle - \mathcal{R}(x)$$

where $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is a convex regularizer. The FTRL regret analysis involves a fictitious algorithm called *Be the Regularized Leader* (BTRL), which looks at the loss vector one step ahead and plays what FTRL would play at the next time step.

Theorem 2.1. *The FTRL with regularizer $\mathcal{R}$ has the regret bound:*

$$\mathsf{Regret}(\mathsf{FTRL})_T \leq \underbrace{\sup_{x \in \mathcal{X}} \mathcal{R}(x)}_{\textit{BTRL regret}} + \underbrace{\sum_{t=1}^{T} \langle x_{t+1} - x_t, g_t \rangle}_{\textit{stability term}}$$

The first term is the regret of BTRL, which only suffers the difference introduced by $\mathcal{R}(x)$. The second term is the regret due to instability of the prediction.

2.2.3 Follow the Perturbed Leader

Follow the Perturbed Leader (FTPL) sets $f(x, G) = \langle x, G + z \rangle$ where z is a random vector from a distribution $\mathcal{D}$. Since losses are always linear in the learner's action, the

expected regret of FTPL is equal to the regret of the *expected* version of FTPL algorithm, which plays

$$x_t = \mathbb{E}_{z\sim\mathcal{D}}\left[\operatorname*{argmin}_{x\in\mathcal{X}}\langle x, G_{t-1}+z\rangle\right].$$

Also note that since we assume an oblivious adversary that does not adapt to learner's randomness, only a single sample of z is required.

Similarly to the FTRL, the FTPL regret analysis involves a fictitious algorithm called *Be the Perturbed Leader* (BTPL), which looks at the loss vector one step ahead and plays x_{t+1}^{FTPL} at round t. By inductive argument, (Kalai and Vempala 2005) shows that BTPL suffers a small regret that does not grow in T but only in the magnitude of the noise and the size of $\mathcal{X}$:

$$\begin{aligned}\mathbb{E}[\mathsf{Regret}(\mathsf{BTPL})_T] &\le \mathbb{E}_{z\sim\mathcal{D}}[\sup_{x\in\mathcal{X}}\langle x, z\rangle] \\ &\le \|\mathcal{X}\|\mathbb{E}_{z\sim\mathcal{D}}[\|z\|_*]. \end{aligned} \tag{2.3}$$

Thus, we get the counterpart of Theorem 2.1:

Theorem 2.2. *The FTPL with distribution $\mathcal{D}$ has the regret bound:*

$$\mathbb{E}[\mathsf{Regret}(\mathsf{FTPL})_T] \le \underbrace{\|\mathcal{X}\|\mathbb{E}_{z\sim\mathcal{D}}[\|z\|_*]}_{\textit{BTPL regret}} + \underbrace{\mathbb{E}\left[\sum_{t=1}^{T}\langle x_{t+1}-x_t, g_t\rangle\right]}_{\textit{stability term}}$$

for any arbitrary dual norm pairs $(\|\cdot\|, \|\cdot\|_\star)$.

The stability term, however, is extremely hard to analyze in this case due to the stochastic nature of the algorithm. Much of the existing literature focuses on directly analyzing the stability term by clever algebra tricks that do not generalize across different distributions. For example, exponential (Kalai and Vempala 2005), random-walk (Devroye et al. 2013), and dropout (Erven et al. 2014) noise have been shown to achieve low regret for the experts problem (defined in Section 2.3.1), all using distribution-specific arguments.

In contrast, the analysis techniques presented in this dissertation are generically applicable to a wide range of distributions. In fact, our main results, such as Theorem 4.15, *characterize* a family of optimal distributions.

Prior to this work, Rakhlin et al. (2012) developed the first generic analysis framework

for FTPL. They view FTPL as the minimax strategy against a random approximation of the worst-case adversary and offered some intuitions on what distributions are optimal for different settings. We compare their results to ours throughout the paper wherever applicable.

2.3 Canonical Problems

2.3.1 Experts Problem

The experts problem is an instance of OLO where $\mathcal{X} = \Delta^N$ and $\mathcal{Y} = \mathcal{B}_1(\mathbb{R}^N, \|\cdot\|_\infty)$, the unit ball in the ℓ_∞-norm. The use of the term *expert* originated from Littlestone and Warmuth (1994)'s formulation of the problem that the learner iteratively updates a belief distribution over a set of experts, each of whom makes a prediction on every round.

The minimax regret is $\sqrt{(T/2)\log N}$ for the loss/gain setting (Cesa-Bianchi and Lugosi 2006, Chapter 7). For the loss-only setting, Freund and Schapire (1997) proved that the Weighted Majority algorithm achieves the first-order regret bound of $\sqrt{2L_T^* \log N} + \log N$, which is asymptotically minimax optimal (Vovk 1998).

Rakhlin et al. (2012) showed that FTPL with any symmetric noise distribution, with a proper scaling, achieves the optimal $O(\sqrt{T \log N})$ regret for the loss/gain setting. In Section 3.5, we prove $O(\sqrt{\sum_{t=1}^T \|g_t\|_\infty^2 \log N})$ bounds for several exponential family distributions. Although the two bounds are asymptotically equivalent in the worst case where $\|g_t\|_\infty = 1$, our bound does not require the knowledge of T in advance and is stronger against a sequence of small gain vectors. This is due to the fact that the game-theoretic analysis framework Rakhlin et al. (2012) must reason recursively from the last step, assuming the worst case in each step.

For the loss-only setting, Kalai and Vempala (2005) used the negative exponential noise and Erven et al. (2014) used the dropout noise to achieve the optimal $O(\sqrt{L_T^* \log N} + \log N)$ regret. In Section 4.3, we prove a generic sufficient condition on the noise distribution for this optimal regret bound.

2.3.2 Euclidean Balls Problem

The Euclidean balls problem is an instance of OLO where $\mathcal{X} = \mathcal{Y} = \mathcal{B}_r(\mathbb{R}^N, \|\cdot\|_2)$. In this dissertation, we only consider the *unit Euclidean balls* problem, where $r = 1$. Abernethy et al. (2008) showed that the minimax optimal regret (under the gain formulation)

is precisely $\frac{1}{2}\sqrt{\sum_{t=1}^{T} \|g_t\|_2^2}$. Note that this minimax regret has no explicit dependence on the dimension N. The regret bound of the same order can be achieved by FTRL with the Euclidean norm as regularizer, also known as Online Gradient Descent (Zinkevich 2003).

The optimal regret bound with FTPL was unknown for a long time until Rakhlin et al. (2012, Lemma 9) proved that FTPL with uniform distribution on the surface of the unit sphere has a regret at most $4\sqrt{2T}$. We prove in Section 3.4 that FTPL with i.i.d. Gaussian noise has a regret at most $2\sqrt{1 + \sum_{t=1}^{T} \|g_t\|_2^2}$, closely matching the minimax regret.

2.3.3 Online PCA

The OLO framework naturally extends to the space of symmetric matrices. For $A, B \in \mathrm{S}^{\mathrm{N}^N}$, the *matrix inner product* is the dot product between matrices flattened to vectors:

$$\langle A, B \rangle = \mathrm{Tr}(AB) = \sum_{i,j=1}^{N} A_{ij} B_{ij}.$$

To motivate the Online PCA problem, first consider the online data compression problem defined as follows. Let P_t be a rank k-matrix, which is a low-rank approximation of a (potentially full-rank) covariance matrix Σ_t. We define the *compression loss*, the loss in precision due to the approximation, to be

$$\mathrm{Tr}((I - P_t)\Sigma_t) = \mathrm{Tr}(\Sigma_t) - \mathrm{Tr}(P_t \Sigma_t).$$

Since the first term is independent of the learner's action, we can equivalently say that the learner gains reward of $\mathrm{Tr}(P_t \Sigma_t)$ every round for faithfully preserving the data. The Online PCA is an abstraction of this problem in which Σ_t is no longer restricted to a valid covariance matrix.

As a formal definition, *Online k-Dense PCA* is an instance of OLO where $\mathcal{X} = \{A \in \mathrm{S}_+^{\mathrm{N}} : \|\lambda(A)\|_1 \leq k\}$ and $\mathcal{Y} = \{A \in \mathrm{S}^{\mathrm{N}} : \|\lambda(A)\|_\infty \leq 1\}$. The minimax regret is $O(k\sqrt{T \log(N/k)})$ for the loss/gain setting, and $O(\sqrt{L^* k \log \frac{N}{k}} + k \log \frac{N}{k})$ for the loss-only setting (Nie et al. 2013). Both bounds are achieved by FTRL with Von Neumann entropy, which is also known as Online Matrix Exponentiated Gradient algorithm (Nie et al. 2013). In this dissertation, we only consider the Online 1-Dense PCA problem, and simply call it Online Dense PCA.

An interesting fact is that the minimax regret for Online Dense PCA is independent

of whether the adversary may change the eigensystems between iterations or not. The Online Dense PCA and the experts problem have the same minimax regret, even though the reduction only works in one direction from experts problem to Online Dense PCA.

The *Online k-Sparse PCA* problem is a special case of Online Dense PCA where $\mathcal{Y} = \{aa^\top : a \in \mathbb{R}^N, \|a\|_2 = 1\}$, restricted to rank-1 matrices. The minimax optimal regret for this problem is $O(\sqrt{Tk \log \frac{N}{k}})$. That is, for $k = 1$, the sparsity assumption on the reward matrices does not change the problem complexity.

Whether there exists a computationally efficient algorithm that achieves the minimax optimal regret without requiring a full eigendecomposition every round had been a long-standing open question posed by Warmuth and Kuzmin (2010). Allen-Zhu and Li (2017) recently solved this problem by reducing the effective dimensionality of the matrix problem to dimension 3, but the question still remains whether it can be solved using FTPL, which only requires the computation of maximum eigenvector each round. The best known regret bound using FTPL is $O(\sqrt{TN})$ for the dense case (Garber et al. 2015; Kotłowski and Warmuth 2015), and $O(\sqrt[4]{N}\sqrt{kL_T^* \log T})$ for the sparse case (Dwork et al. 2014).

For Online Dense PCA, we show that there is an FTPL algorithm that achieves $O(\sqrt[4]{N}\sqrt{L_T^* \log T})$ regret, which is generally an improvement over the best-known FTPL bound of $O(\sqrt{TN})$ regret. For the Online Sparse PCA, we show that there is a simple FTPL algorithm that achieves the optimal regret, partially resolving the open problem.

2.4 Adversarial Multi-Armed Bandits

Adversarial Multi-Armed Bandits (MAB) problem is a *partial information* variant of the loss-only experts problem. The two main differences are: (a) the learner is required to *sample* an action $i_t \in \{1, \ldots, N\}$ according to a chosen probability vector $p_t \in \Delta^N$, and (b) the learner observes only the scalar ℓ_{t,i_t} and receives no information regarding the losses/gains for the other coordinates of ℓ. The limited feedback is what makes MAB significantly more challenging than the vanilla experts problem.

The MAB problem is useful for a wide range of applications including medical experiment design (Gittins 1996), automated poker playing strategies (Van den Broeck et al. 2009), and hyperparameter tuning (Pacula et al. 2012). For the survey of work on MAB, see the summary paper by Bubeck and Cesa-Bianchi (2012).

The minimax regret for this problem is $O(\sqrt{NT})$ (Bubeck et al. 2012), and it is also

possible to achieve the first-order regret of $O(\sqrt{NL_T^* \log N})$ (Allenberg et al. 2006; Neu 2015). In Section 3.7, we show that FTPL with a wide variety of distributions enjoys nearly optimal regret of $O(\sqrt{TN \log N})$. In Section 4.5, we show a similar statement for the first-order regret bound.

2.4.1 Connection to the Experts Problem

There is a standard reduction from adversarial MAB to the experts problem. Assume we have an algorithm $\mathcal{A}$, which outputs a probability vector $x_t \in \Delta^N$ given a sequence of (potentially unbounded) loss vectors; for example, $\mathcal{A}$ can be any algorithm for the experts problem that we consider in this dissertation.

Now we design an MAB algorithm, which performs two steps every round. In the *decision step*, the MAB algorithm uses $\mathcal{A}$ as a subroutine to obtain x_t. Then, we sample $i_t \sim p_t$, where p_t is a transformation of x_t. In the *estimation step*, we construct an *estimated* loss vector $\hat{l}_t$ from ℓ_{t,i_t}, which is the only observed coordinate. The estimated loss is now fed into $\mathcal{A}$.

All known MAB algorithms in the literature follow this template. For example, the well-known EXP3 algorithm (Bubeck et al. 2012) uses the Follow the Regularized Leader with entropy regularizer as its subroutine for choosing x_t. Then, it samples i_t directly from x_t (i.e., $p_t = x_t$) and then uses the unbiased importance weighting scheme for estimation: $\hat{l}_t = \ell_{t,i_t} / p_{t,i_t}$.

2.4.2 Implementation of Follow the Perturbed Leader

The *expected* version of FTPL algorithm for the experts setting, which returns a full probability vector instead of a single sample, can be used to generate a sequence $p_{1:t}$. The problem is in the estimation step, there is generally no closed form for $p_{1:t}$. However, we assume there is a close approximation for it such that the regret due to the approximation error is subsumed in the regret bounds. Indeed, Geometric Resampling (GR) technique by Neu and Bartók (2013) is one such method.

CHAPTER III

Follow the Perturbed Leader Analysis via Convex Duality

3.1 Preliminaries

3.1.1 Convex Analysis

For this section, let f be a differentiable, closed, and proper convex function with $\mathrm{dom}(f) \subseteq \mathbb{R}^N$. We say that f is *L-Lipschitz continuous* (or simply *Lipschitz*) with respect to a norm $\|\cdot\|$ when f satisfies $|f(x) - f(y)| \leq L\|x - y\|$ for all $x, y \in \mathrm{dom}(f)$.

The *Bregman divergence* $D_f(y, x)$ is the gap between $f(y)$ and the linear approximation of $f(y)$ around x. Formally, $D_f(y, x) = f(y) - f(x) - \langle \nabla f(x), y - x \rangle$. We say that f is *β-strongly convex* with respect to a norm $\|\cdot\|$ if we have $D_f(y, x) \geq \frac{\beta}{2}\|y - x\|^2$ for all $x, y \in \mathrm{dom}\, f$. Similarly, f is said to be *β-strongly smooth* with respect to a norm $\|\cdot\|$ if we have $D_f(y, x) \leq \frac{\beta}{2}\|y - x\|^2$ for all $x, y \in \mathrm{dom}\, f$.

The Bregman divergence measures how fast the gradient changes, or equivalently, how large the second derivative is. In fact, we can bound the Bregman divergence by analyzing the Hessian, as the following adaptation of Abernethy et al. (2013, Lemma 4.6) shows.

Lemma 3.1. *Let f be a twice-differentiable convex function with $\mathrm{dom}\, f \subseteq \mathbb{R}^N$, and let $\|\cdot\|$ be an arbitrary norm. Assume that there exists B such that for every $x \in \mathrm{dom}\, f$,*

$$\sup_{v:\|v\|\leq 1} v^\top \nabla^2 f(x) v \leq B.$$

Then, $D_f(x + v, x) \leq B\|v\|^2/2$ for any $x, x + v \in \mathrm{dom}\, f$.

The *Fenchel conjugate* of f is defined as $f^\star(x) = \sup_{w \in \mathrm{dom}(f)} \{\langle w, x \rangle - f(w)\}$, and it is a dual mapping that satisfies $f = (f^\star)^\star$. If f is differentiable and strictly convex we also

have $\nabla f^\star \in \mathrm{dom}(f)$. The notions of strong convexity and strong smoothness are *dual* to each other. That is, f is β-strongly convex with respect to a norm $\|\cdot\|$ if and only if $f^\star$ is $\frac{1}{\beta}$-strongly smooth with respect to the dual norm $\|\cdot\|_\star$. For more details and proofs, readers are referred to an excellent survey by Shalev-Shwartz (2012).

In this dissertation, we slightly abuse the term *gradient* to refer to any sub-gradient of a function.

3.2 Gradient-Based Prediction Algorithm

Recall that the comparator term in the regretdefinition (2.1) is a function of the final cumulative reward vector G_T. We define the *baseline potential function* for a given OLO problem to be

$$\Phi(G) := \max_{x \in \mathcal{X}} \langle x, G \rangle$$

so that the learner's performance is compared against $\Phi(G_T)$.

In convex analysis literature, this function is called the *support function* of $\mathcal{X}$. For a bounded compact set $\mathcal{X}$, the support function of $\mathcal{X}$ has useful properties.(Rockafellar 1997, Section 13).

Lemma 3.2. *Let f be the support function of a bounded compact set $\mathcal{X}$. Then, f has the following properties:*

- *It is positively homogeneous: $f(\alpha x) = \alpha f(x)$ for all $\alpha > 0$.*
- *It is sub-additive: $f(x) + f(y) \geq f(x + y)$.*
- *It is Lipschitz continuous with respect to any norm $\|\cdot\|$, where the Lipschitz constant is equal to $\sup_{x \in \mathcal{X}} \|x\|_*$.*

Note that the *gradient* of the baseline potential function is

$$\nabla \Phi(G) = \arg\max_{x \in \mathcal{X}} \langle x, G \rangle.$$

By evaluating this gradient at G_{t-1}, we recover the FTL decision rule. Similarly, we can see that FTRL decision rule is playing the gradients of a slight modification of the baseline potential function, as in the following lemma:

Lemma 3.3. *Let $\mathcal{R}$ be a convex function and let $f(G) = \max_{x \in \mathcal{X}} \{\langle x, G \rangle - \mathcal{R}(x)\}$. Then, $x^* = \arg\max_{x \in \mathcal{X}} \{\langle x, G \rangle - \mathcal{R}(x)\}$ is the (sub-)gradient of f at G.*

Proof. It suffices to note that

$$f(G) + \langle x^*, G' - G \rangle = \langle x^*, G' \rangle - \mathcal{R}(x^*) \leq f(G'). \qquad \square$$

In this perspective, we can define a family of algorithms that play the gradients of a function as in Algorithm 1, which we name Gradient-based Prediction Algorithm (GBPA). We note that Cesa-Bianchi and Lugosi (2006, Theorem 11.6) presented a similar algorithm, but our formulation is simpler as it eliminates all dual mappings.

Input: $\mathcal{X}, \mathcal{Y} \subseteq \mathbb{R}^N$
Require: convex potentials $\widetilde{\Phi}_1, \ldots, \widetilde{\Phi}_T : \mathbb{R}^N \to \mathbb{R}$, with $\nabla \widetilde{\Phi}_t(G) \in \mathcal{X},\ \forall G$
Initialize: $G_0 = 0$
for *t = 1 to T* **do**
  The learner plays $w_t = \nabla \widetilde{\Phi}_t(G_{t-1})$
  The adversary reveals $g_t \in \mathcal{Y}$
  The learner receives a reward of $\langle w_t, g_t \rangle$
  Update the cumulative gain vector: $G_t = G_{t-1} + g_t$
end

Algorithm 1: Gradient-Based Prediction Algorithm (GBPA)

3.2.1 Regret Analysis

We begin with a generic result on the regret of GBPA.

Lemma 3.4 (GBPA Regret)**.** *Let Φ be the baseline potential function for an online linear optimization problem. The regret of the GBPA can be decomposed as follows:*

$$\begin{aligned} \mathsf{Regret} = \sum_{t=1}^{T} \Bigg(\underbrace{\left(\widetilde{\Phi}_t(G_{t-1}) - \widetilde{\Phi}_{t-1}(G_{t-1})\right)}_{\textit{overestimation penalty}} + \underbrace{D_{\widetilde{\Phi}_t}(G_t, G_{t-1})}_{\textit{divergence penalty}} \Bigg) \\ + \underbrace{\Phi(G_T) - \widetilde{\Phi}_T(G_T)}_{\textit{underestimation penalty}}, \end{aligned} \tag{3.1}$$

where $\widetilde{\Phi}_0 \equiv \Phi$.

Proof. We note that since $\widetilde{\Phi}_0(0) = 0$,

$$\begin{aligned}
\widetilde{\Phi}_T(G_T) &= \textstyle\sum_{t=1}^{T} \widetilde{\Phi}_t(G_t) - \widetilde{\Phi}_{t-1}(G_{t-1}) \\
&= \textstyle\sum_{t=1}^{T} \Big(\big(\widetilde{\Phi}_t(G_t) - \widetilde{\Phi}_t(G_{t-1})\big) + \big(\widetilde{\Phi}_t(G_{t-1}) - \widetilde{\Phi}_{t-1}(G_{t-1})\big) \Big) \\
&= \textstyle\sum_{t=1}^{T} \Big(\big(\langle \nabla \widetilde{\Phi}_t(G_{t-1}), g_t \rangle + D_{\widetilde{\Phi}_t}(G_t, G_{t-1})\big) \\
&\qquad + \big(\widetilde{\Phi}_t(G_{t-1}) - \widetilde{\Phi}_{t-1}(G_{t-1})\big) \Big),
\end{aligned}$$

where the last equality holds because:

$$\widetilde{\Phi}_t(G_t) - \widetilde{\Phi}_t(G_{t-1}) = \langle \nabla \widetilde{\Phi}_t(G_{t-1}), g_t \rangle + D_{\widetilde{\Phi}_t}(G_t, G_{t-1}).$$

We now have

$$\begin{aligned}
\mathsf{Regret}_T &:= \Phi(G_T) - \sum_{t=1}^{T} \langle w_t, g_t \rangle \\
&= \Phi(G_T) - \textstyle\sum_{t=1}^{T} \langle \nabla \widetilde{\Phi}_t(G_{t-1}), g_t \rangle \\
&= \Phi(G_T) - \widetilde{\Phi}_T(G_T) + \textstyle\sum_{t=1}^{T} D_{\widetilde{\Phi}_t}(G_t, G_{t-1}) + \widetilde{\Phi}_t(G_{t-1}) - \widetilde{\Phi}_{t-1}(G_{t-1}),
\end{aligned}$$

which completes the proof. □

We point out a couple of important facts about Lemma 3.4:

1. If $\widetilde{\Phi}_1 \equiv \cdots \equiv \widetilde{\Phi}_T$, then the overestimation penalty sums up to $\widetilde{\Phi}_1(0) - \widetilde{\Phi}(0) = \widetilde{\Phi}_T(0) - \Phi(0)$.

2. If $\widetilde{\Phi}_t$ is β-strongly smooth with respect to $\|\cdot\|$, the divergence penalty at t is at most $\frac{\beta}{2}\|g_t\|^2$.

The above lemma proves an *equality*, which breaks down the regret into two sources. The Bregman divergence of $\widetilde{\Phi}_t$ captures the fact that the GBPA always ascends along the gradient that is one step behind. The adversary can exploit this and play g_t to induce a large *gap* between $\widetilde{\Phi}_t(G_t)$ and the linear approximation of $\widetilde{\Phi}_t(G_t)$ around G_{t-1}. The learner can reduce this gap by choosing a *smooth* $\widetilde{\Phi}_t$ whose gradient changes slowly. On the other hand, the overestimation and underestimation penalty terms prevent the learner from achieving a low regret by choosing an arbitrarily smooth $\widetilde{\Phi}_t$.

In short, the GBPA achieves low regret if the potential function $\widetilde{\Phi}_t$ gives a favorable tradeoff between the two sources of regret. This tradeoff is captured by the following definition of *smoothing parameters*, adapted from Beck and Teboulle (2012, Definition 2.1).

Definition 3.5. *Let f be a closed proper convex function. A collection of functions $\{\tilde{f}_\eta : \eta \in \mathbb{R}_+\}$ is said to be an η-*smoothing *of f with* smoothing parameters *$(\alpha, \beta, \|\cdot\|)$, if for every $\eta > 0$:*

1. *There exists real numbers α_1 (underestimation bound) and α_2 (overestimation bound) such that*
$$\sup_{G \in \mathrm{dom}(f)} f(G) - \tilde{f}_\eta(G) \leq \alpha_1 \eta \quad \text{and} \quad \sup_{G \in \mathrm{dom}(f)} \tilde{f}_\eta(G) - f(G) \leq \alpha_2 \eta$$
with $\alpha_1 + \alpha_2 = \alpha$.
2. *$\tilde{f}_\eta$ is $\frac{\beta}{\eta}$-strongly smooth with respect to $\|\cdot\|$.*

We say α is the deviation parameter, *and β is the* smoothness parameter.

A straightforward application of Lemma 3.4 gives the following statement:

Corollary 3.6. *Let Φ be the baseline potential for an online linear optimization problem. Suppose $\{\widetilde{\Phi}_\eta\}$ is an η-smoothing of Φ with parameters $(\alpha, \beta, \|\cdot\|)$. Then, the GBPA run with $\widetilde{\Phi}_1 \equiv \cdots \equiv \widetilde{\Phi}_T \equiv \widetilde{\Phi}_\eta$ enjoys the following regret bound,*

$$\mathsf{Regret}_{\leq} \alpha\eta + \frac{\beta}{2\eta} \sum_{t=1}^{T} \|g_t\|^2.$$

Choosing η to optimize the bound gives $\mathsf{Regret}_{\leq} \sqrt{2\alpha\beta \sum_{t=1}^{T} \|g_t\|^2}$.

In OLO, we often consider the settings where the reward vectors $g_1, \ldots, g_t$ are constrained in norm, i.e., $\|g_t\| \leq r$ for all t. In such settings, the regret grows in $O(r\sqrt{\alpha\beta T})$ for the optimal choice of η. The product of smoothing parameters $\alpha\beta$ is, therefore, at the core of the GBPA regret analysis.

3.2.2 Follow the Regularized Leader as a Gradient Based Prediction Algorithm

Define the *regularized potential* as follows:

$$\widetilde{\Phi}(G) = \mathcal{R}^\star(G) = \max_{x \in \mathcal{X}} \{\langle x, G \rangle - \mathcal{R}(x)\} \tag{3.2}$$

where $\mathcal{R} : \mathcal{X} \to \mathbb{R}$ is some strictly convex function. The class of FTRL algorithms can be viewed precisely as an instance of GBPA where the potential is chosen according to (3.2).

In the convex optimization community, this technique has been referred to as *inf-conv smoothing* of Φ with $\mathcal{R}^*$ (Beck and Teboulle 2012), due to the following equality:

$$\mathcal{R}^\star(G) = \inf_{G'} \left\{ \Phi(G') + \mathcal{R}^\star(G - G') \right\}.$$

3.2.3 Follow the Perturbed Leader as a Gradient Based Prediction Algorithm

An important smoothing technique for this chapter is *stochasting smoothing*, which is the convolution of a function with a probability density function.

Definition 3.7 (Stochastic Smoothing)**.** *Let $f : \mathbb{R}^N \to \mathbb{R}$ be a function. We define $\tilde{f}(\cdot; \mathcal{D}_\eta)$ to be the* stochastic smoothing *of f with distribution $\mathcal{D}$ and scaling parameter $\eta > 0$. The function value at G is obtained as:*

$$\tilde{f}(G; \mathcal{D}_\eta) := \mathbb{E}_{z' \sim \mathcal{D}_\eta}[f(G + z')] = \mathbb{E}_{z \sim \mathcal{D}}[f(G + \eta z)],$$

where we adopt the convention that if z has distribution $\mathcal{D}$ then the distribution of ηz is denoted by $\mathcal{D}_\eta$.

The technique of *stochastic smoothing* has been well-studied in the optimization literature for gradient-free optimization algorithms (Glasserman 1991; Yousefian et al. 2010) and accelerated gradient methods for non-smooth optimizations (Duchi et al. 2011).

Let $\mathcal{D}$ be a probability distribution over $\mathbb{R}^N$ with a well-defined density everywhere. Consider the GBPA run with a stochastic smoothing of the baseline potential:

$$\forall t, \widetilde{\Phi}_t(G) = \widetilde{\Phi}(G; \mathcal{D}_{\eta_t}) = \mathbb{E}_{z \sim \mathcal{D}} \Big[\max_{x \in \mathcal{X}} \langle x, G + \eta_t z \rangle \Big]. \tag{3.3}$$

Then, from the convexity of $G \mapsto \max_{x \in \mathcal{X}} \langle x, G + \eta_t z \rangle$ (for any fixed z), we can swap the expectation and gradient (Bertsekas 1973, Proposition 2.2) and evaluate the gradient at $G = G_{t-1}$ to obtain

$$\nabla \widetilde{\Phi}_t(G_{t-1}) = \mathbb{E}_{z \sim \mathcal{D}} \Big[\arg\max_{x \in \mathcal{X}} \langle x, G_{t-1} + \eta_t z \rangle \Big]. \tag{3.4}$$

Taking a single random sample of arg max inside expectation is equivalent to the decision rule of FTPL; the GBPA on a stochastically smoothed potential can thus be seen as

playing the *expected action* of FTPL. Since the learner gets a linear reward in online linear optimization, the regret of the GBPA on a stochastically smoothed potential is equal to the *expected regret* of FTPL. For this reason, we will use the terms FTPL and GBPA with stochastic smoothing interchangeably.

One very useful property of stochastic smoothing is that as long as $\mathcal{D}$ has a support over $\mathbb{R}^N$ and has a differentiable probability density function μ, $\tilde{f}$ is always differentiable. To see this, we use the change of variable technique:

$$\tilde{f}(G;\mathcal{D}) = \int f(G+z)\mu(z)\,dz = \int f(\tilde{G})\mu(\tilde{G}-G)\,d\tilde{G},$$

and it follows that

$$\begin{aligned}\nabla_G \tilde{f}(G;\mathcal{D}) &= -\int f(\tilde{G})\nabla_G \mu(\tilde{G}-G)\,d\tilde{G},\\ \nabla_G^2 \tilde{f}(G;\mathcal{D}) &= \int f(\tilde{G})\nabla_G^2 \mu(\tilde{G}-G)\,d\tilde{G}.\end{aligned} \tag{3.5}$$

This change of variable trick leads to the following useful expressions for the first and second derivatives of $\tilde{f}$ in case the density $\mu(G)$ is proportional to $\exp(-\nu(G))$ for a sufficiently smooth ν.

Lemma 3.8 (Exponential Family Smoothing)**.** *Suppose $\mathcal{D}$ is a distribution over $\mathbb{R}^N$ with a probability density function μ of the form $\mu(G) = \exp(-\nu(G))/Z$ for some normalization constant Z. Then, for any twice-differentiable ν, we have*

$$\begin{aligned}\nabla \tilde{f}(G) &= \mathbb{E}[f(G+z)\nabla_z \nu(z)],\\ \nabla^2 \tilde{f}(G) &= \mathbb{E}[f(G+z)\left(\nabla_z \nu(z)\nabla_z \nu(z)^T - \nabla_z^2 \nu(z)\right)].\end{aligned} \tag{3.6}$$

Furthermore, if f is convex, we have

$$\nabla^2 \tilde{f}(G) = \mathbb{E}[\nabla f(G+z)\nabla_z \nu(z)^T].$$

Proof. If ν is twice-differentiable, $\nabla\mu = -\mu\cdot\nabla\nu$ and $\nabla^2\mu = \left(\nabla\nu\nabla\nu^T - \nabla^2\nu\right)\mu$. Plugging these in (3.5) and using the substitution $z = \tilde{G} - G$ immediately gives the first two claims of the lemma. For the last claim, we first directly differentiate the expression for $\nabla\tilde{f}$ in (3.6) by swapping the expectation and gradient. This is justified because f is convex (and is hence differentiable almost everywhere) and μ is absolutely continuous w.r.t. Lebesgue measure everywhere (Bertsekas 1973, Proposition 2.3). □

Notes on estimation penalty If the perturbation used has mean zero, it follows from Jensen's inequality that the stochastic smoothing will overestimate the convex function Φ. Hence, for mean zero perturbations, the underestimation penalty is always non-positive. When the scaling parameter η_t changes every iteration, the overestimation penalty becomes a sum of T terms. The following lemma shows that we can collapse them into one since the baseline potential Φ in OLO problems is sub-additive: $\Phi(G + H) \leq \Phi(G) + \Phi(H)$.

Lemma 3.9. *Let $\Phi : \mathbb{R}^N \to \mathbb{R}$ be the baseline potential function of an OLO problem. Let $\mathcal{D}$ be a continuous distribution with mean zero and support $\mathbb{R}^N$. Consider the GBPA with $\widetilde{\Phi}_t(G) = \widetilde{\Phi}(G; \mathcal{D}_{\eta_t})$ for $t = 0, \ldots, T$ where $(\eta_1, \ldots, \eta_T)$ is a non-decreasing sequence of non-negative numbers. Then the overestimation penalty has the following upper bound,*

$$\sum_{t=1}^{T} \widetilde{\Phi}_t(G_{t-1}) - \widetilde{\Phi}_{t-1}(G_{t-1}) \leq \eta_T \mathbb{E}_{u \sim \mathcal{D}}[\Phi(u)],$$

and the underestimation penalty is non-positive which gives gives a regret bound of

$$\mathsf{Regret} \leq \eta_T \mathbb{E}_{u \sim \mathcal{D}}[\Phi(u)] + \sum_{t=1}^{T} D_{\widetilde{\Phi}_t}(G_t, G_{t-1}). \tag{3.7}$$

Proof. By virtue of the fact that Φ is a support function, it is also sub-additive and satisfies the triangle inequality (Lemma 3.2). Hence we can see that, for any $0 < \eta' \leq \eta$,

$$\begin{aligned}\widetilde{\Phi}(G; \mathcal{D}_\eta) - \widetilde{\Phi}(G; \mathcal{D}_{\eta'}) &= \mathbb{E}_{u \sim \mathcal{D}}[\Phi(G + \eta u) - \Phi(G + \eta' u)] \\ &\leq \mathbb{E}_{u \sim \mathcal{D}}[\Phi((\eta - \eta')u)] = (\eta - \eta')\mathbb{E}_{u \sim \mathcal{D}}[\Phi(u)],\end{aligned}$$

where the final line follows from the positive homogeneity of Φ. Since we implicitly assume that $\widetilde{\Phi}_0 \equiv \Phi$ we can set $\eta_0 = 0$. We can then conclude that

$$\sum_{t=1}^{T} \widetilde{\Phi}_t(G_{t-1}) - \widetilde{\Phi}_{t-1}(G_{t-1}) \leq \left(\sum_{t=1}^{T} \eta_t - \eta_{t-1} \right) \mathbb{E}_{u \sim \mathcal{D}}[\Phi(u)] = \eta_T \mathbb{E}_{u \sim \mathcal{D}}[\Phi(u)],$$

which completes the proof. □

3.2.4 Connection between Follow the Perturbed Leader and Follow the Regularized Leader

Now that we have seen that FTRL and FTPL can be viewed as a certain type of smoothing operation, a natural question one might ask is: to what extent are stochastic smoothing and inf-conv smoothing related? That is, can we view FTRL and FTPL as really two sides of the same coin? The answer here is "partially yes" and "partially no":

1. When $\mathcal{X}$ is 1-dimensional then (nearly) every instance of FTRL can be seen as a special case of FTPL, and vice versa. In other words, stochastic smoothing and inf-conv smoothing are effectively one and the same, and we describe this equivalence in detail below.

2. For problems of dimension larger than 1, every instance of FTPL can be described as an instance of FTRL. More precisely, if we have a distribution $\mathcal{D}_\eta$ which leads to a stochastically smoothed potential $\widetilde{\Phi}(\cdot) = \widetilde{\Phi}(\cdot; \mathcal{D}_\eta)$, then we can always write the gradient of $\widetilde{\Phi}(\cdot)$ as the solution of an FTRL optimization. That is,

$$\nabla\widetilde{\Phi}(G, \mathcal{D}_\eta) = \arg\max_{x \in \mathcal{X}} \langle x, \theta \rangle - \mathcal{R}(x) \quad \text{where} \quad \mathcal{R}(x) := \widetilde{\Phi}^\star(x),$$

   and we recall that $\widetilde{\Phi}^\star$ denotes the Fenchel Conjugate. In other words, the perturbation $\mathcal{D}$ induces an implicit regularizer defined as the cojugate of $\mathbb{E}_{z \sim \mathcal{D}}[\max_{g \in \mathcal{X}} \langle g, G \rangle]$

3. In general, however, stochastic smoothing is not as general as inf-conv smoothing. FTPL is in some sense less general than FTRL, as there are examples of regularizers that can not be "induced" via a specific perturbation. One particular case is given by Hofbauer and Sandholm (2002).

We now give a brief description of the equivalence between stochastic smoothing and inf-conv smoothing for the 1-dimensional case.

On the near-equivalence between FTRL and FTPL in one dimension. Consider a one-dimensional online linear optimization prediction problem where the player chooses an action x_t from $\mathcal{X} = [0, 1]$ and the adversary chooses a reward g_t from $\mathcal{Y} = [0, 1]$. This can be interpreted as a two-expert setting; the player's action $w_t \in \mathcal{X}$ is the probability

of following the first expert and g_t is the net excess reward of the first expert over the second. The baseline potential for this setting is $\widetilde{\Phi}(G) = \max_{x\in[0,1]} wG$.

Let us consider an instance of FTPL with a continuous distribution $\mathcal{D}$ whose cumulative density function (cdf) is $F_{\mathcal{D}}$. Let $\widetilde{\Phi}$ be the smoothed potential function (Equation 3.3) with distribution $\mathcal{D}$. Its derivative is

$$\widetilde{\Phi}'(G) = \mathbb{E}\left[\arg\max_{x\in[0,1]} x(G+u)\right] = \mathbb{P}[u > -G] \tag{3.8}$$

because the maximizer is unique with probability 1. Notice, crucially, that the derivative $\widetilde{\Phi}'(G)$ is exactly the expected solution of our FTPL instance. Moreover, by differentiating it again, we see that the second derivative of $\widetilde{\Phi}$ at G is exactly the pdf of $\mathcal{D}$ evaluated at $(-G)$.

We can now precisely define the mapping from FTPL to FTRL. Our goal is to find a convex regularization function $\mathcal{R}$ such that $\mathbb{P}(u > -G) = \arg\max_{x\in[0,1]} (xG - \mathcal{R}(x))$. Since this is a one-dimensional convex optimization problem, we can differentiate for the solution. The characterization of $\mathcal{R}$ is:

$$\mathcal{R}(x) - \mathcal{R}(0) = -\int_0^x \varphi_{\mathcal{D}}^{-1}(1-z)\mathrm{d}z. \tag{3.9}$$

Note that the cdf $\varphi_{\mathcal{D}}(\cdot)$ is indeed invertible since it is a strictly increasing function.

The inverse mapping is just as straightforward. Given a regularization function $\mathcal{R}$ well-defined over $[0,1]$, we can always construct its Fenchel conjugate $\mathcal{R}^\star(G) = \sup_{x\in[0,1]}\langle x, G\rangle - \mathcal{R}(x)$. The derivative of $\mathcal{R}^\star$ is an increasing convex function, whose infimum is 0 at $G = -\infty$ and supremum is 1 at $G = +\infty$. Hence, $\mathcal{R}^\star$ defines a cdf of the perturbation distribution that exactly reproduces FTRL corresponding to $\mathcal{R}$.

3.3 Generic Bounds

In this section, we show how the general result in Corollary 3.6, combined with stochastic smoothing results from the existing literature, painlessly yield regret bounds for two generic settings: one in which the learner/adversary sets are bounded in ℓ_∞/ℓ_1 norms and another in which they are bounded in the standard Euclidean (i.e., ℓ_2) norm.

Kalai and Vempala (2005, Theorem 1.1) showed that FTPL with exponential distribution and uniform distribution over hypercube are universal OLO algorithms. These distri-

butions have a simple and well-behaved probability density function that can be directly analyzed. This analysis, however, does not generalize to other continuous distributions even with a slightly more complex probability density function, such as Gaussian.

Our smoothing analysis framework lets us apply the convex optimization tools to the regret analysis in a painless manner. As a result, we prove the following new result that FTPL with Gaussian is a universal OLO algorithm.

Theorem 3.10. *Consider GBPA run with a potential $\widetilde{\Phi}_t(G) = \widetilde{\Phi}(G; \mathcal{D}_\eta)$ where $\mathcal{D}$ is the uniform distribution on the unit ℓ_2 ball. Then we have,*

$$\mathsf{Regret} \leq \frac{1}{2\eta}\sqrt{N}\|\mathcal{X}\|_2 \sum_{t=1}^{T} \|g_t\|_2^2 + \eta\|\mathcal{X}\|_2.$$

If we choose $\mathcal{D}$ to be the standard multivariate Gaussian distribution, then we have,

$$\mathsf{Regret} \leq \frac{1}{2\eta}\|\mathcal{X}\|_2 \sum_{t=1}^{T} \|g_t\|_2^2 + \eta\sqrt{N}\|\mathcal{X}\|_2.$$

In either case, optimizing over η we get $\mathsf{Regret} \leq \|\mathcal{X}\|_2 \sqrt[4]{N} \sqrt{2\sum_{t=1}^{T} \|g_t\|_2^2}$.

Proof. The baseline potential function Φ is $\|\mathcal{X}\|_2$-Lipschitz with respect to $\|\cdot\|_2$. Duchi et al. (2011, Lemma E.2) show that the stochastic smoothing of Φ with the uniform distribution on the Euclidean unit ball is an η-smoothing with parameters

$$\left(\|\mathcal{X}\|_2, \|\mathcal{X}\|_2\sqrt{N}, \|\cdot\|_1\right).$$

Further, Duchi et al. (2011, Lemma E.3) shows that the stochastic smoothing of Φ with the standard Gaussian distribution is an η-smoothing with parameters

$$\left(\|\mathcal{X}\|_2\sqrt{N}, \|\mathcal{X}\|_2, \|\cdot\|_1\right).$$

The result now follows from Corollary 3.6. □

We also revisit the FTPL with uniform noise on hypercube and improve the constant factors in (Kalai and Vempala 2005, Theorem 1.1.a)

Theorem 3.11. *Consider GBPA run with a potential $\widetilde{\Phi}_t(G) = \widetilde{\Phi}(G; \mathcal{D}_\eta)$ where $\mathcal{D}$ is the uniform*

distribution on the unit ℓ_∞ ball. Then we have,

$$\mathsf{Regret}_T \le \frac{1}{2\eta}\|\mathcal{X}\|_\infty \sum_{t=1}^T \|g_t\|_1^2 + \eta\frac{\|\mathcal{X}\|_\infty N}{2}.$$

Choosing η to optimize the bound gives $\mathsf{Regret}_T \le \|\mathcal{X}\|_\infty\sqrt{N\sum_{t=1}^T \|g_t\|_1^2}$.

Proof. The baseline potential function Φ is $\|\mathcal{X}\|_\infty$-Lipschitz with respect to $\|\cdot\|_1$. By Corollary 3.6, it suffices to prove that the stochastic smoothing of Φ with the uniform distribution on the unit ℓ_∞ ball is an η-smoothing with parameters

$$\left(\frac{\|\mathcal{X}\|_\infty N}{2}, \|\mathcal{X}\|_\infty, \|\cdot\|_1\right).$$

These smoothing parameters have been indeed shown to hold by Duchi et al. (2011, Lemma E.1). □

3.4 Unit Euclidean Balls Problem

The generic bound of Theorem 3.10 results in the suboptimal $O(\sqrt[4]{N}\sqrt{\sum_{t=1}^T \|g_t\|_2^2})$ regret bound for the unit Euclidean balls problem. In this section, we will show that a more careful analysis of the smoothing parameters for this setting yields the minimax-optimal regret bound up to a constant factor. Prior to this work, Rakhlin et al. (2012) proved that FTPL with the uniform distribution over the surface of the unit sphere enjoys regret bound of $4\sqrt{2T}$. We improve not only the constant factor but the dependence on time to $\sqrt{\sum_{t=1}^T \|g_t\|_2^2}$ which is always less than or equal to T.

Note that the baseline potential function is $\Phi(G) = \max_{x\in\mathcal{X}}\langle x, G\rangle = \|G\|_2$.

Theorem 3.12. *Let Φ be the baseline potential for the Euclidean balls setting. The GBPA run with $\widetilde{\Phi}_t(\cdot) = \widetilde{\Phi}(\cdot; \mathcal{N}(0, I)_{\eta_t})$ for all t has regret at most*

$$\mathsf{Regret}_{\le} \eta_T\sqrt{N} + \tfrac{1}{2\sqrt{N}}\textstyle\sum_{t=1}^T \frac{1}{\eta_t}\|g_t\|_2^2. \tag{3.10}$$

If the algorithm selects $\eta_t = \sqrt{\sum_{s=1}^T \|g_s\|_2^2/(2N)}$ for all t, we have

$$\mathsf{Regret}_{\le}\sqrt{2\textstyle\sum_{t=1}^T \|g_t\|_2^2}.$$

If the algorithm selects η_t adaptively according to $\eta_t = \sqrt{(1 + \sum_{s=1}^{t-1} \|g_s\|_2^2))/N}$, we have

$$\mathsf{Regret} \leq 2\sqrt{1 + \sum_{t=1}^{T} \|g_t\|_2^2}$$

Proof. The proof is mostly similar to that of Theorem 3.15. In order to apply Lemma 3.4, we need to upper bound (i) the overestimation and underestimation penalty, and (ii) the Bregman divergence.

The Gaussian smoothing always overestimates a convex function, so it suffices to bound the overestimation penalty. Furthermore, it suffices to consider the fixed η_t case due to Lemma 3.1. The overestimation penalty can be upper-bounded as follows:

$$\begin{aligned}\widetilde{\Phi}_T(0) - \widetilde{\Phi}(0) &= \mathbb{E}_{u\sim\mathcal{N}(0,I)}\|G + \eta_T u\|_2 - \|G\|_2 \\ &\leq \eta_T \mathbb{E}_{u\sim\mathcal{N}(0,I)}\|u\|_2 \leq \eta_T \sqrt{\mathbb{E}_{u\sim\mathcal{N}(0,I)}\|u\|_2^2} = \eta_T\sqrt{N}.\end{aligned}$$

The first inequality is from the triangle inequality, and the second inequality is from the concavity of the square root.

For the divergence penalty, note that $\max_{v:\|g\|_2=1} g^T(\nabla^2\widetilde{\Phi})g \leq \|\lambda(\nabla^2\widetilde{\Phi})\|_\infty$, which we bound in Lemma 3.13. The final step is to apply Lemma 3.1. □

Lemma 3.13. *Let Φ be the baseline potential for the Euclidean balls setting. Then, for all $G \in \mathbb{R}^N$ and $\eta > 0$, the Hessian matrix of the Gaussian smoothed potential satisfies*

$$\nabla^2\widetilde{\Phi}(G; \mathcal{N}(0,I)_\eta) \preceq \tfrac{1}{\eta\sqrt{N}} I.$$

Proof. The Hessian of the Euclidean norm $\nabla^2\Phi(G) = \|G\|_2^{-1} I - \|G\|_2^{-3} GG^T$ diverges near $G = 0$. Expectedly, the maximum curvature is at origin even after Gaussian smoothing (See Appendix A.1). So, it suffices to prove

$$\nabla^2\widetilde{\Phi}(0) = \mathbb{E}_{u\sim\mathcal{N}(0,I)}[\|u\|_2(uu^T - I)] \preceq \sqrt{\tfrac{1}{N}} I,$$

where the Hessian expression is from Lemma 3.8.

By symmetry, all off-diagonal elements of the Hessian are 0. Let $Y = \|u\|^2$, which is

Chi-squared with N degrees of freedom. So,

$$\begin{aligned}\mathrm{Tr}(\mathbb{E}[\|u\|_2(uu^T - I)]) &= \mathbb{E}[\mathrm{Tr}(\|u\|_2(uu^T - I))] = \mathbb{E}[\|u\|_2^3 - N\|u\|_2] \\ &= \mathbb{E}[Y^{\frac{3}{2}}] - N\mathbb{E}[Y^{\frac{1}{2}}]\end{aligned}$$

Using the Chi-squared moment formula (Simon 2002, p. 13):

$$\mathbb{E}[Y^k] = \frac{2^k\Gamma(\frac{N}{2}+k)}{\Gamma(\frac{N}{2})},$$

the above becomes:

$$\frac{2^{\frac{3}{2}}\Gamma(\frac{3}{2}+\frac{N}{2})}{\Gamma(\frac{N}{2})} - \frac{N2^{\frac{1}{2}}\Gamma(\frac{1}{2}+\frac{N}{2})}{\Gamma(\frac{N}{2})} = \frac{\sqrt{2}\Gamma(\frac{1}{2}+\frac{N}{2})}{\Gamma(\frac{N}{2})}. \tag{3.11}$$

From the log-convexity of the Gamma function,

$$\log\Gamma\left(\tfrac{1}{2}+\tfrac{N}{2}\right) \le \tfrac{1}{2}\left(\log\Gamma\left(\tfrac{N}{2}\right) + \log\Gamma\left(\tfrac{N}{2}+1\right)\right) = \log\Gamma\left(\tfrac{N}{2}\right)\sqrt{\tfrac{N}{2}}.$$

Exponentiating both sides, we obtain

$$\Gamma\left(\tfrac{1}{2}+\tfrac{N}{2}\right) \le \Gamma\left(\tfrac{N}{2}\right)\sqrt{\tfrac{N}{2}},$$

which we apply to Equation 3.11 and get $\mathrm{Tr}(\nabla^2\widetilde{\Phi}(0)) \le \sqrt{N}$. To complete the proof, note that by symmetry, each entry must have the same expected value, and hence it is bounded by $\sqrt{1/N}$. □

3.5 Experts Problem

In our framework we have used language of maximizing gain, in contrast to the more common theme of minimizing loss. However, the loss-only setting can be easily obtained by simply changing the domain $\mathcal{Y}$ to contain only vectors with negative-valued coordinates.

3.5.1 Exponential Weights Algorithm as a Follow the Pertured Leader

The most well-known and widely used algorithm in the experts setting is the *Exponential Weights Algorithm* (EWA), often referred to as the *Multiplicative Weights Algorithm* and strongly related to the classical *Weighted Majority Algorithm* (Littlestone and Warmuth 1994). On round t, EWA specifies a set of unnormalized weights based on the cumulative gains thus far,

$$\tilde{w}_{t,i} := \exp(\eta G_{t-1,i}) \qquad i = 1, \dots, N,$$

where $\eta > 0$ is a parameter. The learner's distribution on this round is then obtained by normalizing $\tilde{w}_t$

$$w_{t,i} := \frac{\tilde{w}_{t,i}}{\sum_{j=1}^{N} \tilde{w}_{t,j}} \qquad i = 1, \dots, N. \tag{3.12}$$

More recent perspectives of EWA have relied on an alternative interpretation via an optimization problem. Indeed the weights obtained in Eqn. 3.12 can be equivalently obtained as follows,

$$x_t = \arg\max_{x \in \Delta^N} \left\{ \langle \eta G_{t-1}, w \rangle - \sum_{i=1}^{N} w_i \log w_i \right\}.$$

We have cast the exponential weights algorithm as an instance of FTRL where the regularization function $\mathcal{R}$ corresponds to the *negative entropy function*, $\mathcal{R}(w) := \sum_i w_i \log w_i$. Applying Lemma 3.4 one can show that EWA obtains a regret of order $\sqrt{T \log N}$.

A third interpretation of EWA is obtained via the notion of stochastic smoothing (perturbations) using the *Gumbel distribution*:

$$\begin{aligned} \mu(z) &:= e^{-(z+e^{-z})} && \text{is the PDF of the standard Gumbel; and} \\ \Pr(Z \le z) &= e^{-e^{-z}} && \text{is the CDF of the standard Gumbel.} \end{aligned}$$

The Gumbel distribution has several natural properties, including for example that it is *max-stable*: the maximum value of several Gumbel-distributed random variables is itself distributed according to a Gumbel distribution[1]. But another nice fact is that the distribution of the maximizer of N fixed values perturbed with Gumbel noise leads to an exponentially-weighted distribution. Precisely, if we have a values $v_1, \dots, v_N$, and we draw n IID samples $Z_1, \dots, Z_N$ from the standard Gumbel, then a straightforward calcu-

[1] Above we only defined the standard Gumbel, but in general the Gumbel has both a scaling and shift parameter.

lus exercise gives that

$$\Pr\left[v_i + Z_i = \max_{j=1,\ldots,N}\{v_j + Z_j\}\right] = \frac{\exp(v_i)}{\sum_{j=1,\ldots,N}\exp(v_j)} \qquad i = 1,\ldots,N.$$

What we have just arrived at is that EWA is indeed an instance of FTPL with Gumbel-distributed noise. This was described by Adam Kalai in personal communication, and later Warmuth (2009) expanded it into a short note available online. However, the result appears to be folklore in the area of probabilistic choice models, and it is mentioned briefly by Hofbauer and Sandholm (2002).

3.5.2 Exponential Family Perturbations

We will now apply our stochastic smoothing analysis to derive bounds on a class of algorithms for the Experts Setting using three different perturbations: the *Exponential*, *Gaussian*, and *Gumbel*. The latter noise distribution generates an algorithm which is equivalent to EWA, as discussed above, but we prove the same bound using new tools. Note, however that we use a mean-zero Gumbel whereas the standard Gumbel has mean 1.

The key lemma for the GBPA analysis is Lemma 3.4, which decomposes the regret into overestimation, underestimation, and divergence penalty. By Lemma 3.9, the underestimation is less than or equal to 0 and the overestimation penalty is upper-bounded by $\mathbb{E}_{z\sim\mathcal{D}}\left[\max_{i=1,\ldots,N} z_i\right]$. This expectation for commonly used distributions $\mathcal{D}$ is well-studied in extreme value theory.

In order to upper bound the divergence penalty, it is convenient to analyze the Hessian matrix, which has a nice structure in the experts setting. We will be especially interested in bounding the trace of this Hessian.

Lemma 3.14. *Let Φ be the baseline potential for the N-experts setting, and $\mathcal{D}$ be a continuous distribution with a differentiable probability density function $\mu_{\mathcal{D}}$. We will consider the potential $\widetilde{\Phi}(G) = \widetilde{\Phi}(G;\mathcal{D}_\eta)$. If for some constant β we have a bound $\mathrm{Tr}(\nabla^2\widetilde{\Phi}(G)) \le \beta/\eta$ for every G, then it follows that*

$$D_{\widetilde{\Phi}}(G+g, G) \le \beta\|g\|_\infty^2/\eta. \tag{3.13}$$

Proof. The Hessian exists because $\mu_{\mathcal{D}}$ is differentiable (Equation 3.5). Let H denote the Hessian matrix of the stochastic smoothing of Φ, i.e., $H(\cdot) = \nabla^2\widetilde{\Phi}(\cdot;\mathcal{D}_\eta)$. In order to apply Lemma 3.1, we must bound $\sum_{i,j}|H_{ij}|$, the sum of absolute values of all entries of H.

We claim two properties on H:

1. Diagonal entries are non-negative and off diagonal entries are non-positive.

2. Each row or column sums up to 0.

All diagonal entries of H are non-negative because $\widetilde{\Phi}$ is convex. Note that $\nabla_i \widetilde{\Phi}$ is the probability that the i-th coordinate of $G + z$ is the maximum coordinate, and an increase in the j-th of G where $j \neq i$ cannot increase that probability; hence, the off-diagonal entries of H are non-positive. To prove the second claim, note that the gradient $\nabla \widetilde{\Phi}$ is a probability vector, whose coordinates always sum up to 1. Thus, each row (or each column) must sum up to 0.

It follows from these properties that $\sum_{i,j} |H_{ij}| \leq 2\mathrm{Tr}(H)$, as desired. □

The above result will be very convenient in proving bounds on the divergence penalty associated with different noise distributions. In particular, assume we have a noise distribution with exponential form, then IID sample $z = (z_1, \ldots, z_n)$ has density $\mu(z) \propto \prod_i \exp(-\nu(z_i))$. Now applying Lemma 3.8 we have a nice expression for the diagonal Hessian values:

$$\begin{aligned} \nabla^2_{ii} \widetilde{\Phi}(G; \mathcal{D}_\eta) &= \frac{1}{\eta} \mathop{\mathbb{E}}_{(z_1,\ldots,z_n)\sim\mu} \left[\nabla_i \Phi(G + \eta z) \frac{d}{dz_i} \nu(z_i) \right] \\ &= \frac{1}{\eta} \mathop{\mathbb{E}}_{(z_1,\ldots,z_n)\sim\mu} \left[\mathbf{1}\{i = i^*(G + \eta z)\} \frac{d\nu(z_i)}{dz_i} \right]. \end{aligned} \tag{3.14}$$

The above formula now gives us a natural bound on the trace of the Hessian for the three distributions of interest.

- **Laplace:** For this distribution we have $\nu(z) = |z| \implies \frac{d\nu(z)}{dz} = \mathrm{sign}(z)$, where the sign function returns +1 if the argument is positive, -1 if the argument is negative, and 0 otherwise. Then we have

$$\begin{aligned} \mathrm{Tr}(\nabla^2 \widetilde{\Phi}(G)) &= \frac{1}{\eta} \mathop{\mathbb{E}}_{(z_1,\ldots,z_n)\sim\mu} \left[\textstyle\sum_{i=1}^N \mathbf{1}\{i = i^*(G + \eta z)\} \frac{d\nu(z_i)}{dz_i} \right] \\ &= \frac{1}{\eta} \mathop{\mathbb{E}}_{z} \left[\textstyle\sum_{i=1}^N \mathbf{1}\{i = i^*(G + \eta z)\} \mathrm{sign}(z_i) \right] \\ &\leq \frac{1}{\eta} \mathop{\mathbb{E}}_{z} \left[\textstyle\sum_{i=1}^N \mathbf{1}\{i = i^*(G + \eta z)\} \right] = \frac{1}{\eta}. \end{aligned}$$

- **Gumbel:** Here, using zero-mean Gumbel, we have $\nu(z) = z + 1 + e^{-z-1} \implies \frac{d\nu(z)}{dz} = 1 - e^{-z-1}$. Applying the same arguments we obtain

$$\begin{aligned}\mathrm{Tr}(\nabla^2\widetilde{\Phi}(G)) &= \frac{1}{\eta}\mathop{\mathbb{E}}_{z}\left[\textstyle\sum_{i=1}^{N}\mathbf{1}\{i = i^*(G+\eta z)\}(1-e^{-z_i-1})\right] \\ &\leq \frac{1}{\eta}\mathop{\mathbb{E}}_{z}\left[\textstyle\sum_{i=1}^{N}\mathbf{1}\{i = i^*(G+\eta z)\}\right] = \frac{1}{\eta}.\end{aligned}$$

- **Gaussian:** Here we have $\nu(z) = \frac{z^2}{2} \implies \frac{d\nu(z)}{dz} = z$. Bounding the sum of diagonal Hessian terms requires a slightly different trick:

$$\begin{aligned}\mathrm{Tr}(\nabla^2\widetilde{\Phi}(G)) &= \frac{1}{\eta}\mathop{\mathbb{E}}_{z}\left[\textstyle\sum_{i=1}^{N}\mathbf{1}\{i = i^*(G+\eta z)\}z_i\right] \\ &= \frac{1}{\eta}\mathop{\mathbb{E}}_{z}\left[z_{i^*(G+\eta z)}\right] \leq \frac{1}{\eta}\mathop{\mathbb{E}}_{z}[\max_i z_i] \leq \frac{\sqrt{2\log N}}{\eta}.\end{aligned}$$

  where the last inequality follows according to moment generating function arguments given below.

To obtain regret bounds, all that remains is a bound on the overestimation penalty. As we showed in Lemma 3.9, the overestimation penalty is upper bounded as $\eta\mathbb{E}_{z\sim\mathcal{D}}[\Phi(z)] = \eta\mathbb{E}[\max_i z_i]$. We can bound this quantity using moment generating functions. Let $s > 0$ be some parameter and notice

$$s\mathbb{E}[\max_i z_i] \leq \log\mathbb{E}[\exp(s\max_i z_i)] \leq \log\sum_i \mathbb{E}[\exp(sz_i)] \leq \log N + \log m(s)$$

where $m(s)$ is the *moment generating function*[2] (mgf) of the distribution $\mathcal{D}$ (or an upper bound thereof). The statement holds for any positive choice of s in the domain of $m(\cdot)$, hence we have

$$\mathbb{E}_{z\sim\mathcal{D}}[\Phi(z)] \leq \inf_{s>0}\frac{\log N + \log m(s)}{s}. \tag{3.15}$$

- **Laplace:** The mgf of the standard Laplace is $m(s) = \frac{1}{1-s}$. Choosing $s = \frac{1}{2}$ gives us that $\mathbb{E}[\max_i z_i] \leq 2\log 2N$.

- **Gumbel:** The mgf of the mean-zero Gumbel is $m(s) = \Gamma(1-s)e^{-s}$. Choosing $s = 1/2$ gives that $\mathbb{E}[\max_i z_i] \leq 2\log 2N$ since $m(0.5) < 2$.

[2]The mgf of a distribution $\mathcal{D}$ is the function $m(s) := \mathbb{E}_{X\sim\mathcal{D}}[\exp(sX)]$.

- **Gaussian:** The mgf of the standard Gaussian is $m(s) = \exp(s^2/2)$. Choosing $s = \sqrt{2\log N}$ gives $\mathbb{E}[\max_i z_i] \leq \sqrt{2\log N}$.

Theorem 3.15. *Let Φ be the baseline potential for the experts setting. Suppose we GBPA run with $\widetilde{\Phi}_t(\cdot) = \widetilde{\Phi}(\cdot; \mathcal{D}_\eta)$ for all t where the mean-zero distribution $\mathcal{D}$ is such that $\mathbb{E}_{z\sim\mathcal{D}}[\Phi(z)] \leq \alpha$ and $\forall G, \mathrm{Tr}(\nabla^2\widetilde{\Phi}(G)) \leq \beta/\eta$. Then we have*

$$\mathsf{Regret}_{\leq} \eta\alpha + \frac{\beta T}{\eta}.$$

Choosing η to optimize the bound gives $\mathsf{Regret}_{\leq} 2\sqrt{\alpha\beta T}$. In particular, for Laplace, (mean-zero) Gumbel and Gaussian perturbations, the regret bound becomes $2\sqrt{2T\log 2N}$, $2\sqrt{2T\log 2N}$ and $2\sqrt{2T\log N}$ respectively.

Proof. Result follows by plugging in bounds into Lemma 3.4. Mean-zero perturbations imply that the underestimation penalty is zero. The overestimation penalty is bounded by $\eta\alpha$ and the divergence penalty is bounded by $\beta T/\eta$ because of Lemma 3.14 and the assumption that $\|g_t\|_\infty \leq 1$. Our calculations above showed that for the Laplace, (mean-zero) Gumbel and Gaussian perturbations, we have $\alpha = 2\log 2N$, $2\log 2N$ and $\sqrt{2\log N}$ respectively. Furthermore, we have $\beta = 1, 1$ and $\sqrt{2\log N}$ respectively. □

3.6 Online PCA

3.6.1 Spectral Functions

Many important matrix functions, including matrix norms, are *spectral*, which means that they are symmetric functions of the eigenvalues. We say that F is a *spectral extension* of a vector function f, i.e., $F = f \circ \lambda$ for some $f : \mathbb{R}^N \to \mathbb{R}$. The spectral extension of the vector norm $\|\cdot\|_p$ is called the *Schattern-p norm*, denoted by $\|\cdot\|_{\lambda_p}$.

Spectral functions are invariant to unitary transformations, i.e.,

$$F(VAV^T) = F(A), \text{ for any unitary } V. \tag{3.16}$$

Furthermore, at points where f is differentiable, the gradient has the same eigenvectors as A:

$$\nabla(f \circ \lambda)(A) = U^T \nabla f(\mathrm{diag}(\lambda(A)))U, \tag{3.17}$$

where U is a unitary matrix such that $A = U\text{diag}(\lambda(A))U^T$. It follows that

$$\nabla F(VAV^T) = V\nabla F(A)V^T, \text{ for any unitary matrix } V. \tag{3.18}$$

For the proof, see (Lewis 1996, Corollary 3.14) or (Baes 2007, Corollary 31)).

3.6.2 FTPL with Gaussian Orthogonal Ensemble

Gaussian Orthogonal Ensemble (GOE) is a distribution over real symmetric matrices whose upper triangular entries are i.i.d. normal random variables with mean zero and variance $1/2$ and diagonal entries are i.i.d. standard normal (and also independent of the upper triangular entries). Alternatively, if Y is an $N \times N$ random matrix with i.i.d. Gaussian entries, then $(Y + Y^T)/\sqrt{2}$ follows GOE. The density measure μ of GOE on the space of real symmetric matrices can thus be written as

$$\mu_{\text{GOE}}(Z) \propto \exp\left(-\sum_{i<j} Z_{ij}^2 - \sum_i \frac{Z_{ii}^2}{2}\right),$$

Using the fact that $(ZZ^\top)_{ii} = \sum_{i,j} Z_{ij}^2$, we can express μ more concisely:

$$\mu_{\text{GOE}}(Z) = C\exp(-\text{Tr}(ZZ^\top)/2).$$

The above expression shows an extremely useful property of GOE that it is a *Unitary Invariant Ensemble (UIE)*: for any matrix A in its support and a unitary matrix U, the density is equal for A and U^TAU.

Theorem 3.16. *Let Φ be the baseline potential for the Online Dense PCA ($k = 1$). The GBPA run with $\widetilde{\Phi}(G) = \widetilde{\Phi}(G; \text{GOE}_\eta)$ for all t has regret $\sqrt{N}(\eta + \eta^{-1})\sum_{t=1}^T \|g_t\|_\infty^2$.*

Proof. Duchi et al. (2011, Lemma 9) can be easily generalized (See Appendix A.2) to an arbitrary norm to show that

$$\|\lambda(\nabla\widetilde{\Phi}(A) - \nabla\widetilde{\Phi}(B))\|_1 \le \eta^{-1}\|\lambda(A-B)\|_2 \le \eta^{-1}\sqrt{N}\|\lambda(A-B)\|_\infty \tag{3.19}$$

which implies that $D_{\widetilde{\Phi}}(G_t, G_{t-1}) \le \eta^{-1}\sqrt{N}$. Also, the GOE smoothing always overestimates $\widetilde{\Phi}$ because $\widetilde{\Phi}(A + \eta Z) \le \widetilde{\Phi}(A) + \eta\widetilde{\Phi}(Z)$. Thus, the overestimation penalty is at most $\eta\mathbb{E}_{Z\sim\text{GOE}}[\widetilde{\Phi}(Z)] = \eta\sqrt{N}$. By plugging into Lemma 3.4, we obtain the desired regret bound. □

3.7 Multi-Armed Bandits

3.7.1 Gradient-Based Prediction Algorithms for the Multi-Armed Bandits

We give a generic template for constructing MAB strategies in Algorithm 2, and we emphasize that this template can be viewed as a bandit reduction to the (full information) GBPA framework. Randomization is used for making decisions and for *estimating* the losses via importance sampling.

Require: fixed convex potential $\widetilde{\Phi} : \mathbb{R}^N \to \mathbb{R}$, with $\nabla\widetilde{\Phi} \subset \text{interior}(\Delta^N)$.
Require: Adversary selects (hidden) seq. of loss vectors $g_1, \ldots, g_T \in [-1, 0]^N$
Initialize: $\hat{G}_0 = 0$
for *t = 1 to T* **do**
  Sampling: Learner chooses i_t according to dist. $p(\hat{G}_{t-1}) = \nabla\widetilde{\Phi}(\hat{G}_{t-1})$
  Cost: Learner "gains" g_{t,i_t}, and observes this value
  Estimation: Learner produces estimate of gain vector, $\hat{G}_t := \frac{g_{t,i_t}}{p_{i_t}(\hat{G}_{t-1})}\mathbf{e}_{i_t}$
  Update: $\hat{G}_t = \hat{G}_{t-1} + \hat{G}_t$
end

Algorithm 2: GBPA Template for Multi-Armed Bandits.

Nearly all proposed methods have relied on this particular algorithmic blueprint. For example, the EXP3 algorithm of Auer et al. (2003) proposed a more advanced version of the Exponential Weights Algorithm (discussed in Section 3.5) to set the sampling distribution $p(\hat{G}_{t-1})$, where the only real modification is to include a small probability of uniformly sampling the arms.[3] But EXP3 more or less fits the template we propose in Algorithm 2 when we select $\widetilde{\Phi}(\cdot) = \mathbb{E}_{z\sim\text{Gumbel}}\Phi(G + \eta z)$. We elaborated on the connection between EWA and Gumbel perturbations in Section 3.5.

Lemma 3.17. *The baseline potential for this setting is* $\Phi(G) \equiv \max_i G_i$ *so that we can write the expected regret of GBPA*$(\widetilde{\Phi})$ *as*

$$\mathbb{E}\text{Regret}_T = \Phi(G_T) - \mathbb{E}[\textstyle\sum_{t=1}^T \langle \nabla\widetilde{\Phi}(\hat{G}_{t-1}), g_t\rangle].$$

[3]One of the conclusions we may draw from this section is that the uniform sampling of EXP3 is not necessary when we are only interested in expected-regret bounds and we focus on negative gains (that is, where $\hat{G}_t \in [-1, 0]^N$). It has been suggested that the uniform sampling may be necessary in the case of positive gains, although this point has not been resolved to the authors' knowledge.

Then, the expected regret of GBPA($\widetilde{\Phi}$) can be written as:

$$\mathbb{E}\mathsf{Regret}_T \leq \mathbb{E}_{i_1,\ldots,i_T}\Bigg[\underbrace{\Phi(\hat{G}_T) - \widetilde{\Phi}(\hat{G}_T)}_{\textit{underestimation penalty}} + \sum_{t=1}^{T}\underbrace{\mathbb{E}_{i_t}[D_{\widetilde{\Phi}}(\hat{G}_t, \hat{G}_{t-1})|\hat{G}_{t-1}]}_{\textit{divergence penalty}}\Bigg] + \underbrace{\widetilde{\Phi}(0) - \Phi(0)}_{\textit{overestimation penalty}} \tag{3.20}$$

where the expectations are over the sampling of $i_t, t = 1, \ldots, T$.

Proof. Let $\widetilde{\Phi}$ be a valid convex function for GBPA. Consider GBPA($\widetilde{\Phi}$) run on the loss sequence $g_1, \ldots, g_T$. The algorithm produces a sequence of estimated losses $\hat{G}_1, \ldots, \hat{G}_T$. Now consider GBPA-FI($\widetilde{\Phi}$), which is GBPA($\widetilde{\Phi}$) run with the full information on the deterministic loss sequence $\hat{G}_1, \ldots, \hat{G}_T$ (there is no estimation step, and the learner updates $\hat{G}_t$ directly). The regret of this run can be written as

$$\Phi(\hat{G}_T) - \textstyle\sum_{t=1}^{T}\langle \nabla\widetilde{\Phi}(\hat{G}_{t-1}), \hat{G}_t\rangle \tag{3.21}$$

and $\Phi(G_T) \leq \mathbb{E}[\Phi(\hat{G}_T)]$ by the convexity of Φ. □

It is clear that $\nabla\widetilde{\Phi}$ is in the probability simplex, and note that

$$\begin{aligned}\frac{\partial\widetilde{\Phi}}{\partial G_i} &= \mathbb{E}_{Z_1,\ldots,Z_N}\mathbf{1}\{G_i + Z_i > G_j + Z_j, \forall j \neq i\} \\ &= \mathbb{E}_{\tilde{G}_{j^*}}[\mathbb{P}_{Z_i}[Z_i > \tilde{G}_{j^*} - G_i]] = \mathbb{E}_{\tilde{G}_{j^*}}[1 - F(\tilde{G}_{j^*} - G_i)]\end{aligned} \tag{3.22}$$

where $\tilde{G}_{j^*} = \max_{j\neq i} G_j + Z_j$ and F is the cdf of Z_i. The unbounded support condition guarantees that this partial derivative is non-zero for all i given any G. So, $\widetilde{\Phi}(G; \mathcal{D})$ satisfies the requirements of Algorithm 2.

3.7.2 Differential Consistency

Recall that for the full information experts setting, if we have a uniform bound on the trace of $\nabla^2\widetilde{\Phi}$, then we immediately have a finite regret bound. In the bandit setting, however, the regret (Lemma 3.17) involves terms of the form $D_{\widetilde{\Phi}}(\hat{G}_{t-1} + \hat{G}_t, \hat{G}_{t-1})$, where the incremental quantity $\hat{G}_t$ can scale as large as *the inverse of the smallest probability* of $p(\hat{G}_{t-1})$. These inverse probabilities are essentially unavoidable, because unbiased estimates of a

quantity that is observed with only probability p must necessarily involve fluctuations that scale as $O(1/p)$.

Therefore, we need a stronger notion of smoothness that counters the $1/p$ factor in $\|\hat{G}_t\|$. We propose the following definition which bounds $\nabla^2\tilde{\tilde{\Phi}}$ in correspondence with $\nabla\tilde{\Phi}$.

Definition 3.18 (Differential Consistency)**.** *For constant $C > 0$, we say that a convex function $f(\cdot)$ is* C-differentially-consistent *if for all $G \in (-\infty, 0]^N$,*

$$\nabla^2_{ii} f(G) \leq C \nabla_i f(G).$$

In other words, the rate in which we decrease p_i should approach 0 as p_i approaches 0. This guarantees that the algorithm reduces the rate of exploration slowly enough. We later show that smoothings obtaining using perturbations with bounded hazard rate satisfy the differential consistency property introduced above (see Lemma 3.22).

We now prove a generic bound that we will use in the following two sections, in order to derive regret guarantees.

Theorem 3.19. *Suppose $\widetilde{\Phi}$ is C-differentially-consistent for constant $C > 0$. Then divergence penalty at time t in Lemma 3.17 can be upper bounded as:*

$$\mathbb{E}_{i_t}[D_{\widetilde{\Phi}}(\hat{G}_t, \hat{G}_{t-1})|\hat{G}_{t-1}] \leq \frac{NC}{2}.$$

Proof. For the sake of clarity, we drop the t subscripts on $\hat{G}$ and $\hat{G}$; we use $\hat{G}$ to denote the cumulative estimate $\hat{G}_{t-1}$, $\hat{G}$ to denote the marginal estimate $\hat{G}_t = \hat{G}_t - \hat{G}_{t-1}$, and g to denote the true loss g_t.

Note that by definition of Algorithm 2, $\hat{G}$ is a sparse vector with one non-zero (and negative) coordinate with value $\hat{G}_{i_t} = g_{t,i_t}/\nabla_{i_t}\widetilde{\Phi}(\hat{G})$. Plus, i_t is conditionally independent given $\hat{G}$. Now we can expand the expectation as

$$\begin{aligned}\mathbb{E}_{i_t}[D_{\widetilde{\Phi}}(\hat{G}+\hat{G}, \hat{G})|\hat{G}] &= \sum_i \mathbb{P}[i_t = i]\mathbb{E}[D_{\widetilde{\Phi}}(\hat{G}+\hat{G}, \hat{G})|\hat{G}, i_t = i] \\ &= \sum_i \nabla_i\widetilde{\Phi}(\hat{G})\mathbb{E}[D_{\widetilde{\Phi}}(\hat{G}+\hat{G}, \hat{G})|\hat{G}, i_t = i]. \qquad (3.23)\end{aligned}$$

For each term in the sum on the right hand side, the conditional expectation given $\hat{G}$ is

now,

$$\mathbb{E}[D_{\widetilde{\Phi}}(\hat{G}+\hat{G},\hat{G})|\hat{G},i_t=i] = D_{\widetilde{\Phi}}\left(\hat{G}+\frac{g_i}{\nabla_i\widetilde{\Phi}(\hat{G})}\mathbf{e}_i,\hat{G}\right) = \frac{g_i^2}{2(\nabla_i\widetilde{\Phi}(\hat{G}))^2}\nabla^2_{ii}\widetilde{\Phi}(J_i)$$

where J_i is some vector on the line segment joining $\hat{G}$ and $\hat{G}+\frac{g_i}{\nabla_i\widetilde{\Phi}(\hat{G})}\mathbf{e}_i$. Using differential consistency, we have $\nabla^2_{ii}\widetilde{\Phi}(J_i) \leq C\nabla_i\widetilde{\Phi}(J_i)$. Note that J_i agrees with $\hat{G}$ in all coordinates except coordinate i where it is at most $\hat{G}_i$. Note that this conclusion depends crucially on the *loss-only assumption* that $g_i \leq 0$. Convexity of $\widetilde{\Phi}$ guarantees that ∇_i is a non-decreasing function of coordinate i. Therefore, $\nabla_i\widetilde{\Phi}(J_i) \leq \nabla_i\widetilde{\Phi}(\hat{G})$. This means that

$$\mathbb{E}[D_{\widetilde{\Phi}}(\hat{G}+\hat{G},\hat{G})|\hat{G},i_t=i] \leq C\frac{g_i^2}{2(\nabla_i\widetilde{\Phi}(\hat{G}))^2}\nabla_i\widetilde{\Phi}(\hat{G}) \leq \frac{C}{2\nabla_i\widetilde{\Phi}(\hat{G})},$$

since $g_i^2 \leq 1$. Plugging this into (3.23), we get

$$\mathbb{E}_{i_t}[D_{\widetilde{\Phi}}(\hat{G}+\hat{G},\hat{G})|\hat{G}] \leq \sum_i \nabla_i\widetilde{\Phi}(\hat{G})\frac{C}{2\nabla_i\widetilde{\Phi}(\hat{G})} = \frac{NC}{2}. \qquad \square$$

3.7.3 Hazard Rate analysis

Despite the fact that perturbation-based multi-armed bandit algorithms provide a natural randomized decision strategy, they have seen little applications mostly because they are hard to analyze. But one should expect general results to be within reach: the EXP3 algorithm can be viewed through the lens of perturbations, where the noise is distributed according to the Gumbel distribution. Indeed, an early result of Kujala and Elomaa (2005) showed that a near-optimal MAB strategy comes about through the use of exponentially-distributed noise, and the same perturbation strategy has more recently been utilized in the work of Neu and Bartók (2013) and Kocák et al. (2014). However, a more general understanding of perturbation methods has remained elusive. For example, would Gaussian noise be sufficient for a guarantee? What about, say, the Weibull distribution?

In this section, we show that the performance of the GBPA($\widetilde{\Phi}(G;\mathcal{D})$) can be characterized by the *hazard function* of the smoothing distribution $\mathcal{D}$. The hazard rate is a standard tool in survival analysis to describe failures due to aging; for example, an increasing hazard rate models units that deteriorate with age while a decreasing hazard rate models units that improve with age (a counter intuitive but not illogical possibility). To the best of our knowledge, the connection between hazard rates and design of adversarial bandit

algorithms has not been made before.

Definition 3.20 (Hazard rate function)**.** *Assume we are given a distribution $\mathcal{D}$ whose probability density function is given by $\mu_{\mathcal{D}}$ and whose cumulative density function is given by $\varphi_{\mathcal{D}}$. The hazard rate function of $\mathcal{D}$ is*

$$\mathrm{haz}_{\mathcal{D}}(x) := \frac{\mu_{\mathcal{D}}(x)}{1 - \varphi_{\mathcal{D}}(x)}.$$

We will use $\mathrm{haz}_{\mathcal{D}}$ *without an argument to denote the supremal value obtained by* $\mathrm{haz}_{\mathcal{D}}$ *on its domain; we drop the subscript $\mathcal{D}$ when it is clear from the context.*

For the rest of the section, we assume that $F(x) < 1$ for all finite x, so that $\mathrm{haz}_{\mathcal{D}}$ is well-defined everywhere. This assumption is for the clarity of presentation but is not strictly necessary.

Theorem 3.21. *The regret of the GBPA for multi-armed bandits (Algorithm 2) with $\widetilde{\Phi}(G; \mathcal{D}_\eta) = \mathbb{E}_{Z_1,\ldots,Z_n \sim \mathcal{D}} \max_i \{G_i + \eta Z_i\}$ is at most:*

$$\underbrace{\eta \mathbb{E}_{Z_1,\ldots,Z_n \sim \mathcal{D}} \left[\max_i Z_i \right]}_{\textit{overestimation penalty}} + \underbrace{\frac{N \sup h_{\mathcal{D}}}{\eta} T}_{\textit{divergence penalty}}$$

Proof. Due to the convexity of Φ, the underestimation penalty is non-positive. The overestimation penalty is clearly at most $\mathbb{E}_{Z_1,\ldots,Z_n \sim D}[\max_i Z_i]$, and Lemma 3.22 proves the $N(\sup \mathrm{haz}_{\mathcal{D}})$ upper bound on the divergence penalty.

It remains to prove the tuning parameter η. Suppose we scale the perturbation Z by $\eta > 0$, i.e., we add ηZ_i to each coordinate. It is easy to see that $\mathbb{E}[\max_{i=1,\ldots,n} \eta X_i] = \eta \mathbb{E}[\max_{i=1,\ldots,n} X_i]$. For the divergence penalty, let F_η be the CDF of the scaled random variable. Observe that $F_\eta(t) = F(t/\eta)$ and thus $f_\eta(t) = \frac{1}{\eta} f(t/\eta)$. Hence, the hazard rate scales by $1/\eta$, which completes the proof. □

Lemma 3.22. *Consider implementing GBPA with potential function*

$$\widetilde{\Phi}(G; \mathcal{D}_\eta) = \mathbb{E}_{Z_1,\ldots,Z_n \sim \mathcal{D}} \max_i \{G_i + \eta Z_i\}.$$

The divergence penalty on each round is at most $N\mathrm{haz}_{\mathcal{D}}$.

Proof. Recall the gradient expression in Equation 3.22. We upper bound the i-th diagonal entry of the Hessian, as follows. First, let where $\tilde{G}_{j^*} = \max_{j \neq i} \{G_j + Z_j\}$ which is a random

variable independent of Z_i. Now,

$$\begin{aligned}
\nabla^2_{ii}\widetilde{\Phi}(G;\mathcal{D}_\eta) &= \frac{\partial}{\partial G_i}\mathbb{E}_{\tilde{G}_{j^*}}[1-F(\tilde{G}_{j^*}-G_i)] = \mathbb{E}_{\tilde{G}_{j^*}}\left[\frac{\partial}{\partial G_i}(1-F(\tilde{G}_{j^*}-G_i))\right] \\
&= \mathbb{E}_{\tilde{G}_{j^*}} f(\tilde{G}_{j^*}-G_i) \\
&= \mathbb{E}_{\tilde{G}_{j^*}}[h(\tilde{G}_{j^*}-G_i)(1-F(\tilde{G}_{j^*}-G_i))] \qquad (3.24) \\
&\leq (\sup h)\mathbb{E}_{\tilde{G}_{j^*}}[1-F(\tilde{G}_{j^*}-G_i)] \\
&= (\sup h)\nabla_i\widetilde{\Phi}(G).
\end{aligned}$$

We have just established that $\widetilde{\Phi}$ is differentially consistent with parameter $C = \mathrm{haz}_{\mathcal{D}}$. We apply Theorem 3.19 and the proof is complete. □

Corollary 3.23. *Algorithm 2 run with $\widetilde{\Phi}$ that is obtained by smoothing Φ using any of the distributions in Table 3.1 (restricted to a certain range of parameters) has an expected regret of order $O(\sqrt{TN\log N})$.*

Table 3.1: Distributions that give $O(\sqrt{TN\log N})$ regret FTPL algorithm. The parameterization follows Wikipedia pages for easy lookup. We denote the Euler constant (≈ 0.58) by γ_0.

Distribution	$\sup_x \mathrm{haz}_{\mathcal{D}}(x)$	$\mathbb{E}[\max_{i=1}^N Z_i]$	Parameters
Gumbel($\mu=1,\beta=1$)	1 as $x\to 0$	$\log N+\gamma_0$	N/A
Frechet ($\alpha>1$)	at most 2α	$N^{1/\alpha}\Gamma(1-1/\alpha)$	$\alpha=\log N$
Weibull($\lambda=1,k\leq 1$)	k at $x=0$	$O(\left(\frac{1}{k}\right)!(\log N)^{\frac{1}{k}})$	$k=1$
Pareto($x_m=1,\alpha$)	α at $x=0$	$\alpha N^{1/\alpha}/(\alpha-1)$	$\alpha=\log N$
Gamma($\alpha\geq 1,\beta$)	β as $x\to\infty$	$\log N+(\alpha-1)\log\log N-\log\Gamma(\alpha)+\beta^{-1}\gamma_0$	$\beta=\alpha=1$

CHAPTER IV

Follow the Perturbed Leader Analysis via Differential Privacy

The core idea from the previous chapter is that the second derivatives can be used as a measure of *stability* of online learning algorithms. Another area of research for which stability is a core component is *differential privacy* (DP). As Dwork and Roth (2014) observed, "differential privacy is enabled by stability and ensures stability". The natural robustness of DP algorithms has found many useful applications, most notably to preventing false discovery in statistical analysis (Bassily et al. 2016; Cummings et al. 2016; Dwork et al. 2015; Nissim and Stemmer 2015).

The utility of DP as a stability notion for analyzing *specific* online learning algorithms also has been noted before. The connection between Exponential Weights Algorithms and DP has been known since the early stages of DP literature; see, for example, Dwork and Roth (2014, Section 11.2). Dwork et al. (2014) showed that for online sparse PCA, FTPL with Gaussian Orthogonal Ensemble can be seen as an extension of Gaussian Mechanism, one of the two fundamental DP algorithms.

In this chapter, we generalize such observations and take a systematic approach to using the DP framework to design and analyze FTPL algorithms. We define the term *one-step privacy* as a relaxation of DP and show that it is a sufficient condition for low regret. Leveraging the powerful tools developed in the DP literature, we effortlessly derive generic first-order regret bounds for FTPL, which are notoriously hard to obtain.

We stress that this chapter does *not* study the design of low-regret algorithms that satisfy the privacy condition; indeed there is already substantial existing work along these lines (Agarwal and Singh 2017; Jain et al. 2012; Thakurta and Smith 2013; Tossou and Dimitrakakis 2017). Our goal is instead to show that, in and of itself, the DP methodology

is quite well-suited to design randomized learning algorithms with excellent guarantees.

4.1 Preliminaries

4.1.1 Differential Privacy

We introduce the basic definitions and properties of differential privacy (DP). For a more comprehensive overview, see an excellent survey by Dwork and Roth (2014). Following the convention in the privacy literature, we use the word *mechanism* to refer to a stochastic mapping.

We will define DP in terms of a distance measure between random distributions as in (Dwork et al. 2010).

Definition 4.1. *Let Y, Z be random variables taking values in $\mathbb{R}^N$. The* δ-approximate max divergence *of Y and Z is:*

$$D_\infty^\delta(Y, Z) = \sup_{B \subseteq \mathbb{R}^N : \mathbb{P}[Y \in B] > \delta} \log \frac{\mathbb{P}[Y \in B] - \delta}{\mathbb{P}[Z \in B]} \tag{4.1}$$

When $\delta = 0$, we drop the superscript δ.

Note that the max divergence is *not* a metric, because it is asymmetric and does not satisfy the triangle inequality.

Definition 4.2. *We say that a mechanism $\mathcal{M}$ is* (ϵ, δ)-differentially private (DP) *with respect to a set S if for every $a, a' \in \mathrm{dom}(\mathcal{M})$ such that $a' - a \in S$, we have*

$$D_\infty^\delta(\mathcal{M}(a), \mathcal{M}(a')) \leq \epsilon.$$

If $\mathcal{M}$ is $(\epsilon, 0)$-DP, we simply say it is ϵ-DP.

The DP definition requires a uniform bound on the max divergence. Because we primarily analyze DP mechanisms that take a vector as input, it is useful to allow the bound to scale in the distance between two inputs.

Definition 4.3 (Lipschitz Privacy)**.** *We say that a mechanism $\mathcal{M}$ is (ϵ, δ)-Lipschitz private with respect to a norm $\|\cdot\|$ (and S) if for all $a, a' \in \mathrm{dom}(\mathcal{M})$ (such that $a - a' \in S$),*

$$D_\infty^\delta(\mathcal{M}(a), \mathcal{M}(a')) \leq \epsilon \|a - a'\|.$$

Note that S is now an optional part of the definition.

An important property of DP is the *post-processing immunity*, which means that the composition of a DP mechanism with any function(s) is still DP.

Lemma 4.4 (Post-Processing Immunity)**.** *For any random variables Y and Z,*

$$D_\infty^\delta(f(Y), f(Z)) \leq D_\infty^\delta(Y, Z).$$

Dwork et al. 2010 provided the following alternative characterization of δ-approximate max divergence. The condition (iii) does not appear in the original statement, but follows from their proof.

Lemma 4.5. *(Dwork et al. 2010, Lemma 2.1.1) Let Y, Z be random variables over $\mathcal{B}$ with probability density function μ_Y, μ_Z respectively. Then, $D_\infty^\delta(Y, Z) \leq \epsilon$, if and only if there exits a random variable Y' such that*

(i) $\sup_{B \subseteq \mathcal{B}} |\mathbb{P}[Y \in B] - \mathbb{P}[Y' \in B]| \leq \delta$,

(ii) $D_\infty^\delta(Y', Z) \leq \epsilon$, *and*

(iii) $\mu_Y(b) \leq \mu_{Y'}(b)$ *if and only if* $\mu_Y(b) \leq e^\epsilon \mu_Z(b)$.

In short, we can alter Y into Y' by moving no more than δ probability mass from $\{b \in \mathcal{B} : \mu_Y(b) > e^\epsilon \mu_Z(b)\}$ to $\{b \in \mathcal{B} : e^\epsilon \mu_Y(b) < \mu_Z(b)\}$ such that $D_\infty(Y', Z)$ is bounded.

As an immediate corollary, we obtain the following characterization of DP that is useful when we want to directly analyze the ratios between probability density functions.

Theorem 4.6. *Let Y, Z be random variables. Then, $D_\infty^\delta(Y, Z) \leq \epsilon$ if and only if*

$$\mathbb{P}_{b \sim Y}\left[\log \frac{\mu_Y[b]}{\mu_Z[b]} > \epsilon\right] \leq \delta \tag{4.2}$$

where $\mathbb{P}$ is over the random sample b from $\mathcal{Y}$.

We now state a result showing that if Y and Z are close in max divergence then the expectations of a bounded function of Y and Z are also close.

Lemma 4.7. *Let Y and Z be random variables taking values in $\mathcal{B}$ such that $D_\infty^\delta(Y, Z) \leq \epsilon$. Then for any non-negative function $f : \mathcal{B} \to [0, F]$, we have*

$$\mathbb{E}[f(Y)] \leq e^\epsilon \mathbb{E}[f(Z)] + \delta F \tag{4.3}$$

Proof. Let $B = \{b \in \mathcal{B} : \mu_Y(b) > e^\epsilon \mathbb{P}[Z = b]\}$, and $B^C = \mathcal{B} - B$. Let Y' be the random variable satisfying the conditions of Lemma 4.5.

$$\begin{aligned}
\mathbb{E}[f(Y)] &= \int_S f(b)\mu_Y(b)db + \int_{B^C} f(b)\mu_Y(b)db \\
&= \int_B f(b)\left(\mu_Y(b) - \mu_{Y'}(b)\right)db + \int_B f(b)\mu_{Y'}(b)db + \int_{B^C} f(b)\mu_Y(b)db \\
&\leq F\left(\mathbb{P}[Y \in B] - \mathbb{P}[Y' \in B]\right) + \int_{\mathcal{B}} f(b)\mu_{Y'}(b)db \\
&\leq F\delta + \int_{\mathcal{B}} f(b)e^\epsilon \mu_Z(b)db = F\delta + e^\epsilon \mathbb{E}[Z]
\end{aligned}$$

where the second-to-last inequality is due to Lemma 4.5.(iii) and the last inequality is due to Lemma 4.5.(i) □

Lemma 4.8. *Let Y and Z be random variables taking values in $\mathcal{B}$ such that $D_\infty^\delta(Y, Z) \leq \epsilon$. Then for any bounded function $f : \mathcal{B} \to [-F, F]$, we have*

$$|\mathbb{E}[f(Y)] - \mathbb{E}[f(Z)]| \leq (e^\epsilon + \delta - 1)F. \tag{4.4}$$

Proof. The proof is very similar to that of Lemma 4.7 Let $B = \{b \in \mathcal{B} : \mathbb{P}[Y = b] > e^\epsilon \mathbb{P}[Z = b]\}$, and $B^C = \mathcal{B} - B$. Let Y' be the random variable satisfying the conditions of Lemma 4.5. As a shorthand notation, define $\mu_{Y-Z}(b) = \mu_Y(b) - \mu_Z(b)$.

$$\begin{aligned}
|\mathbb{E}[f(Y)] - \mathbb{E}[f(Z)]| &= \left|\int_B f(b)\left(\mu_{Y-Y'}(b) + \mu_{Y'-Z}(b) + \mu_{Y-Z}(b)\right)db\right| \\
&\leq \left|\int_B f(b)\mu_{Y-Y'}(b)db + \int_{\mathcal{B}} f(b)\mu_{Y'-Z}(b)db\right| \\
&\leq F\left|\int_B \mu_{Y-Y'}(b)db\right| + F\left|\int_{\mathcal{B}} \mu_{Y'-Z}(b)db\right| \\
&\leq F(\delta + e^\epsilon - 1).
\end{aligned}$$

□

This work focuses on DP mechanisms that add a random noise to the input. We introduce two such mechanisms that are fundamental building blocks in the DP literature. The proofs can be found in the Appendix.

Lemma 4.9 (Laplace Mechanism)**.** *Define $\mathcal{M} : \mathbb{R}^N \to \mathbb{R}^N$ such that $\mathcal{M}(a) = a + Z$ where $Z \sim \text{Lap}(\frac{u}{\epsilon})^N$ is a vector of N i.i.d. samples from Laplace distribution with scaling parameter u/ϵ. Then, $\mathcal{M}$ is ϵ-DP with respect to $\{a : \|a\|_1 \leq u\}$, and ϵ-Lipschitz private with respect to $\|\cdot\|_1$ and $\mathbb{R}^N$.*

Lemma 4.10 (Gaussian Mechanism)**.** *Define* $\mathcal{M} : \mathbb{R}^N \to \mathbb{R}^N$ *such that* $\mathcal{M}(\tilde{A}) = \tilde{A} + Z$ *where* Z *is a vector of* N *i.i.d. samples from Gaussian distribution with standard deviation* $\sigma = (\epsilon^{-1}u)2\log(2/\delta)I$*. Then,* $\mathcal{M}$ *is* (ϵ, δ)*-DP with respect to* $\{\tilde{A} : \|\tilde{A}\|_2 \leq u\}$*.*

Compared to the Laplace mechanism, the Gaussian mechanism permits a greater change in the input (as $\|a\|_2 \leq \|a\|_1$), but achieves a weaker privacy guarantee with $\delta > 0$.

4.1.2 One-Step Privacy

Note that we can write FTPL as a composition of a (stochastic) mechanism $\mathcal{M} : \mathbb{R}^N \to \mathbb{R}^N$ that adds a noise to the input, and a (deterministic) linear optimization oracle $\mathcal{O} : \mathbb{R}^N \to \mathcal{X}$:

$$x_t^{\text{FTPL}} = \mathcal{O}(\mathcal{M}(L_{t-1})),$$

Similarly, write BTPL as

$$x_t^{\text{BTPL}} = x_{t+1}^{\text{FTPL}} = \mathcal{O}(\mathcal{M}(L_{t-1} + \ell_t)),$$

Suppose that $\mathcal{M}$ is (ϵ, δ)-DP with respect to $\mathcal{Y}$. By the post processing immunity (Lemma 4.4), we have for every t that

$$\begin{aligned} D_\infty^\delta(x_t^{\text{FTPL}}, x_t^{\text{BTPL}}) =& D_\infty^\delta(\mathcal{O}(\mathcal{M}(L_{t-1})), \mathcal{O}(\mathcal{M}(L_t))) \\ =& D_\infty^\delta(\mathcal{M}(L_{t-1}), \mathcal{M}(L_t)) \leq \epsilon. \end{aligned}$$

This shows that x_t^{FTPL} and x_t^{BTPL} follow very similar distributions, which implies they will suffer similar total regret.

Note that this is *not* equivalent to saying that FTPL algorithm is DP, which would imply that the distribution over the whole sequence of outputs $x_{1:T}$ is robust against a small change in the loss sequence. The following definition of *one-step privacy* highlights this distinction.

Definition 4.11 (One-step privacy)**.** *An online learning algorithm is* (ϵ, δ)-one-step differentially private *if there exists* ϵ, δ *such that* $D_\infty^\delta(x_t, x_{t+1}) \leq \epsilon$ *for all* $t = 1, \ldots, T$ *given any loss sequence.*

One-step privacy is a powerful condition on the stability of FTPL (and online learning algorithms in general), from which we can derive generic regret bounds. The following

theorem, relating privacy to regret, provides a powerful tool which we further develop in this chapter.

Theorem 4.12. *If $\mathcal{A}$ is (ϵ, δ)-one-step DP for a loss-only OLO problem with $\epsilon \leq 1$, its expected regret is at most:*

$$2\epsilon L_T^* + 3\mathbb{E}[\mathsf{Regret}(\mathcal{A}^+)_T] + \delta \|\mathcal{X}\| \sum_{t=1}^T \|\ell_t\|_\star$$

where $\mathcal{A}^+$ is a fictitious algorithm plays at time t what $\mathcal{A}$ would play at time $t+1$.

Proof. Using Lemma 4.7, we have for every t,

$$\mathbb{E}[\langle x_t, \ell_t \rangle)] \leq e^\epsilon \mathbb{E}[\langle x_{t+1}, \ell_t \rangle] + \delta \|\mathcal{X}\| \|\ell_t\|_\star.$$

By summing over t, we have

$$\begin{aligned}\mathbb{E}\left[\sum_{t=1}^T \mathsf{Loss}(\mathcal{A})_t\right] &\leq e^\epsilon \mathbb{E}[\textstyle\sum_{t=1}^T \mathsf{Loss}(\mathcal{A}^+)_t] + \delta \sum_{t=1}^T \|\mathcal{X}\| \|\ell_t\|_\star \\ &\leq e^\epsilon (L_T^* + \mathbb{E}[\mathsf{Regret}(\mathcal{A}^+)_T]) + \delta \textstyle\sum_{t=1}^T \|\mathcal{X}\| \|\ell_t\|_\star.\end{aligned}$$

Subtract L_T^* from each side and get:

$$(e^\epsilon - 1) L_T^* + e^\epsilon \mathbb{E}[\mathsf{Regret}(\mathcal{A}^+)_T] + \delta \|\mathcal{X}\| \textstyle\sum_{t=1}^T \|\ell_t\|_\star.$$

To complete the proof, we use the trivial upper bounds $e^\epsilon \leq 1 + 2\epsilon \leq 3$, which hold for $\epsilon \leq 1$. □

In this chapter, we only consider the loss-only settings where the above result suffices. For completeness, however, we provide a similar statement for the loss/gain setting based on the additive bound from Lemma 4.8.

Theorem 4.13. *If an online learning algorithm $\mathcal{A}$ is (ϵ, δ)-uniformly-one-step DP for an OLO problem with $\epsilon < 1$, then its expected regret is at most:*

$$\textstyle\sum_{t=1}^T (2\epsilon + \delta) \|\mathcal{X}\| \|\ell_t\|_\star + \mathbb{E}[\mathsf{Regret}(\mathcal{A}^+)_T]$$

where $\mathcal{A}^+$ is a fictitious algorithm plays at time t what $\mathcal{A}$ would play at time $t+1$.

4.2 Generic Bounds

With the DP framework, we can improve Theorem 3.10 and establish that FTPL with Gaussian noise in fact enjoys the first-order regret bound that scales in L_T^* (disregarding logarithmic factors). Put differently, FTPL with Gaussian noise is able to *adapt* to the input if there is a strong signal for the best action, a property that was not discovered in previous analysis.

Theorem 4.14. *Consider a* loss-only *OLO problem. Let $R = \|\mathcal{X}\|_2\|\mathcal{Y}\|_2$. FTPL with Gaussian noise achieves expected regret of order $O\big(\sqrt[4]{N}\sqrt{RL_T^* \log T} + \sqrt{N}R\log T\big)$.*

Proof. Let $\sigma = \epsilon^{-1}\|\mathcal{Y}\|_2 2\log(2/\delta)$, where ϵ, δ will be determined later. By Lemma 4.10 and Lemma 4.4, FTPL with $\mathcal{N}(0, \sigma I)$ is (ϵ, δ)-one-step DP with respect to $\mathcal{Y}$. Also note that the regret bound for the Gaussian BTPL is $\sigma\|\mathcal{X}\|_2\sqrt{2N}$.

We now apply Theorem 4.12 and get the regret bound of:

$$2\epsilon L_T^* + 5\sigma\|\mathcal{X}\|_2\sqrt{N} + \delta\|\mathcal{X}\|_2\textstyle\sum_{t=1}^T \|\ell_t\|_2 \le 2\epsilon L_T^* + 10\epsilon^{-1}R\sqrt{N}\log(2/\delta) + \delta TR.$$

Set $\delta = (2TR)^{-1}$, so that the last term becomes a constant. Then, choose $\epsilon = \min(\sqrt[4]{N}\sqrt{(R\log T)/L_T^*}, 1)$. If $\epsilon = 1$, then we must have $L_T^* \le \sqrt{N}R\log T$, then, which gives $O(\sqrt{N}R\log T)$ regret. Otherwise, we obtain $O(\sqrt[4]{N}\sqrt{RL_T^*\log T})$ regret. □

When $L_T^* \ll \|\mathcal{Y}\|_2 T$, our bound is a major improvement over Theorem 3.10. For example, when $L_T^* = O(R\sqrt{T})$, our bound gives $O(R\sqrt[4]{NT})$.

4.3 Experts Problem

We first state our main result, which provides a generic sufficient condition for the distributions that FTPL can use to match the optimal first-order regret.

Theorem 4.15. *For the loss-only experts setting, FTPL with Laplace, Gumbel, Frechet, Weibull, and Pareto noise (i.i.d. for each of N coordinates), with a proper choice of distribution parameters, all achieve $O(\sqrt{L_T^* \log N} + \log N)$ expected regret.*

Although we are not the first to find FTPL with the above regret bound, L_T^* bound for FTPL with any of the mentioned noise is not found in the literature, with the exception of Gumbel noise that is equivalent to Exponential Weights. In fact, previous FTPL

algorithms with L_T^* regret bound all relied on one-sided perturbation that *subtract* from the cumulative loss; Kalai and Vempala (2005) used the negative exponential noise and Erven et al. (2014) used the dropout noise that is effectively a negative multinomial noise.

Symmetric distributions, on the other hand, were previously shown to achieve only $O(\sqrt{T})$ regret: such as Gaussian noise Section 3.5, random-walk noise (Devroye et al. 2013), and a large family of symmetric noises (Rakhlin et al. 2012). Our DP-based analysis shows that such discrepancy was merely due to the lack of proper analysis tools.

4.3.1 Connections between One-Step-Privacy and Bounded Hazard Rates

Let $\mathcal{D}$ be an absolutely continuous distribution over $\mathbb{R}$ with probability density function $\mu_{\mathcal{D}}$ and cumulative density function $\Phi_{\mathcal{D}}$. Let $\tilde{f}_{\mathcal{D}}$ and $\mathcal{M}_{\mathcal{D}}$ be functions from $\mathbb{R}^N$ to $\mathbb{R}$ defined as $\tilde{f}_{\mathcal{D}}(x) = \mathbb{E}[\max_{i\in[N]}(x_i + Z_i)]$ and $\mathcal{M}_{\mathcal{D}}(x) = x + Z$ respectively, where Z in both definitions is a vector of N i.i.d. samples from $\mathcal{D}$.

In Section 3.7, we showed that if $\text{haz}_{\mathcal{D}} \leq \epsilon$, then $\tilde{f}_{\mathcal{D}}$ is ϵ-differentially consistent. We extend this result to connect hazard rate to Lipschitz privacy.

Proposition 4.16. *If $\tilde{f}_{\mathcal{D}}$ is differentially consistent, then the mapping from $a \in \mathbb{R}^N$ to a random sample drawn from $\nabla\tilde{f}_{\mathcal{D}}(a)$ is ϵ-Lipschitz private with respect to $\|\cdot\|_1$.*

Proof. First, note that the second derivative vector $\nabla^2_{i\cdot}\tilde{f}_{\mathcal{D}} = (\nabla^2_{i1}\tilde{f}_{\mathcal{D}}, \ldots, \nabla^2_{iN}\tilde{f}_{\mathcal{D}})$ satisfies that the i-th coordinate is the only positive coordinate, and that its coordinates add up to 0. So, $\|\nabla^2_{i\cdot}\tilde{f}_{\mathcal{D}}\|_\infty = \nabla^2_{ii}\tilde{f}_{\mathcal{D}}$.

Define $q_i(u) = \nabla_i\tilde{f}_{\mathcal{D}}(a' + (a - a')u)$. Its derivative is

$$\begin{aligned}
q_i'(u) &= \langle\nabla^2_{i\cdot}\tilde{f}_{\mathcal{D}}(a + (a - a')u), a' - a\rangle \\
&\leq \|\nabla^2_{i\cdot}\tilde{f}_{\mathcal{D}}(a + (a - a')u)\|_\infty\|a' - a\|_1 \\
&\leq \nabla^2_{ii}\tilde{f}_{\mathcal{D}}(a + (a' - a)u)\|a' - a\|_1 \\
&\leq \epsilon\nabla_i\tilde{f}_{\mathcal{D}}(a + (a' - a)u)\|a' - a\|_1 \\
&= \epsilon q_i(u)\|a' - a\|_1
\end{aligned}$$

The last inequality is from our differential consistency assumption. It follows that for any $u \in [0, 1]$, we have

$$\frac{q_i'(u)}{q_i(u)} = \frac{d}{du}\log(q_i(u)) \leq \epsilon\|a' - a\|_1$$

and therefore

$$\ln \frac{\nabla_i \tilde{f}_{\mathcal{D}}(a)}{\nabla_i \tilde{f}_{\mathcal{D}}(a')} = \log q_i(1) - \log q_i(0) = \int_0^1 \frac{d}{du} \log(q_i(u))\, du \leq \epsilon \|a' - a\|_1. \qquad \square$$

Note that $\nabla \tilde{f}_{\mathcal{D}}(x)$ is always a probability vector and in particular, $\nabla_i \tilde{f}_{\mathcal{D}}(x)$ is the probability that FTPL algorithm would play $\mathbf{e}_i$ given a cumulative loss vector x. Thus the above proposition in fact proves that FTPL is one-step private.

Corollary 4.17. *If* $\text{haz}_{\mathcal{D}} \leq \epsilon$*, then FTPL with* $\mathcal{D}^N$ *(sampling* N *i.i.d. samples from* $\mathcal{D}$ *to generate noise) is* ϵ*-one-step Lipschitz private.*

4.3.2 Optimal Family of FTPL Algorithms

We will now prove Theorem 4.15. All listed distributions have max hazard rate of ϵ (for the parameter choice, see Table 3.1). From Corollary 4.17 and post-processing immunity (Lemma 4.4), we conclude that FTPL with any of the listed distributions is ϵ-Lipschitz private with respect to the L_1-norm. The loss set for experts setting, however, is bounded in the L_∞-norm.

To address this gap, we will show that from the privacy perspective, the worst case is when ℓ_t has only one non-zero element and thus $\|\ell_t\|_1 = \|\ell_t\|_\infty$. Note that in the experts setting, the output of FTPL is always a vertex of the simplex. Consider an arbitrary noise vector Z. If $L_{t,i} + Z_i < L_{t,j} + Z_j$, then $L_{t,i} + z_i < L_{t,j} + Z_j + \alpha$ for any $\alpha > 0$. So, $\{Z \in \mathbb{R}^N : \mathbf{e}_i = \mathcal{O}(L_t + Z)\} \subseteq \{Z \in \mathbb{R}^N : \mathbf{e}_i = \mathcal{O}(L_t + Z + \ell^{(-i)})\}$ for any loss vector $\ell^{(-i)} \in \mathcal{Y}$ whose i-th coordinate is zero. In other words, adding any loss to coordinates other than i can only increase the probability of playing $\mathbf{e}_i$. So, for any fixed $\ell_{1,t-1} \in \mathcal{Y}^{t-1}$,

$$\begin{aligned}
\sup_{i \in [N], \ell_t \in \mathcal{Y}} \frac{\mathbb{P}[x_t^{\text{FTPL}} = \mathbf{e}_i]}{\mathbb{P}[x_{t+1}^{\text{FTPL}} = \mathbf{e}_i]} &= \sup_{i \in [N]} \frac{\mathbb{P}[x_t^{\text{FTPL}} = \mathbf{e}_i]}{\inf_{\ell_t \in \mathcal{Y}} \mathbb{P}[x_{t+1}^{\text{FTPL}} = \mathbf{e}_i]} \\
&= \sup_{i \in [N]} \frac{\mathbb{P}[x_t^{\text{FTPL}} = \mathbf{e}_i]}{\inf_{\ell_t : \|\ell_t\|_1 \leq 1} \mathbb{P}[x_{t+1}^{\text{FTPL}} = \mathbf{e}_i]} \\
&= \sup_{\ell_t : \|\ell_t\|_1 \leq 1} \sup_{i \in [N]} \frac{\mathbb{P}[x_t^{\text{FTPL}} = \mathbf{e}_i]}{\mathbb{P}[x_{t+1}^{\text{FTPL}} = \mathbf{e}_i]}.
\end{aligned}$$

The BTPL regret is of order $(\log N)/\epsilon$ for all distributions. Applying Theorem 4.12 with $\epsilon = \min(\sqrt{\mathsf{Regret}(\mathsf{BTPL})_T}/L_T^*, 1)$ completes the proof.

4.4 Online PCA

The general intuition from the DP-based regret analysis is that in order to achieve one-step privacy with respect to a loss set bounded in some norm, our noise distribution's density function must decay exponentially in the same norm. This motivates our *Laplace-on-Diagonal Orthogonal Invariant Ensemble* (LOD). The LOD with scaling parameter $1/\epsilon$ has probability density function $p(Z) \propto \exp(-\epsilon\|\lambda(Z)\|_1)$.

Lemma 4.18. *LOD mechanism, defined as $\mathcal{M} : \mathbb{R}^{N\times N} \to \mathbb{R}^{N\times N}$ with $\mathcal{M}(A) = A + Z$, where Z is a sample from* $\mathrm{LOD}(u/\epsilon)$*, is ϵ-differentially private with respect to the set $\{X \in \mathbb{R}^{N\times N} : \|\lambda(X)\|_1 \leq u\}$.*

Proof. We will prove this by showing a generic reduction technique to the vector case. In particular, suppose that a distribution $\mathcal{D}$ over matrices has density function of the form $p(Z) = Cq(\|\lambda(Z)\|)$ for a normalizing constant C, arbitrary function of vectors q, and some norm $\|\cdot\|$. Then, we will show that the privacy guarantee of distribution $\mathcal{D}'$ over vectors whose density function is some constant times q extends to the matrices.

Let A, A', B be matrices. Then,

$$\frac{p(B-A)}{p(B-A')} = \frac{q(\|\lambda(B-A)\|)}{q(\|\lambda(B-A')\|)}.$$

By triangle inequality, $\|\lambda(B-A)\| - \|\lambda(B-A')\| \leq \|\lambda(A-A')\|$. So,

$$\sup_{\substack{A,A',B\in\mathbb{R}^{N\times N}\\ \|\lambda(A-A')\|\leq u}} \frac{p(B-A)}{p(B-A')} \leq \sup_{\substack{a,a'\in\mathbb{R}^N\\ \|a-a'\|\leq u}} \frac{q(a)}{q(a')}.$$

Hence, if adding a noise from $\mathcal{D}'$ achieves ϵ-DP with respect to a set of vectors bounded in $\|\cdot\|$, then adding a noise from $\mathcal{D}$ achieves ϵ-DP with respect to a set of matrices bounded in $\|\lambda(\cdot)\|$. □

To sample from LOD, first sample an orthogonal matrix U from N-dimensional Haar measure (uniform over all N-by-N orthogonal matrices), sample a vector Λ iid from Laplace distribution, and finally take $U^\top \Lambda U$. This requires $O(N^2 \log N)$ time using butterfly matrices Genz 1998 and it is performed only once; for oblivious adversaries, sampling once at the beginning and sampling fresh samples every round is equivalent as far as the expected regret is concerned (Kalai and Vempala 2005).

Theorem 4.19. *FTPL with LOD achieves* $O(\sqrt{L^* k \log(N/k)} + k \log(N/k))$ *expected regret on Online k-Sparce PCA.*

Proof. By post-processing immunity (Lemma 4.4), it follows from Lemma 4.18 that FTPL using LOD$(1/\epsilon)$ is one-step private for Online Sparse PCA. For the BTPL regret, we have an upper bound of $k(1 + \log N/k))/\epsilon$ (Appendix). To complete the proof, we apply Theorem 4.12 with $\epsilon = \min(\sqrt{k(1 + \log(N/k))/L^*}, 1)$. □

Applying LOD to the dense case, however, would result in an extra $\sqrt{N}$ factor in the regret bound, matching the regret bounds of rank-1 matrix perturbation algorithm by Garber et al. (2015). Instead, we use the *Gaussian-on-Diagonal Orthogonal Invariant Ensemble* (GOD) which has density function of

$$p(Z) \propto \exp(-\|\lambda(Z)\|_2^2/(2\sigma^2)).$$

Similarly to LOD, we can independently sample the eigenvectors from Haar measure and eigenvalues from multivariate Gaussian.

Theorem 4.20. *FTPL with GOD achieves* $O\left(\sqrt[4]{N \log N \log T}\sqrt{L_T^*} + \sqrt{N \log T \log N}\right)$ *regret on Online Dense PCA.*

Proof. Using the same arguments in the proof of Lemma 4.18 and the alternative characterization of (ϵ, δ)-DP (Theorem 4.6), we can extend the guarantees of the Gaussian mechanism (Lemma 4.10) to the matrices. It follows that FTPL with GOD with $\sigma = 2\sqrt{N}\log(2/\delta)/\epsilon$ is (ϵ, δ)-private with respect to $\{A \in \mathbb{R}^{N \times N} : \|\lambda(A)\|_2 \leq \sqrt{N}\}$, which contains the loss set $\mathcal{Y}$.

For the BTPL regret bound, we have $\mathbb{E}[\|\lambda(Z)\|_\infty] = O(\sigma\sqrt{\log N})$.

We apply Theorem 4.12 with $\delta = T^{-1}$, and $\epsilon = \min(\sqrt[4]{N \log T \log N}/\sqrt{L^*}, 1)$ to complete the proof. □

4.5 Adversarial Multi-Armed Bandits

In MAB, the standard importance sampling scheme for unbiased estimates has $(1/p_{t,i_t})$ scaling, which produces estimated loss vectors that are unbounded. The problem with applying the DP tools to MAB is that the DP mechanisms are originally designed to protect the privacy of an *individual* in the dataset. As a result, they are not well-suited

to provide stability guarantees when there is a large change in the input, let alone an unbounded change.

In this section we see two different methods to address this issue. The first method is mixing in the uniform distribution to lower bound $p_{t,i}$, which in turn upper bounds $(1/p_{t,i_t})$. The second method is changing the estimation scheme to produce a *biased* estimate.

For this section, we define additional notations as follows. For a fixed sequence of estimated losses $\hat{l}_{1:t}$, consider running only the *decision step* of a bandit algorithm $\mathcal{A}$ on $\hat{l}_{1:t}$ as if it is a full information (experts) setting; we use $\hat{\mathsf{Loss}}(\mathcal{A})_T$ to denote the loss accumulated in this run. For deriving an expected regret of $\mathcal{A}$, we would then take the expectation of $\hat{\mathsf{Loss}}(\mathcal{A})_T$ over all randomness in sampling i_t. Because i_t is conditionally independent given $i_{1:t-1}$, it suffices to consider each time step separately and sum over $t \in [T]$. For a more detailed exposition of this, see (Abernethy et al. 2012).

4.5.1 Mixing in Uniform Distributions

Given x_t as an output from an experts algorithm $\mathcal{A}$, our bandit algorithm $\bar{\mathcal{A}}_\gamma$ samples an arm i_t from

$$p_t = x_t(1-\gamma) + (\tfrac{1}{N}, \ldots, \tfrac{1}{N})\gamma \tag{4.5}$$

(for some $\gamma \geq 0$), and use the standard inverse propensity weighting estimation:

$$\hat{l}_t = \frac{\ell_{t,i_t}}{p_{t,i_t}} \mathbf{e}_{i_t}. \tag{4.6}$$

Theorem 4.21. *Assume $\mathcal{A}$ is an ϵ-one-step Lipschitz private algorithm for the experts setting. Then the bandit algorithm $\bar{\mathcal{A}}_\gamma$ for any $\gamma \geq \epsilon$ has expected regret at most*

$$\mathbb{E}R_T + \gamma T + \sum_{t=1}^{T} \epsilon \|\ell_t\|_2^2$$

where R_T is the regret of $\mathcal{A}^+$, a fictitious algorithm that plays at time t what $\mathcal{A}$ would play at time $t+1$, given full information of $\ell_{1:t}$.

Proof. Let $x_{1:t}$ be $\mathcal{A}$'s output, and $p_{1:t}$ be defined as in (4.5). First note that because $\hat{l}_t$ is an unbiased estimate of ℓ, $\mathbb{E}\mathsf{Loss}(\mathcal{A}_\gamma)_T = \mathbb{E}\hat{\mathsf{Loss}}(\mathcal{A}_\gamma)_T$. Hence, it is sufficient to consider the expected regret on the estimated loss sequence.

From the one-step privacy of $\mathcal{A}$, it follows $p_{t,i}/p_{t+1,i} \leq x_{t,i}/x_{t+1,i} \leq \exp(\epsilon\|\hat{l}_t\|_1)$, and

thus

$$p_{t+1,i} \geq \exp(-\epsilon\|\hat{l}_t\|_1)p_{t,i} \geq (1-\epsilon\|\hat{l}_t\|_1)p_{t,i}$$

where the last inequality follows from $\epsilon\|\hat{l}_t\|_1 \leq \epsilon/\gamma \leq 1$.

For any $\hat{l}_t$, we have:

$$\langle p_t - p_{t+1}, \hat{l}_t \rangle = \sum_{i=1}^N (p_{t,i} - p_{t+1,i})\hat{l}_{t,i} \leq \sum_{i=1}^N \epsilon p_{t,i}\hat{l}_{t,i}^2.$$

Taking the expectation of the above over $i \sim p_t$, the sum disappears because only the sample coordinate has non-zero value in $\hat{l}_t$, and we have

$$\sum_{i=1}^N p_{t,i}\left(\epsilon p_{t,i}\left(\frac{\ell_{t,i}}{p_{t,i}}\right)^2\right) = \sum_{i=1}^N \epsilon\ell_{t,i}^2 = \epsilon\|\ell_t\|_2^2.$$

We can thus conclude $\mathbb{E}[\hat{\mathsf{Loss}}(\bar{\mathcal{A}}_\gamma)_t - \hat{\mathsf{Loss}}(\bar{\mathcal{A}}_\gamma^+)_t]] \leq \epsilon\|\ell_t\|_2^2$. To complete the proof, we combine it with a trivial upper bound $\hat{\mathsf{Loss}}(\bar{\mathcal{A}}^+)_t \leq \hat{\mathsf{Loss}}(\mathcal{A}^+)_t + \gamma$. □

Applying Corollary 4.17, we obtain the following result. The regret bound and the distributions used are the same as Theorem 3.23, but the algorithms are different.

Corollary 4.22. *For MAB, FTPL with Laplace, Gumbel, Frechet, Weibull, and Pareto noise (i.i.d. for each of N coordinates) with a proper choice of distribution parameters (Table 3.1), uniform distribution mixing* (4.5)*, and estimation scheme* (4.6)*, all achieve an expected regret of order* $O\left(\sqrt{\log N \sum_{t=1}^T (\|\ell_t\|_2^2 + 1)}\right)$.

4.5.2 Biased Sampling

Neu (2015) proposed the FTPL-TRIX algorithm, which achieves first-order regret bound of $O(\sqrt{NL_T^* \log N})$ for combinatorial bandits problem, which includes MAB as a special case. We focus on the multi-armed bandits to simplify formulation so that it is easy to see how the DP framework is being applied. Our results in this section, however, can be extended to combinatorial settings.

The base algorithm for FTPL-TRIX is FTPL with a truncated Exponential distribution. From the privacy perspective, truncation of the noise distribution converts a ϵ-one-step DP algorithm to a weaker (ϵ, δ)-one-step DP algorithm. There is now a δ probability that the distribution over algorithm's prediction changes rapidly, in a multiplicative sense, within a single step.

The estimation scheme, on the other hand, may benefit from the fact that (ϵ, δ)-one-step DP allows to change from small $p_{t,i}$ to $p_{t+1,i} = 0$, which would avoid having large values of $\hat{l}_{t,i}$.

The δ parameter can be tuned to attain the optimal tradeoff between algorithmic stability and input stability, and it is indeed crucial in achieving the first-order regret bound. We formalize this intuition in the following theorem. The proof mostly follows (Neu 2015) and can be found in the Appendix.

Theorem 4.23. *Let $\mathcal{A}$ be an FTPL algorithm that is (ϵ, δ)-Lipschitz one-step privacy with respect to $\{\ell : \|\ell\|_1 \leq 1/\epsilon\}$. That is, there exists a set $\mathcal{Z}$ such that $\mathbb{P}[Z \notin \mathcal{Z}] \leq \delta$,*

$$\log \frac{\mathbb{P}[L_{t-1} + Z = b]}{\mathbb{P}[L_t + Z = b]} \leq \epsilon \|\ell_t\|_1 \text{ for all } Z \in \mathcal{Z}.$$

Furthermore, suppose $\sup_{Z \in \mathcal{Z}} \|Z\|_\infty \leq B$. Then, $\mathcal{A}$ applied to MAB with a biased *estimation scheme*

$$\hat{\ell}_{t,i_t} = \frac{\ell_{t,i_t}}{p_{t,i_t} + \epsilon} \tag{4.7}$$

has expected regret at most

$$\mathbb{E}\|Z\|_\infty + \delta T + (2\epsilon + \delta) N (L_T^* + B + \epsilon^{-1}).$$

Note that the existence of $\mathcal{Z}$ is simply a restated definition of (ϵ, δ)-Lipschitz privacy (Theorem 4.6) for FTPL. The only extra condition that the δ-probability event happens at the tail of the distribution is not prohibitive, as reasonable noise-adding private mechanisms should attempt to reduce the amount of noise and concentrate around zero. For example, the Laplace and Gaussian mechanism both satisfy this property.

Proof. If $Z \notin \mathcal{Z}$, then we use the trivial regret bound T. In expectation, this becomes the δT term. Let $\bar{\mathcal{A}}$ be the algorithm that resamples Z until $Z \in \mathcal{Z}$, and $p_1, \ldots, p_t$ be its output.

We will first show that the biased estimate $\hat{L}_{T,i}$ stays close to L_T^* for every i. Let τ be the last time in which i was chosen by $\bar{\mathcal{A}}$. Since $p_{\tau,i} > 0$, $\hat{L}_{\tau-1,i} \leq \hat{L}_{\tau-1,i^*} + B \leq \hat{L}_T^* + B$. Hence,

$$\hat{L}_{T,i} = \hat{L}_{\tau-1,i} + \hat{l}_{\tau,i} \leq \hat{L}_T^* + B + 1/\epsilon \leq L_T^* + B + 1/\epsilon. \tag{4.8}$$

The last inequality holds because $\hat{L}$ underestimates the true cumulative loss: $\mathbb{E}[\hat{l}_{t,i}] = p_{t,i} \frac{\ell_{t,i}}{p_{t,i} + \epsilon} \leq \ell_{t,i}$.

Now suppose the following chain of inequality holds true:

$$\mathbb{E}\mathsf{Loss}(\mathcal{A})_t = \mathbb{E}\hat{\mathsf{Loss}}(\mathcal{A})_t + \epsilon \sum_{i=1}^{N} \hat{l}_{t,i} \tag{4.9}$$

$$\leq \mathbb{E}\hat{\mathsf{Loss}}(\bar{\mathcal{A}})_t + (\epsilon + \delta) \sum_{i=1}^{N} \hat{l}_{t,i} \tag{4.10}$$

$$\leq \mathbb{E}\hat{\mathsf{Loss}}(\bar{\mathcal{A}}^+)_t + (2\epsilon + \delta) \sum_{i=1}^{N} \hat{l}_{t,i} \tag{4.11}$$

where the expectation is over algorithm's randomness.

Then, by summing over t and subtracting by L^*, we have

$$\mathbb{E}\mathsf{Regret}(\mathcal{A})_T \leq \mathsf{Regret}(\bar{\mathcal{A}}^+)_T + (2\epsilon + \delta) \sum_{i=1}^{N} \hat{L}_{T,i}.$$

Combined with (4.8), it follows

$$\mathbb{E}\mathsf{Regret}(\mathcal{A})_T \leq \mathsf{Regret}(\bar{\mathcal{A}}^+)_T + (2\epsilon + \delta)N(L_T^* + B + 1/\epsilon),$$

as desired.

It remains to prove the inequalities (4.9)-(4.11). We will first prove (4.9):

$$\begin{aligned}\mathbb{E}\hat{\mathsf{Loss}}(\mathcal{A})_t &= \sum_{i=1}^{N} p_{t,i}\hat{l}_{t,i} = p_{t,i_t}\frac{\ell_{t,i_t}}{p_{t,i_t}+\epsilon} = \ell_{t,i_t} - \epsilon\frac{\ell_{t,i_t}}{p_{t,i_t}+\epsilon} = \ell_{t,i_t} - \epsilon\hat{l}_{t,i_t} = \ell_{t,i_t} - \epsilon\sum_{i=1}^{N}\hat{l}_{t,i} \\ &= \mathbb{E}\mathsf{Loss}(\mathcal{A})_t - \epsilon\sum_{i=1}^{N}\hat{l}_{t,i}\end{aligned}$$

Next, we prove (4.10), which follows from the fact that $\mathcal{A}$ and $\bar{\mathcal{A}}$ differ with probability at most δ. When they differ with probability δ, the $\sum_{i=1}^{N}\hat{l}_{t,i}$ is the trivial bound for the difference in the losses, and therefore

$$\mathbb{E}\hat{\mathsf{Loss}}(\mathcal{A})_t - \mathbb{E}\hat{\mathsf{Loss}}(\bar{\mathcal{A}})_t \leq \delta\sum_{i=1}^{N}\hat{l}_{t,i}$$

Finally, we will prove (4.11) using the same argument as in the proof of Theorem 4.21:

$$\langle p_t - p_{t+1}, \hat{l}_t\rangle \leq \textstyle\sum_{i=1}^{N}\epsilon p_{t,i}\hat{l}_{t,i}^2 \leq \sum_{i=1}^{N}\epsilon\hat{l}_{t,i}$$

□

Corollary 4.24. *EXP3, with the change of its unbiased estimator to the biased estimator* (4.7)*, has regret of order* $O(\sqrt{L^* N \log N} + N \log N \log T))$.

Proof. Note that EXP3 uses FTRL with entropy regularizer as its subroutine, which is identical to FTPL with Gumbel distribution (Hofbauer and Sandholm 2002). The Gumbel distribution (with mean 0, $\beta = \epsilon$) has hazard rate at most ϵ, which makes it ϵ-Lipschitz private. Hence, we can arbitrarily choose B for applying Theorem 4.23. Set $B = (\log \frac{NT}{\sqrt{L^*}})/\epsilon$ and $\mathcal{Z} = \{Z : \|Z\|_\infty \leq B\}$. Note for each coordinate Z_i of Z, we have

$$\mathbb{P}[Z_i \geq B] = 1 - \exp(-\exp(-\epsilon B) \leq \exp(-\epsilon B) = \frac{\sqrt{L^*}}{NT}.$$

By union bound, $\mathbb{P}[Z \notin \mathcal{Z}] \leq N\mathbb{P}[Z_1 \geq B] \leq \sqrt{L^* T}$.

The BTPL regret is $(1 + \log N)/\epsilon$. By choosing $\epsilon = \sqrt{\log N}/\sqrt{L^* N}$, we obtain the claimed regret bound. □

Corollary 4.25. *FTPL with Gaussian noise with the biased estimator* (4.7) *has regret of order* $O(\sqrt{L^* N \log N} \sqrt[4]{\log T} + N \log T))$.

BIBLIOGRAPHY

BIBLIOGRAPHY

Abernethy, Jacob, Peter L. Bartlett, Alexander Rakhlin, and Ambuj Tewari (2008). "Optimal Stragies and Minimax Lower Bounds for Online Convex Games". In: *Conference on Learning Theory (COLT)*.

Abernethy, Jacob, Yiling Chen, and Jennifer Wortman Vaughan (2013). "Efficient Market Making via Convex Optimization, and a Connection to Online Learning". In: *ACM Transactions on Economics and Computation*.

Abernethy, Jacob, Elad Hazan, and Alexander Rakhlin (2012). "Interior-point methods for full-information and bandit online learning". In: *IEEE Transactions on Information Theory*.

Agarwal, Naman and Karan Singh (2017). "The Price of Differential Privacy For Online Learning". In: *International Conference on Machine Learning (ICML)*.

Allen-Zhu, Zeyuan and Yuanzhi Li (2017). "Follow the Compressed Leader: Faster Online Learning of Eigenvectors and Faster MMWU". In: *International Conference on Machine Learning (ICML)*.

Allenberg, Chamy, Peter Auer, Laszlo Gyorfi, and György Ottucsák (2006). "Hannan consistency in on-line learning in case of unbounded losses under partial monitoring". In: *International Conference on Algorithmic Learning Theory (ALT)*.

Auer, Peter, Nicolò Cesa-Bianchi, Yoav Freund, and Robert E. Schapire (2003). "The Nonstochastic Multiarmed Bandit Problem". In: *SIAM Journal of Computuataion*.

Baes, Michel (2007). "Convexity and differentiability properties of spectral functions and spectral mappings on Euclidean Jordan algebras". In: *Linear Algebra and its Applications* 422, pp. 664–700.

Bassily, Raef, Kobbi Nissim, Adam Smith, Thomas Steinke, Uri Stemmer, and Jonathan Ullman (2016). "Algorithmic Stability for Adaptive Data Analysis". In: *ACM Symposium on Theory of Computing (STOC)*.

Beck, Amir and Marc Teboulle (2012). "Smoothing and First Order Methods: A Unified Framework." In: *SIAM Journal on Optimization* 22.2, pp. 557–580.

Bertsekas, Dimitri P. (1973). "Stochastic optimization problems with nondifferentiable cost functionals". English. In: *Journal of Optimization Theory and Applications*.

Bubeck, Sébastien and Nicolò Cesa-Bianchi (2012). "Regret Analysis of Stochastic and Nonstochastic Multi-armed Bandit Problems". In: *Foundations and Trends in Machine Learning*.

Bubeck, Sébastien, Nicolò Cesa-Bianchi, and Sham Kakade (2012). "Towards minimax policies for online linear optimization with bandit feedback". In: *Conference on Learning Theory (COLT)*.

Cesa-Bianchi, Nicolò and Gábor Lugosi (2006). *Prediction, Learning, and Games*. Cambridge University Press. ISBN: 978-0-521-84108-5.

Cummings, Rachel, Katrina Ligett, Kobbi Nissim, Aaron Roth, and Zhiwei Steven Wu (2016). "Adaptive learning with robust generalization guarantees". In: *Conference on Learning Theory (COLT)*.

Devroye, Luc, Gábor Lugosi, and Gergely Neu (2013). "Prediction by random-walk perturbation". In: *Conference on Learning Theory (COLT)*.

Duchi, John, Peter L. Bartlett, and Martin J. Wainwright (2011). "Randomized smoothing for stochastic optimization". In: *arXiv preprint arXiv:1103.4296*.

Dwork, Cynthia, Vitaly Feldman, Moritz Hardt, Toniann Pitassi, Omer Reingold, and Aaron Leon Roth (2015). "Preserving statistical validity in adaptive data analysis". In: *ACM Symposium on Theory of Computing (STOC)*.

Dwork, Cynthia and Aaron Roth (2014). "The algorithmic foundations of differential privacy". In: *Foundations and Trends in Theoretical Computer Science*.

Dwork, Cynthia, Guy N Rothblum, and Salil Vadhan (2010). "Boosting and differential privacy". In: *IEEE Annual Symposium on Foundations of Computer Science (FOCS)*.

Dwork, Cynthia, Kunal Talwar, Abhradeep Thakurta, and Li Zhang (2014). "Analyze gauss: optimal bounds for privacy-preserving principal component analysis". In: *ACM Symposium on Theory of Computing (STOC)*.

Erven, Tim van, Wojciech Kotlowski, and Manfred K. Warmuth (2014). "Follow the leader with dropout perturbations". In: *Conference on Learning Theory (COLT)*.

Freund, Yoav and Robert E. Schapire (1997). "A Decision-Theoretic Generalization of On-Line Learning and an Application to Boosting". In: *Journal of Computer and System Sciences* 1.

Garber, Dan, Elad Hazan, and Tengyu Ma (2015). "Online learning of eigenvectors". In: *International Conference on Machine Learning (ICML)*.

Genz, Alan (1998). "Methods for generating random orthogonal matrices". In: *Monte Carlo and Quasi-Monte Carlo Methods*, pp. 199–213.

Gittins, John (1996). "Quantitative methods in the planning of pharmaceutical research". In: *Drug Information Journal* 30.2, pp. 479–487.

Glasserman, Paul (1991). *Gradient Estimation Via Perturbation Analysis*. Springer. ISBN: 9780792390954.

Hannan, James (1957). "Approximation to Bayes risk in repeated play". In: *Contributions to the Theory of Games*.

Hinton, Geoffrey E., Nitish Srivastava, Alex Krizhevsky, Ilya Sutskever, and Ruslan Salakhutdinov (2012). "Improving neural networks by preventing co-adaptation of feature detectors". In: *arXiv preprint arXiv:1207.0580*.

Hofbauer, Josef and William H. Sandholm (2002). "On the global convergence of stochastic fictitious play". In: *Econometrica*.

Jain, Prateek, Pravesh Kothari, and Abhradeep Thakurta (2012). "Differentially Private online learning". In: *Conference on Learning Theory (COLT)*.

Kalai, Adam and Santosh Vempala (2005). "Efficient algorithms for online decision problems". In: *Journal of Computer and System Sciences*.

Kocák, Tomáš, Gergely Neu, Michal Valko, and Remi Munos (2014). "Efficient learning by implicit exploration in bandit problems with side observations". In: *Neural Information Processing Systems (NIPS)*.

Kotłowski, Wojciech and Manfred K. Warmuth (2015). "PCA with Gaussian perturbations". In: *arXiv preprint arXiv:1506.04855*.

Kujala, Jussi and Tapio Elomaa (2005). "On following the perturbed leader in the bandit setting". In: *International Conference on Algorithmic Learning Theory (ALT)*. Springer.

Kwon, Joon and Vianney Perchet (2016). "Gains and losses are fundamentally different in regret minimization: The sparse case". In: *Journal of Machine Learning Research (JMLR)*.

Lewis, Adrian S. (1996). "Derivatives of Spectral Functions". In: *Mathematics of Operations Research* 3.

Littlestone, Nick and Manfred K. Warmuth (1994). "The Weighted Majority Algorithm". In: *Information and Computation*.

McMahan, H. Brendan (2011). "Follow-the-Regularized-Leader and Mirror Descent: Equivalence Theorems and L1 Regularization". In: *International Conference on Artificial Intelligence and Statistics (AISTATS)*, pp. 525–533.

Neu, Gergely (2015). "First-order regret bounds for combinatorial semi-bandits". In: *Conference on Learning Theory (COLT)*.

Neu, Gergely and Gábor Bartók (2013). "An efficient algorithm for learning with semi-bandit feedback". In: *International Conference on Algorithmic Learning Theory (ALT)*.

Nie, Jiazhong, Wojciech Kotłowski, and Manfred K. Warmuth (2013). "Online pca with optimal regrets". In: *International Conference on Algorithmic Learning Theory (ALT)*.

Nissim, Kobbi and Uri Stemmer (2015). "On the generalization properties of differential privacy". In: *arXiv preprint arXiv:1504.05800*.

Pacula, Maciej, Jason Ansel, Saman Amarasinghe, and Una-May O'Reilly (2012). "Hyperparameter tuning in bandit-based adaptive operator selection". In: *Applications of Evolutionary Computation*. Springer, pp. 73–82.

Rakhlin, Alexander, Ohad Shamir, and Karthik Sridharan (2012). "Relax and randomize: From value to algorithms". In: *Neural Information Processing Systems (NIPS)*.

Rockafellar, Ralph Tyrell (1997). *Convex Analysis*. Princeton University Press. ISBN: 9780691015866.

Shalev-Shwartz, Shai (2012). "Online Learning and Online Convex Optimization". In: *Foundations and Trends in Machine Learning* 4.2, pp. 107–194. ISSN: 1935-8237.

Simon, Marvin K (2002). *Probability distributions involving Gaussian random variables: A handbook for engineers and scientists*. Springer Science & Business Media.

Srebro, Nati, Karthik Sridharan, and Ambuj Tewari (2011). "On the Universality of Online Mirror Descent". In: *Neural Information Processing Systems (NIPS)*.

Thakurta, Abhradeep Guha and Adam Smith (2013). "(Nearly) optimal algorithms for private online learning in full-information and bandit settings". In: *Neural Information Processing Systems (NIPS)*.

Tossou, Aristide C. Y. and Christos Dimitrakakis (2017). "Achieving Privacy in the Adversarial Multi-Armed Bandit." In: *AAAI Conference on Artificial Intelligence (AAAI)*.

Van den Broeck, Guy, Kurt Driessens, and Jan Ramon (2009). "Monte-Carlo tree search in poker using expected reward distributions". In: *Advances in Machine Learning*. Springer, pp. 367–381.

Vovk, Vladimir (1998). "A game of prediction with expert advice". In: *Journal of Computer and System Sciences* 56.2, pp. 153–173.

Warmuth, Manfred K. (2009). *A perturbation that makes "Follow the leader" equivalent to "Randomized Weighted Majority"*. `http://classes.soe.ucsc.edu/cmps290c/Spring09/lect/10/wmkalai-rewrite.pdf`.

Warmuth, Manfred K. and Dima Kuzmin (2010). "Online Variance Minimization in $O(n^2)$ per Trial?" In: *Conference on Learning Theory (COLT)*.

Yousefian, Farzad, Angelia Nedić, and Uday V. Shanbhag (2010). "Convex nondifferentiable stochastic optimization: A local randomized smoothing technique". In: *Proceedings of American Control Conference (ACC)*.

Zinkevich, Martin (2003). "Online Convex Programming and Generalized Infinitesimal Gradient Ascent." In: *International Conference on Machine Learning (ICML)*.

APPENDICES

APPENDIX A

Omitted Proofs for Chapter III

A.1 Proof that the origin is the worst case (Lemma 3.13)

Proof. Let $\Phi(G) = \|G\|_2$ and η be a positive number. By continuity of eigenvectors, it suffices to show that the maximum eigenvalue of the Hessian matrix of the Gaussian smoothed potential $\widetilde{\Phi}(G; \eta, \mathcal{N}(0, I))$ is decreasing in $\|G\|$ for $\|G\| > 0$.

By Lemma 3.8, the gradient can be written as follows:

$$\nabla\Phi(G; \eta, \mathcal{N}(0, I)) = \frac{1}{\eta}\mathbb{E}_{u\sim\mathcal{N}(0,I)}[u\|G + \eta u\|] \tag{A.1}$$

Let u_i be the i-th coordinate of the vector u. Since the standard normal distribution is spherically symmetric, we can rotate the random variable u such that its first coordinate u_1 is along the direction of G. After rotation, the gradient can be written as

$$\frac{1}{\eta}\mathbb{E}_{u\sim\mathcal{N}(0,I)}\left[u\sqrt{(\|G\| + \eta u_1)^2 + \sum_{k=2}^{N}\eta^2 u_k^2}\right]$$

which is clearly independent of the coordinates of G. The pdf of standard Gaussian distribution has the same value at $(u_1, u_2, \ldots, u_n)$ and its sign-flipped pair $(u_1, -u_2, \ldots, -u_n)$. Hence, in expectation, the two vectors cancel out every coordinate but the first, which is along the direction of G. Therefore, there exists a function α such that $\mathbb{E}_{u\sim\mathcal{N}(0,I)}[u\|G + \eta u\|] = \alpha(\|G\|)G$.

Now, we will show that α is decreasing in $\|G\|$. Due to symmetry, it suffices to consider $G = te_1$ for $t \in \mathbb{R}^+$, without loss of generality. For any $t > 0$,

$$\begin{aligned}
\alpha(t) &= \mathbb{E}[u_1\sqrt{(t+\eta u_1)^2 + u_{\text{rest}}^2})]/t \\
&= \mathbb{E}_{u_{\text{rest}}}[\mathbb{E}_{u_1}[u_1\sqrt{(t+\eta u_1)^2 + b^2}|u_{\text{rest}} = b]]/t \\
&= \mathbb{E}_{u_{\text{rest}}}[\mathbb{E}_{a=\eta|u_1|}[a\Big(\sqrt{(t+a)^2+b^2} - \sqrt{(t-a)^2+B}\Big)|u_{\text{rest}} = b]]/t
\end{aligned}$$

Let $g(t) = \Big(\sqrt{(t+a)^2+B} - \sqrt{(t-a)^2+B}\Big)/t$. Take the first derivative with respect to t, and we have:

$$\begin{aligned}
g'(t) &= \frac{1}{t^2}\left(\sqrt{(t-a)^2+b^2} - \frac{t(t-a)}{\sqrt{(t+a)^2+b^2}} - \sqrt{(t+a)^2+b^2} + \frac{t(t-a)}{\sqrt{(t+a)^2+b^2}}\right) \\
&= \frac{1}{t^2}\left(\frac{a^2+b^2-at}{\sqrt{(t-a)^2+b^2}} - \frac{a^2+b^2+at}{\sqrt{(t+a)^2+b^2}}\right)
\end{aligned}$$

$$\Big((a^2+b^2)-at\Big)^2\Big((t+a)^2+b^2\Big) - \Big((a^2+b^2)+at\Big)^2\Big((t-a)^2+b^2\Big) = -4ab^2t^3 < 0$$

because t, η, u', B are all positive. So, $g(t) < 0$, which proves that α is decreasing in G.

The final step is to write the gradient as $\nabla(\widetilde{\Phi}; \eta, \mathcal{N}(0, I))(G) = \alpha(\|G\|)G$ and differentiate it:

$$\nabla^2 f_\eta(G) = \frac{\alpha'(\|G\|)}{\|G\|}GG^T + \alpha(\|G\|)I$$

The Hessian has two distinct eigenvalues $\alpha(\|G\|)$ and $\alpha(\|G\|) + \alpha'(\|G\|)\|G\|$, which correspond to the eigenspace orthogonal to G and parallel to G, respectively. Since α' is negative, α is always the maximum eigenvalue and it decreases in $\|G\|$. □

A.2 Proof of Equation 3.19

Duchi et al. (2011, Lemma 11) shows that the dual norm $\|\cdot\|_*$ on the left-hand side of Duchi et al. 2011, Equation 39 can be any norm such that f is L_0-Lipschitz with respect to $\|\cdot\|$. The rest of the proof for Duchi et al. 2011, Lemma 9 does not depend on the choice of norm. Since $\lambda_{\max}$ is 1-Lipschitz with respect to $\|\lambda(\cdot)\|_\infty$, we have

$$\|\lambda(\nabla\tilde{\lambda}_{\max}(A) - \nabla\tilde{\lambda}_{\max}(B))\|_1 \leq 1/\eta\|A-B\|_F \leq \sqrt{N}/\eta\|A-B\|_\infty.$$

APPENDIX B

Omitted Proofs for Chapter IV

B.1 Gaussian Mechanism

Theorem B.1.1. *The Gaussian mechanism with* $\sigma = 2\sqrt{\log(2/\delta)}u/\epsilon$ *satisfies* (ϵ, δ)*-privacy with respect to* $\{x : \|x\|_2 \leq u\}$.

The original proof of this theorem can be found in (Dwork and Roth 2014, Theorem A.1), but we include the full proof to use consistent notations as well as to use as a building block for proving Theorem B.1.2.

Proof. Note that due to the spherical symmetry of the normal distribution used for the Gaussian mechanism, it suffices to consider the one-dimensional case. See (Dwork and Roth 2014, Theorem A.1) for the full reduction.

We want to upper bound the following quantity:

$$\log \frac{\exp\left(-\frac{x^2}{2\sigma^2}\right)}{\exp\left(-\frac{(x+u)^2}{2\sigma^2}\right)} = \left|\frac{1}{2\sigma^2}(2xu + u^2)\right|. \tag{B.1}$$

This is bounded by ϵ whenever $x < \sigma^2\epsilon/u - u/2$. We use the tail bound

$$\mathbb{P}[x > t] \leq \exp\left(-\frac{t^2}{2\sigma^2}\right) \tag{B.2}$$

to bound the probability that the privacy loss is not bounded. In other words, we require that

$$\exp\left(-\frac{t^2}{2\sigma^2}\right) \le \delta/2 \iff t^2/(2\sigma^2) > \log(2/\delta)$$

Taking $t = \sigma^2\epsilon/u - u/2$ and setting $\sigma = cu/\epsilon$ we get

$$t^2/(2\sigma^2) = \frac{1}{2}(c^2 - \epsilon + \epsilon^2/(4c^2)) \ge \frac{1}{2}(c^2 - 1)$$

Hence, $c^2 = 4\log(2/\delta)$ satisfies the condition. □

Theorem B.1.2. *The Gaussian mechanism with $\sigma = \sqrt{3\log(2.5/\delta)}/\epsilon$ is (ϵ, δ)-continuous DP with respect to $\|\cdot\|_2$ and $\{x : \|x\|_2 \le 1/\epsilon\}$.*

In particular, if we write the Gaussian mechanism as $\mathcal{M}(a) = a + Z$, then for any $\|a - a'\|_2 \le 1/\epsilon$, we have

$$\log \frac{\mathbb{P}[a + Z = b]}{\mathbb{P}[a' + Z = b]} \le \epsilon\|a - a'\|_2$$

for all $\|Z\|_2 \le \frac{1}{\epsilon}(3\log(2.5/\delta) - 1/2)$, which occurs with probability at least $1 - \delta$.

Proof. From (B.1), it is bounded by ϵu whenever $x < \sigma^2\epsilon - u/2$. It is sufficient to bound the tail probability in the worst case when $u = \epsilon^{-1}$:

$$\mathbb{P}[x > \sigma^2\epsilon - 1/(2\epsilon)] \le \delta$$

Using the tail bound (B.2), it is sufficient to satisfy

$$\frac{1}{2\sigma^2}(\sigma^2\epsilon - 1/(2\epsilon))^2 \ge \log(2/\delta)$$

Setting $\sigma = c/\epsilon$,

$$\frac{1}{2c^2}(c^2 - \frac{1}{2})^2 \ge \log(2/\delta).$$

Since $\delta < 1$, $c^2 = 3\log(2.5/\delta) \ge 3\log(2/\delta) + \frac{1}{2}$ satisfies the above. □

B.2 BTPL Regret

All BTPL regret bounds proven in this paper are based on the following result by Kalai and Vempala (2005):

$$\mathbb{E}[\mathsf{Regret}(\mathsf{BTPL})_T] \le \mathbb{E}_{Z\sim\mathcal{D}} \sup_{x\in\mathcal{X}} \langle x, Z\rangle \le \|\mathcal{X}\| \mathbb{E}_{Z\sim\mathcal{D}}[\|Z\|_*].$$

Lemma B.2.1. *(Neu 2015, Lemma 10) Let $Z_1, \dots, Z_N$ be i.i.d. exponential random variables with unit expectation and let $Z_1^*, \dots, Z_N^*$ be their permutation in decreasing order. Then, for any $1 \le k \le N$,*

$$\mathbb{E}\left[\sum_{i=1}^{k} Z_i^*\right] \le k(\log(N/k) + 1).$$

Corollary B.2.2. *(Experts setting) BTPL algorithm with Laplace distribution with scaling parameter $1/\epsilon$ for the experts setting has expected regret at most $(\log(N) + 1)/\epsilon$.*

Corollary B.2.3. *(k-sparse Online PCA) BTPL algorithm with Laplace-on-Diagonal ensemble for the k-sparse Online PCA problem has expected regret at most $k(\log(N/k) + 1)/\epsilon$.*

Proof. For Online k-Sparse PCA problem, $\sup_{x\in\mathcal{X}} \langle x, Z\rangle$ is the is the sum of k-largest eigenvalues of Z. When $\mathcal{D}$ is LOD ensemble, These eigenvalues follow the Laplace distribution and therefore we can apply Lemma B.2.1. □

Lemma B.2.4. *(Dense PCA) Let $Z_1, \dots, Z_N$ be i.i.d. Gaussian random variables with zero mean and variance σ^2. Then,*

$$\mathbb{E}\left[\max_{i=1,\dots,N} Z_i\right] \le \sigma\sqrt{2\log N}.$$

Lemma B.2.5. *(General OLO) Let $Z_1, \dots, Z_N$ be i.i.d. Gaussian random variables with zero mean and unit variance. Then,*

$$\mathbb{E}[\|(Z_1, \dots, Z_N)\|_2] \le 2\sqrt{N}.$$

Proof. Note that $\|(Z_1, \dots, Z_N)\|_2$ is a χ-distributed random variable, which has mean $\sqrt{2}\Gamma((N+1)/2)/\Gamma(N/2) \le 2\sqrt{N}$. □

B.3 Multi-Armed Bandits

Corollary B.3.1. *FTPL with Laplace noise with biased estimator* (4.7) *has regret of order* $O(\sqrt{L^* N \log N} + N \log TN)$.

Proof. Consider the Laplace distribution with scaling factor $(1/\epsilon)$. When using Laplace noise, the algorithm is $(\epsilon, 0)$-one-step DP. Hence, we can arbitrarily set B for applying Theorem 4.23. Set $B = (\log \frac{NT}{\sqrt{L^*}})/\epsilon$, which gives $\delta = \sqrt{L^*}/T$ with union bound. The BTPL regret is $(1 + \log N)/\epsilon$. By choosing $\epsilon = \sqrt{\log N}/\sqrt{L^* N}$, we obtain the claimed regret bound. □

Corollary B.3.2. *FTPL with Gaussian noise with biased estimator* (4.7) *has regret of order* $O(\sqrt{L^* N \log N} \sqrt[4]{\log T} + N \log T))$.

Proof. Set $\sigma = \sqrt{\log \delta^{-1}}/\epsilon$. Then, Gaussian FTPL satisfies the conditions for Theorem 4.23 with $\delta = 1/T$ and $B = \sigma^2 \epsilon = \epsilon^{-1} \log T$ (See Appendix), and the BTPL regret is $\sigma\sqrt{\log N} = \sqrt{\log N}\sqrt{\log T}/\epsilon$. Set $\epsilon = \sqrt{\log N} \sqrt[4]{\log T}/\sqrt{L_T^* N}$ to get the claimed regret bound. □