10 The Convex Optimization Approach to Regret Minimization

Elad Hazan ehazan@ie.technion.ac.il
Technion - Israel Institute of Technology
Haifa, Israel

A well-studied and general setting for prediction and decision making is regret minimization in games. Recently the design of algorithms in this setting has been influenced by tools from convex optimization. In this chapter we describe the recent framework of online convex optimization which naturally merges optimization and regret minimization. We describe the basic algorithms and tools at the heart of this framework, which have led to the resolution of fundamental questions of learning in games.

10.1 Introduction

In the online decision making scenario, a player has to choose from a pool of available decisions and then incurs a loss corresponding to the quality of the decision made. The regret minimization paradigm suggests the goal of incurring an average loss which approaches that of the best fixed decision in hindsight. Recently tools from convex optimization have given rise to algorithms which are more general, unifying previous results and many times giving new and improved regret bounds.

In this chapter we survey some of the recent developments in this exciting merger of optimization and learning. We start by describing two general templates for producing algorithms and proving regret bounds. The templates are very simple, and unify the analysis of many previous well-known and frequently used algorithms (i.e., multiplicative weights and gradient de-

scent). For the setting of online linear optimization, we also prove that the two templates are equivalent.

After describing the framework and algorithmic templates, we describe some successful applications: characterization of regret bounds in terms of convexity of loss functions, bandit linear optimization, and variational regret bounds.

10.1.1 The Online Convex Optimization Model

In online convex optimization, an online player iteratively chooses a point from a set in Euclidean space denoted $\mathcal{K} \subseteq \mathbb{R}^n$. Following Zinkevich (2003), we assume that the set $\mathcal{K}$ is non-empty, bounded, and closed. For algorithmic efficiency reasons that will be apparent later, we also assume the set $\mathcal{K}$ to be convex.

We denote the number of iterations by T (which is unknown by the online player). At iteration t, the online player chooses $\mathbf{x}_t \in \mathcal{K}$. After committing to this choice, a convex cost function $\mathbf{f}_t : \mathcal{K} \mapsto \mathbb{R}$ is revealed. The cost incurred to the online player is the value of the cost function at the point she committed to $\mathbf{f}_t(\mathbf{x}_t)$. Henceforth we consider mostly *linear* cost functions, and abuse notation to write $\mathbf{f}_t(\mathbf{x}) = \mathbf{f}_t^\top \mathbf{x}$.

The feedback available to the player falls into two main categories. In the full information model, all information about the function $\mathbf{f}_t$ is observable by the player (after incurring the loss). In the "bandit" model, the player observes only the loss $\mathbf{f}_t(\mathbf{x}_t)$ itself.

The regret of the online player using algorithm $\mathcal{A}$ at time T is defined to be the total cost minus the cost of the best fixed single decision, where the best is chosen with the benefit of hindsight. We are usually interested in an upper bound on the worst-case guaranteed regret, denoted

$$\text{Regret}_T(\mathcal{A}) = \sup_{\{\mathbf{f}_1,\ldots,\mathbf{f}_t\}} \left\{ \mathbf{E}[\textstyle\sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t)] - \min_{\mathbf{x}\in\mathcal{K}} \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}) \right\}.$$

Regret is the defacto standard in measuring the performance of learning algorithms.[1]

Intuitively, an algorithm performs well if its regret is sublinear in T, that is, $\text{Regret}_T(\mathcal{A}) = o(T)$, since this implies that "on the average" the algorithm performs as well as the best fixed strategy in hindsight.

1. For some problems it is more natural to talk of the "payoff" given to the online player rather than the cost she incurs. If so, the payoff functions need to be concave and regret is defined analogously.

The running time of an algorithm for online game playing is defined to be the worst-case expected time to produce $\mathbf{x}_t$, for an iteration $t \in [T]$ [2] in a T iteration repeated game. Typically, the running time will depend on n, T and the parameters of the cost functions and underlying convex set.

10.1.2 Examples

10.1.2.1 Prediction from Experts Advice

Perhaps the best-known problem in prediction theory is the "experts problem". The decision maker has to choose from the advice of n given experts. After choosing one, a loss between zero and one is incurred. This scenario is repeated iteratively, and at each iteration the costs of the various experts are arbitrary. The goal is to do as well as the best expert in hindsight.

The online convex optimization problem captures this problem as a special case: the set of decisions is the set of all distributions over n elements (experts), that is the n-dimensional simplex $\mathcal{K} = \Delta_n = \{\mathbf{x} \in \mathbb{R}^n, \sum_i \mathbf{x}_i = 1\ , \mathbf{x}_i \geq 0\}$. Let the cost to the i'th expert at iteration t be denoted by $\mathbf{f}_t(i)$. Then the cost functions are given by $\mathbf{f}_t(x) = \mathbf{f}_t^\top \mathbf{x}$. This is the expected cost of choosing an expert according to distribution $\mathbf{x}$, and happens to be linear.

10.1.2.2 Online Shortest Paths

In the online shortest path problem, the decision maker is given a directed graph $G = (V, E)$ and a source-sink pair $s, t \in V$. At each iteration $t \in [T]$, the decision maker chooses a path $p_t \in \boldsymbol{P}_{s,t}$, where $\boldsymbol{P}_{s,t} \subseteq \{E\}^{|V|}$ is the set of all s, t-paths in the graph. The adversary independently chooses weights on the edges of the graph, given by a function from the edges to the reals $\mathbf{f}_t : E \mapsto \mathbb{R}$, which can be represented as a vector in m-dimensional space: $\mathbf{f}_t \in \mathbb{R}^m$. The decision maker suffers and observes loss, which is the weighted length of the chosen path $\sum_{e \in p_t} \mathbf{f}_t(e)$.

The discrete description of this problem as an experts problem, where we have an expert for every path, presents an efficiency challenge: there are potentially exponentially many paths in terms of the graph representation size. Much work has been devoted to resolving this efficiency issue, and efficient algorithms have been found in this discrete formulation, such as (Takimoto and Warmuth, 2003; Awerbuch and Kleinberg, 2008). However, the optimal regret bound for the bandit version of this problem eluded researchers for some time, and was finally resolved only within the online

2. Here and henceforth we denote the set of integers $\{1, ..., n\}$ by $[n]$.

convex optimization framework (Abernethy et al., 2008; Dani et al., 2008).

The online convex optimization framework suggests an inherently efficient model to capture this problem. Recall the standard description of the set of all distributions over paths (flows) in a graph as a convex set in $\mathbb{R}^m$, with $O(m + |V|)$ constraints. Denote this flow polytope by $\mathcal{K}$. The expected cost of a given flow $\mathbf{x} \in \mathcal{K}$ (distribution over paths) is then a linear function, given by $\mathbf{f}_t^\top \mathbf{x}$, where $\mathbf{f}_t(e)$ is the length of the edge $e \in E$.

10.1.2.3 Portfolio Selection

The universal portfolio selection problem which we briefly describe is due to Cover (1991). At each iteration $t = 1$ to T, the decision maker chooses a distribution of her wealth over n assets $\mathbf{x}_t \in \Delta_n$. The adversary independently chooses market returns for the assets, that is a vector $\mathbf{r}_t \in \mathbb{R}_+^n$ such that each coordinate $\mathbf{r}_t(i)$ is the price ratio for the i'th asset between the iterations t and $t + 1$. The ratio between the wealth of the investor at iterations $t+1$ and t is $\mathbf{r}_t^\top \mathbf{x}_t$, and hence the gain in this setting is defined to be the logarithm of this change ratio in wealth $\log(\mathbf{r}_t^\top \mathbf{x}_t)$. Notice that since $\mathbf{x}_t$ is the distribution of the investor's wealth, even if $\mathbf{x}_{t+1} = \mathbf{x}_t$, the investor may still need to trade in order to adjust for price changes.

The goal of regret minimization, which in this case corresponds to minimizing the difference $\max_{\mathbf{x} \in \Delta_n} \sum_{t=1}^T \log(\mathbf{r}_t^\top \mathbf{x}) - \sum_{t=1}^T \log(\mathbf{r}_t^\top \mathbf{x}_t)$, has an intuitive interpretation. The first term is the logarithm of the wealth accumulated by the distribution $\mathbf{x}^*$. Since this distribution is fixed, it corresponds to a strategy of rebalancing the position after every trading period, and hence is called a *constant rebalanced portfolio*. The second expression is the logarithm of the wealth accumulated by the online decision maker. Hence regret minimization corresponds to maximizing the ratio of investor wealth against wealth of the best benchmark from a pool of investing strategies.

A *universal* portfolio selection algorithm is defined to be one that attains regret converging to zero in this setting. Such an algorithm, albeit requiring exponential time, was first described in Cover (1991). The online convex optimization framework has given rise to much more efficient algorithms based on Newton's method (Hazan et al., 2007).

10.1.3 Algorithms for Online Convex Optimization

Algorithms for online convex optimization can be derived from rich algorithmic techniques developed for prediction in various statistical and machine learning settings. We describe two general algorithmic frameworks from which many previous algorithms can be derived as special cases.

Perhaps the most straightforward approach is for the online player to use whatever decision (point in the convex set) would have been optimal. Formally, let

$$\mathbf{x}_t = \arg\min_{\mathbf{x}\in\mathcal{K}} \sum_{i=1}^{t-1} \mathbf{f}_i(\mathbf{x}).$$

This type of strategy is known as "fictitious play" in economics, and was named "follow the leader" (FTL) by Kalai and Vempala (2005). As Kalai and Vempala point out, this strategy fails miserably in a worst-case sense. That is, its regret can be linear in the number of iterations, as the following example shows. Consider K to be the real line segment between -1 and +1, and $\mathbf{f}_0 = \frac{1}{2}\mathbf{x}$, and let $\mathbf{f}_i$ alternate between $-\mathbf{x}$ and $\mathbf{x}$. The FTL strategy will keep shifting between -1 and $+1$, always making the wrong choice.

Kalai and Vempala proceed to analyze a modification of FTL with added noise to "stabilize" the decision (this modification was originally due to Hannan (1957)). Similarly, much more general and varied twists on this basic FTL strategy can be conjured up, and, as we shall show, also analyzed successfully. This is the essence of the meta-algorithm defined in this section.

Another natural approach for online convex optimization is an iterative approach. Start with some decision $\mathbf{x} \in \mathcal{K}$, and iteratively modify it according to the cost functions that are encountered. Some natural update rules include the gradient update, updates based on a multiplicative rule, on Newton's method, and so forth. Indeed, all of these suggestions make for useful algorithms. But as we shall show, they can all be seen as special cases of the general methodology we analyze next.

10.2 The RFTL Algorithm and Its Analysis

Recall the caveat about straightforward use of follow-the-leader. As in the bad example we have considered, the prediction of FTL may vary wildly from one iteration to the next. This motivates the modification of the basic FTL strategy in order to stabilize the prediction. By adding a *regularization* term, we obtain the RFTL (regularized follow the leader) algorithm.

We proceed to formally describe the RFTL algorithmic template, and analyze it. While the analysis given is optimal asymptotically, we do not give the best constants possible, in order to simplify presentation.

In this section we consider only linear cost functions, $\mathbf{f}(\mathbf{x}) = \mathbf{f}^T\mathbf{x}$. The case of convex cost functions can be reduced to the linear case via the inequality $\mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}^*) \leq \nabla\mathbf{f}_t(\mathbf{x}_t)(\mathbf{x}_t - \mathbf{x}^*)$, and considering the function

$\hat{\mathbf{f}}_t(\mathbf{x}) = \nabla \mathbf{f}_t(\mathbf{x}_t)^\top \mathbf{x}$, which is now linear.

10.2.1 Algorithm Definition

The generic RFTL meta-algorithm is defined below. The regularization function $\mathcal{R}$ is assumed to be strongly convex and smooth such that it has a continuous second derivative.

Algorithm 10.1 RFTL

1: Input: $\eta > 0$, strongly convex regularizer function $\mathcal{R}$, and a convex compact set $\mathcal{K}$
2: Let $\mathbf{x}_1 = \arg\min_{\mathbf{x}\in\mathcal{K}} [\mathcal{R}(\mathbf{x})]$
3: **for** $t = 1$ to T **do**
4: Predict $\mathbf{x}_t$
5: Observe the payoff function $\mathbf{f}_t$
6: Update

$$\mathbf{x}_{t+1} = \arg\min_{\mathbf{x}\in\mathcal{K}} \underbrace{\left[\eta \sum_{s=1}^{t} \mathbf{f}_s^\top \mathbf{x} + \mathcal{R}(\mathbf{x})\right]}_{\Phi_t(\mathbf{x})} \tag{10.1}$$

7: **end for**

10.2.2 Special Cases: Multiplicative Updates and Gradient Descent

Two famous algorithms which are captured by algorithm 10.1 are called the multiplicative update algorithm and the gradient descent method. If $\mathcal{K} = \Delta_n = \{\mathbf{x} \geq 0 \ , \ \sum_i \mathbf{x}(i) = 1\}$, then taking $\mathcal{R}(\mathbf{x}) = \mathbf{x}\log\mathbf{x}$ gives a multiplicative update algorithm, in which

$$\mathbf{x}_{t+1}(i) = \frac{\mathbf{x}_t(i) \cdot e^{\eta \mathbf{f}_t(i)}}{\sum_{i=1}^{n} \mathbf{x}_t(i) \cdot e^{\eta \mathbf{f}_t(i)}}.$$

If $\mathcal{K}$ is the unit ball and $\mathcal{R}(\mathbf{x}) = \|\mathbf{x}\|_2^2$, we get the gradient descent algorithm, in which

$$\mathbf{x}_{t+1} = \frac{\mathbf{x}_t - \eta \mathbf{f}_t}{\|\mathbf{x}_t - \eta \mathbf{f}_t\|_2}.$$

It is possible to derive these special cases by the KKT optimality conditions of equation 10.1. However, we give an easier proof of these facts in the next section, in which we give an equivalent definition of RFTL for the case of linear cost functions.

10.2.3 The Regret Bound

Henceforth we make use of general matrix norms. A PSD matrix $A \succ 0$ gives rise to the norm $\|x\|_A = \sqrt{x^T A x}$. The *dual* norm of this matrix norm is $\|x\|_{A^{-1}} = \|x\|_A^*$. The generalized Cauchy-Schwartz theorem asserts that $x \cdot y \leq \|x\|_A \|y\|_A^*$. We usually take A to be the Hessian of the regularization function $\mathcal{R}(x)$, denoted $\nabla^2 \mathcal{R}(x)$. In this case, we shorthand the notation to be $\|x\|_{\nabla^2 \mathcal{R}(y)} = \|x\|_y$, and similarly $\|x\|_{\nabla^{-2} \mathcal{R}(y)} = \|x\|_y^*$. Denote

$$\lambda = \max_{t, \mathbf{x} \in \mathcal{K}} \mathbf{f}_t^{\intercal} [\nabla^2 \mathcal{R}(\mathbf{x})]^{-1} \mathbf{f}_t \quad , \quad D = \max_{\mathbf{u} \in \mathcal{K}} \mathcal{R}(\mathbf{u}) - \mathcal{R}(\mathbf{x}_1)$$

Notice that both λ and D depend on the regularization function, the convex decision set, and the magnitude of the cost functions.

Theorem 10.1. *Algorithm 10.1 achieves the following bound on the regret for every* $\mathbf{u} \in \mathcal{K}$*:*

$$\mathit{Regret}_T = \sum_{t=1}^{T} \mathbf{f}_t^{\top} (\mathbf{x}_t - \mathbf{u}) \leq 2\sqrt{2\lambda D T} \ .$$

Consider the expert problem, for example: the convex set is the simplex, $\mathcal{R}$ is taken to be the negative entropy function (which corresponds to the multiplicative update algorithm), and the costs are bounded by 1 in each coordinate. Then $\mathbf{f}^{\intercal} [\nabla^2 \mathcal{R}(\mathbf{x})]^{-1} \mathbf{f} = \sum_i \mathbf{f}(i)^2 \mathbf{x}(i) \leq \sum_i \mathbf{x}(i) = 1$, which implies $\lambda \leq 1$. The parameter D in this case is bounded by $\max_{\mathbf{u} \in \Delta} \sum_i \mathbf{u}(i) \log \frac{1}{\mathbf{u}(i)} \leq \log n$. This gives the regret bound $O(\sqrt{T \log n})$, which is known to be tight.[3]

To prove theorem 10.1, we first relate the regret to the stability in prediction. This is formally captured by the FTL-BTL lemma, which holds in the general scenario.

Lemma 10.2 (FTL-BTL lemma)**.** *For every* $\mathbf{u} \in \mathcal{K}$*, the algorithm defined by* (10.1) *enjoys the following regret guarantee*

$$\sum_{t=1}^{T} \mathbf{f}_t^{\intercal} (\mathbf{x}_t - \mathbf{u}) \quad \leq \quad \sum_{t=1}^{T} \mathbf{f}_t^{\intercal} (\mathbf{x}_t - \mathbf{x}_{t+1}) + \frac{1}{\eta} [\mathcal{R}(\mathbf{u}) - \mathcal{R}(\mathbf{x}_1)].$$

We defer the proof of this simple lemma to the appendix, and proceed with the (short) proof of the main theorem.

---

3. In the case of multiplicative updates, as well as in other regularization functions of interest, it is possible to obtain a tighter bound in theorem 10.1: the term λ can be redefined as $\lambda = \max_t \mathbf{f}_t^{\intercal} [\nabla^2 \mathcal{R}(\mathbf{x}_t)]^{-1} \mathbf{f}_t$. The derivation is not in the scope of this survey; see Abernethy et al. (2008) for more details.

Main Theorem. Recall that $\mathcal{R}(x)$ is a convex function and $\mathcal{K}$ is convex. Then, by Taylor expansion (with its explicit remainder term via the mean value theorem) at $\mathbf{x}_{t+1}$, there exists a $\mathbf{z}_t \in [\mathbf{x}_{t+1}, \mathbf{x}_t]$ for which

$$\begin{aligned}\Phi_t(\mathbf{x}_t) &= \Phi_t(\mathbf{x}_{t+1}) + (\mathbf{x}_t - \mathbf{x}_{t+1})^\intercal \nabla \Phi_t(\mathbf{x}_{t+1}) + \frac{1}{2}\|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2_{\mathbf{z}_t} \\ &\geq \Phi_t(\mathbf{x}_{t+1}) + \frac{1}{2}\|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2_{\mathbf{z}_t}\end{aligned}$$

Recall our notation $\|\mathbf{y}\|^2_{\mathbf{z}} = \mathbf{y}^\intercal \nabla^2 \Phi_t(\mathbf{z})\mathbf{y}$, and it follows that $\|\mathbf{y}\|^2_{\mathbf{z}} = \mathbf{y}^\intercal \nabla^2 \mathcal{R}(\mathbf{z})\mathbf{y}$. The inequality above is true because $\mathbf{x}_{t+1}$ is a minimum of Φ_t over $\mathcal{K}$. Thus,

$$\begin{aligned}\|\mathbf{x}_t - \mathbf{x}_{t+1}\|^2_{\mathbf{z}_t} &\leq 2\,\Phi_t(\mathbf{x}_t) - 2\,\Phi_t(\mathbf{x}_{t+1}) \\ &= 2\ (\Phi_{t-1}(\mathbf{x}_t) - \Phi_{t-1}(\mathbf{x}_{t+1})) + 2\eta \mathbf{f}_t^\intercal(\mathbf{x}_t - \mathbf{x}_{t+1}) \\ &\leq 2\,\eta\,\mathbf{f}_t^\intercal(\mathbf{x}_t - \mathbf{x}_{t+1}).\end{aligned}$$

By the generalized Cauchy-Schwartz inequality,

$$\begin{aligned}\mathbf{f}_t^\intercal(\mathbf{x}_t - \mathbf{x}_{t+1}) &\leq \|\mathbf{f}_t\|^*_{\mathbf{z}_t} \cdot \|\mathbf{x}_t - \mathbf{x}_{t+1}\|_{\mathbf{z}_t} \qquad \text{general CS} \\ &\leq \|\mathbf{f}_t\|^*_{\mathbf{z}_t} \cdot \sqrt{2\,\eta\,\mathbf{f}_t^\intercal(\mathbf{x}_t - \mathbf{x}_{t+1})}.\end{aligned} \tag{10.2}$$

Shifting sides and squaring, we get

$$\mathbf{f}_t^\intercal(\mathbf{x}_t - \mathbf{x}_{t+1}) \leq 2\,\eta\,\|\mathbf{f}_t\|^{*\,2}_{\mathbf{z}_t} \leq 2\,\eta\,\lambda.$$

Using this, together with the FTL-BTL lemma, and summing over T periods, we obtain the theorem. Choosing the optimal η, we obtain

$$R_T \leq \min_\eta \left\{ 2\,\eta\lambda\,T + \frac{1}{\eta}[\mathcal{R}(\mathbf{u}) - \mathcal{R}(\mathbf{x}_1)] \right\} \leq 2\sqrt{2\,D\,\lambda\,T}\ .$$

□

10.3 The "Primal-Dual" Approach

The other approach for proving regret bounds, which we call primal-dual, originates from the link-function methodology, as introduced in Grove et al. (2001); Kivinen and Warmuth (2001), and is related to the mirrored descent paradigm in the optimization community. A central concept useful for this method are Bregman divergences, formally defined below.

Definition 10.1. *Denote by $B^{\mathcal{R}}(\mathbf{x}||\mathbf{y})$ the Bregman divergence with respect to the function $\mathcal{R}$, defined as*

$$B^{\mathcal{R}}(\mathbf{x}||\mathbf{y}) = \mathcal{R}(\mathbf{x}) - \mathcal{R}(\mathbf{y}) - (\mathbf{x} - \mathbf{y})^\intercal \nabla \mathcal{R}(\mathbf{y})\ .$$

The primal-dual algorithm is an iterative algorithm, which computes the next prediction using a simple update rule and the previous prediction. The generality of the method stems from the update being carried out in a dual space, where the duality notion is defined by the choice of regularization.

Algorithm 10.2 Primal-dual

Let $\mathcal{K}$ be a convex set
Input: parameter $\eta > 0$, regularizer function $\mathcal{R}(\mathbf{x})$
for $t = 1$ to T **do**
  If $t = 1$, choose $\mathbf{y}_1$ such that $\nabla\mathcal{R}(\mathbf{y}_1) = \mathbf{0}$
  If $t > 1$, choose $\mathbf{y}_t$ such that
    Lazy version: $\nabla\mathcal{R}(\mathbf{y}_t) = \nabla\mathcal{R}(\mathbf{y}_{t-1}) - \eta\,\mathbf{f}_{t-1}$.
    Active version: $\nabla\mathcal{R}(\mathbf{y}_t) = \nabla\mathcal{R}(\mathbf{x}_{t-1}) - \eta\,\mathbf{f}_{t-1}$.
  Project according to $B^{\mathcal{R}}$:

$$\mathbf{x}_t = \arg\min_{\mathbf{x}\in\mathcal{K}} B^{\mathcal{R}}(\mathbf{x}||\mathbf{y}_t)$$

end for

10.3.1 Equivalence to RFTL in the Linear Setting

For the special case of linear cost functions, algorithm 10.2 (lazy version) and RFTL are identical, as we show now. The primal-dual algorithm, however, can be analyzed in a very different way, which is extremely useful in certain online scenarios.

Lemma 10.3. *For linear cost functions, the lazy primal-dual and RFTL algorithms produce identical predictions, that is,*

$$\arg\min_{x\in\mathcal{K}} \left(\mathbf{f}_t^{\intercal}\mathbf{x} + \frac{1}{\eta}\mathcal{R}(\mathbf{x}) \right) = \arg\min_{\mathbf{x}\in\mathcal{K}} B^{\mathcal{R}}(\mathbf{x}||\mathbf{y}_t) \ .$$

Proof. First, observe that the unconstrained minimum

$$\mathbf{x}_t^* \equiv \arg\min_{\mathbf{x}\in\mathbb{R}^n} \left\{ \sum_{s=1}^{t-1} \mathbf{f}_s^{\intercal}\mathbf{x} + \frac{1}{\eta}\mathcal{R}(\mathbf{x}) \right\}$$

satisfies

$$\sum_{s=1}^{t-1} \mathbf{f}_s + \frac{1}{\eta}\nabla\mathcal{R}(\mathbf{x}_t^*) = \mathbf{0} \ .$$

Since $\mathcal{R}(\mathbf{x})$ is strictly convex, there is only one solution for the above

equation and thus $\mathbf{y}_t = \mathbf{x}_t^*$. Hence,

$$\begin{aligned} B^{\mathcal{R}}(\mathbf{x}||\mathbf{y}_t) &= \mathcal{R}(\mathbf{x}) - \mathcal{R}(\mathbf{y}_t) - (\nabla\mathcal{R}(\mathbf{y}_t))^\intercal(\mathbf{x}-\mathbf{y}_t) \\ &= \mathcal{R}(\mathbf{x}) - \mathcal{R}(\mathbf{y}_t) + \eta \sum_{s=1}^{t-1} \mathbf{f}_s^\intercal(\mathbf{x}-\mathbf{y}_t) \ . \end{aligned}$$

Since $\mathcal{R}(\mathbf{y}_t)$ and $\sum_{s=1}^{t-1} \mathbf{f}_s^\intercal \mathbf{y}_t$ are independent of $\mathbf{x}$, it follows that $B^{\mathcal{R}}(\mathbf{x}||\mathbf{y}_t)$ is minimized at the point $\mathbf{x}$ that minimizes $\mathcal{R}(\mathbf{x}) + \eta \sum_{s=1}^{t-1} \mathbf{f}_s^\intercal \mathbf{x}$ over $\mathcal{K}$, which in turn implies that

$$\arg\min_{\mathbf{x}\in\mathcal{K}} B^{\mathcal{R}}(\mathbf{x}||\mathbf{y}_t) = \arg\min_{\mathbf{x}\in K} \left\{ \sum_{s=1}^{t-1} \mathbf{f}_s^\intercal \mathbf{x} + \frac{1}{\eta}\mathcal{R}(\mathbf{x}) \right\} .$$

□

10.3.2 Regret Bounds for the Primal-Dual Algorithm

Theorem 10.4. *Suppose that $\mathcal{R}$ is such that $B_{\mathcal{R}}(\mathbf{x},\mathbf{y}) \geq \|\mathbf{x}-\mathbf{y}\|^2$ for some norm $\|\cdot\|$. Let $\|\nabla \mathbf{f}_t(\mathbf{x}_t)\|^* \leq G_*$ for all t, and $\forall \mathbf{x} \in K$ $B_{\mathcal{R}}(\mathbf{x},\mathbf{x}_1) \leq D^2$. Applying the primal-dual algorithm (active version) with $\eta = \frac{D}{2G_*\sqrt{T}}$, we have*

$$Regret_T \leq DG_*\sqrt{T}$$

Proof. Since the functions $\mathbf{f}_t$ are convex, for any $\mathbf{x}^* \in K$,

$$\mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}^*) \leq \nabla \mathbf{f}_t(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*).$$

The following property of Bregman divergences follows easily from the definition: for any vectors $\mathbf{x}, \mathbf{y}, \mathbf{z}$,

$$(\mathbf{x}-\mathbf{y})^\top(\nabla\mathcal{R}(\mathbf{z}) - \nabla\mathcal{R}(\mathbf{y})) = B_{\mathcal{R}}(\mathbf{x},\mathbf{y}) - B_{\mathcal{R}}(\mathbf{x},\mathbf{z}) + B_{\mathcal{R}}(\mathbf{y},\mathbf{z}).$$

Combining both observations,

$$\begin{aligned} 2(\mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}^*)) &\leq 2\nabla \mathbf{f}_t(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{x}^*) \\ &= \frac{1}{\eta}(\nabla\mathcal{R}(\mathbf{y}_{t+1}) - \nabla\mathcal{R}(\mathbf{x}_t))^\top(\mathbf{x}^* - \mathbf{x}_t) \\ &= \frac{1}{\eta}[B_{\mathcal{R}}(\mathbf{x}^*,\mathbf{x}_t) - B_{\mathcal{R}}(\mathbf{x}^*,\mathbf{y}_{t+1}) + B_{\mathcal{R}}(\mathbf{x}_t,\mathbf{y}_{t+1})] \\ &\leq \frac{1}{\eta}[B_{\mathcal{R}}(\mathbf{x}^*,\mathbf{x}_t) - B_{\mathcal{R}}(\mathbf{x}^*,\mathbf{x}_{t+1}) + B_{\mathcal{R}}(\mathbf{x}_t,\mathbf{y}_{t+1})] \end{aligned}$$

where the last inequality follows from the generalized Pythagorean inequality (see Cesa-Bianchi and Lugosi (2006), lemma 11.3), as $\mathbf{x}_{t+1}$ is the projection

w.r.t the Bregman divergence of $\mathbf{y}_{t+1}$ and $\mathbf{x}^* \in K$ is in the convex set. Summing over all iterations,

$$\begin{aligned} 2\text{Regret} &\leq \frac{1}{\eta}[B_{\mathcal{R}}(\mathbf{x}^*, \mathbf{x}_1) - B_{\mathcal{R}}(\mathbf{x}^*, \mathbf{x}_T)] + \sum_{t=1}^{T} \frac{1}{\eta} B_{\mathcal{R}}(\mathbf{x}_t, \mathbf{y}_{t+1}) \\ &\leq \frac{1}{\eta} D^2 + \sum_{t=1}^{T} \frac{1}{\eta} B_{\mathcal{R}}(\mathbf{x}_t, \mathbf{y}_{t+1}). \end{aligned} \tag{10.3}$$

We proceed to bound $B_{\mathcal{R}}(\mathbf{x}_t, \mathbf{y}_{t+1})$. By definition of the Bregman divergence, and the dual norm inequality stated before,

$$\begin{aligned} B_{\mathcal{R}}(\mathbf{x}_t, \mathbf{y}_{t+1}) + B_{\mathcal{R}}(\mathbf{y}_{t+1}, \mathbf{x}_t) &= (\nabla\mathcal{R}(\mathbf{x}_t) - \nabla\mathcal{R}(\mathbf{y}_{t+1}))^\top (\mathbf{x}_t - \mathbf{y}_{t+1}) \\ &= 2\eta \nabla \mathbf{f}_t(\mathbf{x}_t)^\top (\mathbf{x}_t - \mathbf{y}_{t+1}) \\ &\leq \eta^2 {G^*}^2 + \|\mathbf{x}_t - \mathbf{y}_{t+1}\|^2. \end{aligned}$$

Thus, by our assumption $B_{\mathcal{R}}(\mathbf{x}, \mathbf{y}) \geq \|\mathbf{x} - \mathbf{y}\|^2$, we have

$$B_{\mathcal{R}}(\mathbf{x}_t, \mathbf{y}_{t+1}) \leq \eta^2 G_*^2 + \|\mathbf{x}_t - \mathbf{y}_{t+1}\|^2 - B_{\mathcal{R}}(\mathbf{y}_{t+1}, \mathbf{x}_t) \leq \eta^2 G_*^2.$$

Plugging back into equation (10.3), and by non-negativity of the Bregman divergence, we get

$$\text{Regret} \leq \frac{1}{2}[\frac{1}{\eta} D^2 + \eta T {G_*}_2] \leq D G_* \sqrt{T}$$

by taking $\eta = \frac{D}{2\sqrt{T} G_*}$

□

10.3.3 Deriving the Multiplicative Update and Gradient Descent Algorithms

We stated in section 10.3.2 that by taking $\mathcal{R}$ to be the negative entropy function over the simplex, the RFTL template specializes to become a multiplicative update algorithm. Since we have proved that RFTL is equivalent to the primal-dual algorithm, the same is true for the latter, and the same regret bound applies.

If $\mathcal{R}(\mathbf{x}) = \mathbf{x} \log \mathbf{x}$ is the negative entropy function, then $\nabla\mathcal{R}(\mathbf{x}) = \mathbf{1} + \log \mathbf{x}$, and hence the update rule for the primal-dual algorithm 10.2 (the lazy and adaptive versions are identical in this case) becomes

$$\log \mathbf{y}_t = \log \mathbf{x}_{t-1} - \eta \mathbf{f}_{t-1}$$

or $\mathbf{y}_t(i) = \mathbf{x}_{t-1}(i) \cdot e^{-\eta \mathbf{f}_{t-1}(i)}$. Since the entropy projection corresponds to scaling by the ℓ_1-norm, it follows that $\mathbf{x}_{t+1}(i) = \frac{\mathbf{x}_t(i) \cdot e^{\eta \mathbf{f}_t(i)}}{\sum_{i=1}^{n} \mathbf{x}_t(i) \cdot e^{\eta \mathbf{f}_t(i)}}$.

As for the regret bound, it is well-known that the entropy function satisfies that $B_R(\mathbf{x}, \mathbf{y}) \geq \frac{1}{4}\|\mathbf{x} - \mathbf{y}\|_1^2$ (which is essentially Pinsker's inequality (see (Cover and Thomas, 1991)). Thus, to apply theorem 10.4, we need to bound the the ℓ_∞-norm of the gradients, which corresponds to the maximal cost incurred by the experts. Assume this is bounded by 1, that is, $G_* \leq 1$. The Bregman divergence with respect to $\mathcal{R}$ is the relative entropy, and starting from $\mathbf{x}_1$ being the uniform distribution, it holds that $B_{\mathcal{R}}(\mathbf{x}, \mathbf{x}_1) \leq \log n$ for any $\mathbf{x}$ in the simplex. Thus, by theorem 10.4 the regret of the multiplicative weights algorithm for the experts problem is bounded by $O(\sqrt{T \log n})$.

To derive the online gradient descent algorithm, take $\mathcal{R} = \frac{1}{2}\|\mathbf{x}\|_2^2$. In this case, $\nabla \mathcal{R}(\mathbf{x}) = \mathbf{x}$, and hence the update rule for the primal-dual algorithm 10.2 becomes

$$\mathbf{y}_t = \mathbf{y}_{t-1} - \eta \mathbf{f}_{t-1},$$

and thus when $\mathcal{K}$ is the unit ball,

$$\mathbf{x}_{t+1} = \frac{\mathbf{x}_1 - \eta \sum_{\tau=2}^{t} \mathbf{f}_\tau}{\|\mathbf{x}_1 - \eta \sum_{\tau=2}^{t} \mathbf{f}_\tau\|_2} = \frac{\mathbf{x}_t - \eta \mathbf{f}_t}{\|\mathbf{x}_t - \eta \mathbf{f}_t\|_2}.$$

10.4 Convexity of Loss Functions

In this section we review one of the first consequences of the convex optimization approach to decision making: the characterization of attainable regret bounds in terms of convexity of loss functions. It has long been known that special kinds of loss functions permit tighter regret bounds than other loss functions. For example, in the portfolio selection problem, Cover's algorithm attained regret which depends on the number of iterations T as $O(\log T)$. This is in contrast to online linear optimization or the experts problem, in which $\Theta(\sqrt{T})$ is known to be tight.

In this section we give a simple gradient descent-based algorithm which attains logarithmic regret if the loss functions are *strongly convex*. Interestingly, the naive fictitious play (FTL) algorithm attains essentially the same regret bounds in this special case. Similar bounds are attainable under weaker conditions on the loss functions, which capture the portfolio selection problem, and have led to the efficient algorithm for Cover's problem (Hazan et al., 2007).

We say that a function is α-strongly convex if its second derivative is strictly bounded away from zero. In higher dimensions this corresponds to the matrix inequality $\nabla^2 \mathbf{f}(\mathbf{x}) \succeq \alpha \cdot \mathbb{I}$, where $\nabla^2 \mathbf{f}(\mathbf{x})$ is the Hessian of the function and $A \succeq B$ denotes that the matrix $A - B$ is positive semi-definite.

For example, the squared loss, that is, $\mathbf{f}(\mathbf{x}) = \|\mathbf{x}-\mathbf{a}\|_2^2$, is 1-strongly convex.

Algorithm 10.3 Online gradient descent

1: Input: convex set $\mathcal{K}$, initial point $\mathbf{x}_0 \in \mathcal{K}$, learning rates $\eta_1, ..., \eta_t$
2: **for** $t = 1$ to T **do**
3: Let $\mathbf{y}_t = \mathbf{x}_{t-1} - \eta_{t-1}\nabla \mathbf{f}_{t-1}(\mathbf{x}_{t-1})$
4: Project onto $\mathcal{K}$:

$$\mathbf{x}_t = \arg\min_{\mathbf{x}\in\mathcal{K}} \|\mathbf{x} - \mathbf{y}_t\|_2$$

5: **end for**

The following theorem, proved in Hazan et al. (2007), establishes logarithmic bounds on the regret if the cost functions are strongly convex. Denote by G an upper bound on the Euclidean norm of the gradients.

Theorem 10.5. *The online gradient descent algorithm with stepsizes* $\eta_t = \frac{1}{\alpha t}$ *achieves the following guarantee for all* $T \geq 1$*:*

$$\mathit{Regret}_T(OGD) \;\leq\; \frac{G^2}{2\alpha}(1+\log T).$$

Proof. Let $\mathbf{x}^* \in \arg\min_{\mathbf{x}\in\boldsymbol{P}} \sum_{t=1}^T f_t(\mathbf{x})$. Recall the definition of regret:

$$\mathrm{Regret}_T(OGD) \;= \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=1}^T \mathbf{f}_t(\mathbf{x}^*).$$

Denote $\nabla_t \triangleq \nabla \mathbf{f}_t(\mathbf{x}_t)$. By α-strong convexity, we have

$$\begin{aligned} f_t(\mathbf{x}^*) &\geq f_t(\mathbf{x}_t) + \nabla_t^\top(\mathbf{x}^* - \mathbf{x}_t) + \frac{\alpha}{2}\|\mathbf{x}^* - \mathbf{x}_t\|^2 \\ 2(f_t(\mathbf{x}_t) - f_t(\mathbf{x}^*)) &\leq 2\nabla_t^\top(\mathbf{x}_t - \mathbf{x}^*) - \alpha\|\mathbf{x}^* - \mathbf{x}_t\|^2. \end{aligned} \tag{10.4}$$

Following Zinkevich's analysis, we upper-bound $\nabla_t^\top(\mathbf{x}_t - \mathbf{x}^*)$. Using the update rule for $\mathbf{x}_{t+1}$ and the generalized Pythagorian inequality (Cesa-Bianchi and Lugosi (2006), lemma 11.3), we get

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 \;=\; \|\Pi(\mathbf{x}_t - \eta_{t+1}\nabla_t) - \mathbf{x}^*\|^2 \leq \|\mathbf{x}_t - \eta_{t+1}\nabla_t - \mathbf{x}^*\|^2.$$

Hence,

$$\|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2 \quad\leq\quad \|\mathbf{x}_t - \mathbf{x}^*\|^2 + \eta_{t+1}^2\|\nabla_t\|^2 - 2\eta_{t+1}\nabla_t^\top(\mathbf{x}_t - \mathbf{x}^*).$$

Then, shifting sides,

$$2\nabla_t^\top(\mathbf{x}_t - \mathbf{x}^*) \quad\leq\quad \frac{\|\mathbf{x}_t - \mathbf{x}^*\|^2 - \|\mathbf{x}_{t+1} - \mathbf{x}^*\|^2}{\eta_{t+1}} + \eta_{t+1}G^2. \tag{10.5}$$

Summing (10.5) from $t = 1$ to T. Set $\eta_{t+1} = 1/(\alpha t)$ and, using (10.4), we have

$$\begin{aligned} 2\sum_{t=1}^{T} \mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}^*) &\leq \sum_{t=1}^{T} \|\mathbf{x}_t - \mathbf{x}^*\|^2 \left(\frac{1}{\eta_{t+1}} - \frac{1}{\eta_t} - \alpha \right) + G^2 \sum_{t=1}^{T} \eta_{t+1} \\ &= G^2 \sum_{t=1}^{T} \frac{1}{\alpha t} \leq \frac{G^2}{\alpha}(1 + \log T). \quad \square \end{aligned}$$

10.5 Recent Applications

In this section we describe two recent applications of the convex optimization view to regret minimization which have resolved open questions in the field.

10.5.1 Bandit Linear Optimization

The first application is to the bandit linear optimization problem. Online linear optimization is a special case of online convex optimization in which the loss functions are linear (such as analyzed for the RFTL algorithm). In the bandit version, called bandit linear optimization, the only feedback available to the decision maker is the loss (rather than the entire loss function). This general framework naturally captures important problems such as online routing and online ad-placement for search engine results.

This generalization was put forth by Awerbuch and Kleinberg (2008) in the context of the online shortest path problem. Awerbuch and Kleinberg (2008) gave an efficient algorithm for the problem with a suboptimal regret bound, and conjectured the existence of an efficient and optimal regret algorithm.

The problem attracted much attention in the machine learning community (Flaxman et al., 2005; Dani and Hayes, 2006; Dani et al., 2008; Bartlett et al., 2008). This question was finally resolved in (Abernethy et al., 2008) where an efficient and optimal expected regret algorithm was described. Later Abernethy and Rakhlin (2009) gave an efficient algorithm which also attains this optimal regret bound with high probability. The paper introduced the use of self-concordant barrier functions as a regularization in the RFTL framework. Self-concordant barriers are a powerful tool from optimization which has enabled researchers in operations research to develop efficient polynomial-time algorithms for (offline) convex optimization. The scope of this deep technical issue is beyond this survey, but the resolution of this open question is an excellent example of how the convex optimization approach to regret minimization led to the discovery of powerful tools which in turn resolved fundamental questions in machine learning.

10.5.2 Variational Regret Bounds

A cornerstone of modern machine learning are algorithms for prediction from expert advice, the first example of regret minimization we described. It is already well-established that there exist algorithms that, under fully adversarial cost sequences, attain average cost approaching that of the best expert in hindsight. More precisely, there exist efficient algorithms which attain regret of $O(\sqrt{T \log n})$ in the setting of prediction from expert advice with n experts.

However, a priori it is not clear why online learning algorithms should have high regret (growing with the number of iterations) in an unchanging environment. As an extreme example, consider a setting in which there are only two experts. Suppose that the first expert always incurs cost 1, whereas the second expert always incurs cost $\frac{1}{2}$. One would expect to figure out this pattern quickly, and focus on the second expert, thus incurring a total cost that is at most $\frac{T}{2}$ plus at most a constant extra cost (irrespective of the number of rounds T), thus having only constant regret. However, for a long time all analyses of expert learning algorithms gave only a regret bound of $\Theta(\sqrt{T})$ in this simple case (or very simple variations of it).

More generally, the natural bound on the regret of a "good" learning algorithm should depend on *variation* in the sequence of costs, rather than purely on the number of iterations. If the cost sequence has low variation, we expect our algorithm to be able to perform better.

This intuition has a direct analog in the stochastic setting: here, the sequence of experts' costs is independently sampled from a distribution. In this situation, a natural bound on the rate of convergence to the optimal expert is controlled by the variance of the distribution (low variance should imply faster convergence). This conjecture was formalized by Cesa-Bianchi, Mansour and Stoltz (henceforth the "CMS conjecture") in (Cesa-Bianchi et al., 2007), who assert that *"proving such a rate in the fully adversarial setting would be a fundamental result"*.

The CMS conjecture was proved in the more general case of online linear optimization in Hazan and Kale (2008). Again, the convex optimization view was instrumental in the solution, and taking the general linear optimization view, it was found that a simple geometric argument implies the result. Further work on variational bounds included an extension to the bandit linear optimization setting (Hazan and Kale, 2009a) and to exp-concave loss functions including the problem of portfolio selection (Hazan and Kale, 2009b).

10.6 References

J. Abernethy and A. Rakhlin. Beating the adaptive bandit with high probability. In *Proceedings of the 22nd Annual Conference on Learning Theory*, 2009.

J. Abernethy, E. Hazan, and A. Rakhlin. Competing in the dark: An efficient algorithm for bandit linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory*, pages 263–274, 2008.

B. Awerbuch and R. Kleinberg. Online linear optimization and adaptive routing. *J. Comput. Syst. Sci.*, 74(1):97–114, 2008.

P. L. Bartlett, V. Dani, T. P. Hayes, S. Kakade, A. Rakhlin, and A. Tewari. High-probability regret bounds for bandit online linear optimization. In *Proceedings of the 21st Annual Conference on Learning Theory*, pages 335–342, 2008.

N. Cesa-Bianchi and G. Lugosi. *Prediction, Learning, and Games*. Cambridge University Press, 2006.

N. Cesa-Bianchi, Y. Mansour, and G. Stoltz. Improved second-order bounds for prediction with expert advice. *Machine Learning*, 66(2–3):321–352, 2007.

T. Cover. Universal portfolios. *Math. Finance*, 1(1):1–19, 1991.

T. Cover and J. Thomas. *Elements of Information Theory*. John Wiley, 1991.

V. Dani and T. P. Hayes. Robbing the bandit: Less regret in online geometric optimization against an adaptive adversary. In *Proceedings of the 16th ACM-SIAM Symposium on Discrete Algorithms*, pages 937–943, 2006.

V. Dani, T. Hayes, and S. Kakade. The price of bandit information for online optimization. In J. Platt, D. Koller, Y. Singer, and S. Roweis, editors, *Advances in Neural Information Processing Systems 20*. MIT Press, Cambridge, MA, 2008.

A. Flaxman, A. T. Kalai, and H. B. McMahan. Online convex optimization in the bandit setting: Gradient descent without a gradient. In *Proceedings of the 16th Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 385–394, 2005.

A. J. Grove, N. Littlestone, and D. Schuurmans. General convergence results for linear discriminant updates. *Machine Learning*, 43(3):173–210, 2001.

J. Hannan. Approximation to bayes risk in repeated play. *In M. Dresher, A. W. Tucker, and P. Wolfe, editors, Contributions to the Theory of Games, volume 3*, pages 97–139, 1957.

E. Hazan and S. Kale. Extracting certainty from uncertainty: Regret bounded by variation in costs. In *The 21st Annual Conference on Learning Theory (COLT)*, pages 57–68, 2008.

E. Hazan and S. Kale. Better algorithms for benign bandits. In C. Mathieu, editor, *ACM-SIAM Symposium on Discrete Algorithms (SODA).*, pages 38–47. SIAM, 2009a.

E. Hazan and S. Kale. On stochastic and worst-case models for investing. In *Advances in Neural Information Processing Systems 22*. MIT Press, 2009b.

E. Hazan, A. Agarwal, and S. Kale. Logarithmic regret algorithms for online convex optimization. *Machine Learning*, 69(2-3):169–192, 2007.

A. Kalai and S. Vempala. Efficient algorithms for online decision problems. *Journal of Computer and System Sciences*, 71(3):291–307, 2005.

J. Kivinen and M. K. Warmuth. Relative loss bounds for multidimensional regression problems. *Machine Learning*, 45(3):301–329, 2001.

E. Takimoto and M. K. Warmuth. Path kernels and multiplicative updates. *Journal of Machine Learning Research*, 4:773–818, 2003. special issue on learning theory.

M. Zinkevich. Online convex programming and generalized infinitesimal gradient ascent. In *Proceedings of the 20th International Conference on Machine Learning*, pages 928–936, 2003.

Appendix: The FTL-BTL Lemma

The following proof is essentially due to Kalai and Vempala (2005).

Proof of Lemma 10.2. For convenience, denote by $\mathbf{f}_0 = \frac{1}{\eta}\mathcal{R}$, and assume we start the algorithm from $t = 0$ with an arbitrary $\mathbf{x}_0$. The lemma is now proved by induction on T.

Induction base: Note that by definition, we have that $\mathbf{x}_1 = \arg\min_{\mathbf{x}}\{\mathcal{R}(\mathbf{x})\}$, and thus $\mathbf{f}_0(\mathbf{x}_1) \leq \mathbf{f}_0(\mathbf{u})$ for all $\mathbf{u}$, and $\mathbf{f}_0(\mathbf{x}_0) - \mathbf{f}_0(\mathbf{u}) \leq \mathbf{f}_0(\mathbf{x}_0) - \mathbf{f}_0(\mathbf{x}_1)$.

Induction step: Assume that that for T, we have

$$\sum_{t=0}^{T} \mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{u}) \quad \leq \quad \sum_{t=0}^{T} \mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}_{t+1}),$$

and let us prove for $T+1$. Since $\mathbf{x}_{T+2} = \arg\min_{\mathbf{x}}\{\sum_{t=0}^{T+1} \mathbf{f}_t(\mathbf{x})\}$, we have

$$\begin{aligned}
& \sum_{t=0}^{T+1} \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=0}^{T+1} \mathbf{f}_t(\mathbf{u}) \\
\leq \; & \sum_{t=0}^{T+1} \mathbf{f}_t(\mathbf{x}_t) - \sum_{t=0}^{T+1} \mathbf{f}_t(\mathbf{x}_{T+2}) \\
= \; & \sum_{t=0}^{T} (\mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}_{T+2})) + \mathbf{f}_{T+1}(\mathbf{x}_{T+1}) - \mathbf{f}_{T+1}(\mathbf{x}_{T+2}) \\
\leq \; & \sum_{t=0}^{T} (\mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}_{t+1})) + \mathbf{f}_{T+1}(\mathbf{x}_{T+1}) - \mathbf{f}_{T+1}(\mathbf{x}_{T+2}) \\
= \; & \sum_{t=0}^{T+1} \mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}_{t+1}),
\end{aligned}$$

where in the fourth line we used the induction hypothesis for $\mathbf{u} = \mathbf{x}_{T+2}$. We conclude that

$$\begin{aligned}
& \sum_{t=1}^{T} \mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{u}) \\
\leq \; & \sum_{t=1}^{T} \mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}_{t+1}) + [-\mathbf{f}_0(\mathbf{x}_0) + \mathbf{f}_0(\mathbf{u}) + \mathbf{f}_0(\mathbf{x}_0) - \mathbf{f}_0(\mathbf{x}_1)] \\
= \; & \sum_{t=1}^{T} \mathbf{f}_t(\mathbf{x}_t) - \mathbf{f}_t(\mathbf{x}_{t+1}) + \frac{1}{\eta}\left[\mathcal{R}(\mathbf{u}) - \mathcal{R}(\mathbf{x}_1)\right].
\end{aligned}$$

□