Stochastic-Constrained Stochastic Optimization with Markovian Data

Yeongjong Kim   KIMYJ@KAIST.AC.KR
Department of Mathematical Sciences
KAIST
Daejeon 34141, South Korea

Dabeen Lee   DABEENL@KAIST.AC.KR
Department of Industrial and Systems Engineering
KAIST
Daejeon 34141, South Korea

Abstract

This paper considers stochastic-constrained stochastic optimization where the stochastic constraint is to satisfy that the expectation of a random function is below a certain threshold. In particular, we study the setting where data samples are drawn from a Markov chain and thus are not independent and identically distributed. We generalize the drift-plus-penalty framework, a primal-dual stochastic gradient method developed for the i.i.d. case, to the Markov chain sampling setting. We propose two variants of drift-plus-penalty; one is for the case when the mixing time of the underlying Markov chain is known while the other is for the case of unknown mixing time. In fact, our algorithms apply to a more general setting of constrained online convex optimization where the sequence of constraint functions follows a Markov chain. Both algorithms are adaptive in that the first works without knowledge of the time horizon while the second uses AdaGrad-style algorithm parameters, which is of independent interest. We demonstrate the effectiveness of our proposed methods through numerical experiments on classification with fairness constraints.

Keywords: stochastic-constrained stochastic optimization, constrained online convex optimization, Markov chain stochastic gradient descent, drift-plus-penalty, classification with fairness constraints

1 Introduction

In this paper, we consider the stochastic-constrained stochastic optimization (SCSO) problem

$$\min_{\boldsymbol{x}\in\mathcal{X}} \quad \mathbb{E}_{\boldsymbol{\xi}\sim\mu}[f(\boldsymbol{x},\boldsymbol{\xi})] \quad \text{s.t.} \quad \mathbb{E}_{\boldsymbol{\xi}\sim\mu}[g(\boldsymbol{x},\boldsymbol{\xi})] \leq 0 \qquad \text{(SCSO)}$$

where the expectation is taken with respect to the random parameter $\boldsymbol{\xi}$ over a stationary distribution μ, f and g are convex loss and constraint functions, and $\mathcal{X}$ is a compact domain. This formulation of SCSO has applications in stochastic programming with a risk constraint (Rockafellar and Uryasev, 2000), chance-constrained programming (Nemirovski and Shapiro, 2007), portfolio optimization (Dentcheva and Ruszczynski, 2003), sparse matrix completion (Akhtar et al., 2021), semi-supervised learning (Chapelle et al., 2010), classification with fairness constraints (Zafar et al., 2019; Celis et al., 2019; Donini et al., 2018), Neyman-Pearson classification (Scott and Nowak, 2005; Rigollet and Tong, 2011), ranking with fairness constraints (Celis et al., 2018), recommendation systems with

fairness constraints (Yao and Huang, 2017), scheduling in distributed data centers (Yu et al., 2017), safe reinforcement learning (García et al., 2015), and stochastic simple bilevel optimization (Jalilzadeh et al., 2023; Cao et al., 2023).

Stochastic approximation (SA) algorithms are prevalent solution methods for SCSO. Basically, we run gradient-based algorithms with an oracle providing i.i.d. samples of $f(\boldsymbol{x}, \boldsymbol{\xi}), g(\boldsymbol{x}, \boldsymbol{\xi}), \nabla f(\boldsymbol{x}, \boldsymbol{\xi}), \nabla g(\boldsymbol{x}, \boldsymbol{\xi})$ for a given solution $\boldsymbol{x}$. Lan and Zhou (2020) proposed the cooperative stochastic approximation scheme for SCSO, which is a stochastic extension of Polyak's subgradient method developed for constrained optimization (Polyak, 1967). Xiao (2019) developed the penalized stochastic gradient method that takes the square of the constrained function as a penalty term. Lin et al. (2020) developed a level set-based algorithm for SCSO. Akhtar et al. (2021) studied an augmented Lagrangian-based stochastic primal-dual algorithm, which was inspired by the primal-dual framework for online convex optimization with long-term constraints due to Mahdavi et al. (2012). Furthermore, motivated by recent success in adaptive gradient algorithms, Yan and Xu (2022) considered an adaptive primal-dual stochastic gradient method for SCSO. Zhang et al. (2023) proposed a stochastic variant of the proximal method of multipliers. Zhang et al. (2022) studied another stochastic proximal method of multipliers based on linearization.

SA algorithms for other related problem settings are as follows. Yu et al. (2017) studied online convex optimization with stochastic constraints where the constraint functions are stochastic i.i.d. while the objective functions are arbitrary, for which they proposed the drift-plus-penalty (DPP) algorithm. DPP applies to SCSO given that i.i.d. samples of $g(\boldsymbol{x}, \boldsymbol{\xi})$, $\nabla g(\boldsymbol{x}, \boldsymbol{\xi})$, and $\nabla f(\boldsymbol{x}, \boldsymbol{\xi})$ are available. Wei et al. (2020) developed an extension of DPP, and Lee et al. (2023) provided a projection-free algorithm for the setting. Moreover, SCSO can be formulated as a stochastic saddle-point problem by taking the Lagrangian if certain constraint qualifications hold. Nemirovski et al. (2009); Juditsky et al. (2011) developed stochastic mirror-prox algorithms for general stochastic saddle point problems. Zhao (2022) devised an accelerated stochastic framework for convex-concave saddle-point problems, and Yazdandoost Hamedani et al. (2023) devised a randomized adaptive primal-dual method.

The aforementioned solution methods for SCSO require i.i.d. data samples from the stationary distribution μ when running SA algorithms. However, there are several application scenarios in which sampling from the stationary distribution μ independently and identically is difficult. For example, federated learning serves as an alternative to traditional machine learning systems that require data centralization, with the purpose of improving data privacy (Zhao et al., 2018). The basic framework is that data is stored on individual local devices while the training is governed by a central server. Another related setting is distributed optimization over sensor networks (Rabbat and Nowak, 2004; Lopes and Sayed, 2007; Johansson et al., 2007) and multi-agent systems (Johansson et al., 2008). For these applications, an enormous amount of data is distributed over distinct nodes of a network, for which data exchange and massage passing are between immediate neighboring nodes.

One resolution approach for such application scenarios is Markov chain stochastic gradient descent (SGD) (Johansson et al., 2010; Ram et al., 2009a). Basically, Markov chain SGD takes a random walk over a network of local data centers and updates solutions based on data acquired from the data centers visited. Here, as the data is generated by a Markov random walk, there is inherent dependence and bias between data samples. More generally, the framework can be viewed as a variant of the Markov chain Monte Carlo

method (Andrieu et al., 2003). That said, Markov chain SGD can be applied to other applications domains where data is collected from a Markov process, such as decentralized learning (Yang et al., 2021), robust estimation (Poljak and Tsypkin, 1980; Sarkar and Rakhlin, 2019), and reinforcement learning (Nagaraj et al., 2020; Kowshik et al., 2021).

Following the Markov incremental subgradient methods due to Johansson et al. (2010); Ram et al. (2009a) designed for distributed optimization problems, Markov chain SGD with data sampled from a general Markov process have been studied. Duchi et al. (2012) developed Markov chain SGD with data sampled from an ergodic process. Later, Sun et al. (2018) studied Markov chain SGD for convex and nonconvex problems when the underlying Markov chain is nonreversible. Doan et al. (2020) proposed an accelerated version of Markov chain SGD for both convex and nonconvex settings. Dorfman and Levy (2022) considered the setting where the mixing time of the underlying Markov chain is unknown, for which they developed Markov chain SGD based on the multi-level Monte Carlo estimation scheme (Giles, 2015; Blanchet and Glynn, 2015). Roy et al. (2022) studied nonconvex problems where the transition of the underlying Markov chain is state-dependent, motivated by strategic classification and reinforcement learning. Recent works (Doan, 2023; Even, 2023) established some performance guarantees of Markov chain SGD under minimal assumptions.

Applications of SCSO naturally motivate and necessitate algorithmic frameworks that can handle data sets with inherent dependence and bias. Portfolio optimization in finance takes time series data. Machine learning with fairness constraints processes heterogeneous data sets from disjoint sources. Scheduling in distributed data centers will benefit from distributed optimization technologies. Safe and constrained reinforcement learning can be formulated as SCSO for which the training data is obtained from trajectories of the underlying Markov decision process. Despite this immediate need, no prior work exists for solving SCSO with non-i.i.d. data samples. Motivated by this, the objective of this paper is to develop stochastic approximation algorithms that run with data sampled from a Markov chain.

1.1 Our Contributions

This paper initiates the study of stochastic approximation algorithms for stochastic-constrained stochastic optimization (SCSO) using non-i.i.d. data samples. Inspired by recent advances in Markov chain SGD, we develop primal-dual stochastic gradient methods using data sampled from a Markov chain, which can be viewed as primal-dual variants of Markov chain SGD. Specifically, we extend the drift-plus-penalty algorithm by Yu et al. (2017) that was originally developed for the i.i.d. setting. We adopt the approach of ergodic mirror descent by Duchi et al. (2012) for the case of known mixing time and the framework of Dorfman and Levy (2022) for the setting where the mixing time is unknown.

Our key technical contribution is to develop two variants of the drift-plus-penalty algorithm that can take a sequence of dependent constraint functions. These two algorithms solve constrained online convex optimization where the objective functions can be arbitrary and the constraint functions are generated from a Markov chain. One of them is for the case of known mixing time, while the other is for the case when the mixing time is unknown. We provide regret and constraint violation bounds for the algorithms, which delineate how their performance depends on the mixing time. Based on the regret and constraint violation

	Algorithm 1	Algorithm 1 under Slater's condition	Algorithm 2
Oblivious to τ_{mix}	✗	✗	✔
Regret	$\tilde{O}\left(\tau_{\mathrm{mix}}^{1-\beta}T^{1-\beta}\right)$	$\tilde{O}\left(\sqrt{\tau_{\mathrm{mix}}T}\right)$	adaptive
Constraint violation	$O\left(\tau_{\mathrm{mix}}^{\beta/2}T^{(\beta+1)/2}\right)$	$\tilde{O}\left(\sqrt{\tau_{\mathrm{mix}}T}\right)$	adaptive
Optimality gap	$\tilde{O}\left(\frac{\tau_{\mathrm{mix}}^{1-\beta}}{T^{\beta}} + \frac{\tau_{\mathrm{mix}}^{\beta/2}}{T^{(1-\beta)/2}}\right)$	$\tilde{O}\left(\frac{\sqrt{\tau_{\mathrm{mix}}}}{\sqrt{T}}\right)$	$\tilde{O}\left(\frac{\tau_{\mathrm{mix}}^{1-\beta}}{T^{\beta}}\right)$
Feasibility gap	$\tilde{O}\left(\frac{\tau_{\mathrm{mix}}^{\beta/2}}{T^{(1-\beta)/2}}\right)$	$\tilde{O}\left(\frac{\sqrt{\tau_{\mathrm{mix}}}}{\sqrt{T}}\right)$	$\tilde{O}\left(\frac{\tau_{\mathrm{mix}}^{(2\beta+1)/4}}{T^{(1-\beta)/2}}\right)$

Table 1: Bounds on regret, constraint violation, optimality gap, and feasibility gap under Algorithms 1 and 2 ($\beta \in (0, 1/2]$ is a predetermined algorithm parameter that controls the balance between regret and constraint violation)

guarantees, we analyze the optimality gap and feasibility gap for SCSO. The connection between the constrained online convex optimization formulation and SCSO for our Markovian setting is not as immediate as the i.i.d. setting because the expectation in (SCSO) is taken with respect to the stationary distribution of the underlying Markov chain.

What follows is a more detailed description of our contributions. Our results are also summarized in Table 1.

- In Section 3, we consider the case when the mixing time of the underlying Markov chain is known, for which we develop Algorithm 1, a variant of the drift-plus-penalty algorithm. We first prove that for online convex optimization with Markovian constraints, the regret of Algorithm 1 is $\tilde{O}(\tau_{\mathrm{mix}}^{1-\beta}T^{1-\beta})$ and the constraint violation is $O(\tau_{\mathrm{mix}}^{\beta/2}T^{(\beta+1)/2})$ where τ_{mix} is the mixing time, T is the length of horizon, β can be chosen to be any number in $(0, 1/2]$, and $\tilde{O}$ hides a poly $\log T$ factor. If we further assume that (SCSO) satisfies Slater's constraint qualification, then the regret and constraint violation of Algorithm 1 can be both bounded by $\tilde{O}(\sqrt{\tau_{\mathrm{mix}}T})$. We remark that Algorithm 1 is adaptive in that its parameters are chosen without knowledge of T, unlike the vanilla drift-plus-penalty algorithm. These results generalize the work of Yu et al. (2017).

- Based on the regret and constraint violation analysis for Algorithm 1, we show that the averaging of the sequence of solutions generated by Algorithm 1 guarantees that the optimality gap is bounded by $\tilde{O}(\tau_{\mathrm{mix}}^{1-\beta}T^{-\beta} + \tau_{\mathrm{mix}}^{\beta/2}T^{-(1-\beta)/2})$ while the feasibility gap is bounded by $\tilde{O}(\tau_{\mathrm{mix}}^{\beta/2}T^{-(1-\beta)/2})$. If we further assume that (SCSO) satisfies Slater's constraint qualification, then the optimality gap and feasibility gap of Algorithm 1 can be both bounded by $\tilde{O}(\sqrt{\tau_{\mathrm{mix}}/T})$.

- In Section 4, we propose Algorithm 2, another variant of the drift-plus-penalty algorithm, for the setting where the mixing time is unknown. The parameters of Algorithm 2 are set in an adaptive fashion, as in the AdaGrad method (Duchi et al.,

2011; Levy, 2017). Then we apply Algorithm 2 to constrained online convex optimization where the objective and constraint functions are given by the Multi-level Monte Carlo estimation scheme as in Dorfman and Levy (2022). We provide adaptive regret and constraint violation bounds for Algorithm 2. We note that Algorithm 2 is the first AdaGrad-style adaptive variant of the drift-plus-penalty algorithm. In fact, Algorithm 2 applies to online convex optimization with adversarial constraints (Neely and Yu, 2017) and provides adaptive performance guarantees. We include this result in Appendix C.

- Combining the estimation accuracy bounds on the Multi-level Montel Carlo method by Dorfman and Levy (2022) and our adaptive regret and constraint violation bounds, we prove that the optimality gap under Algorithm 2 is bounded by $\tilde{O}(\tau_{\mathrm{mix}}^{1-\beta}T^{-\beta})$ while the feasibility gap is bounded by $\tilde{O}(\tau_{\mathrm{mix}}^{(2\beta+1)/4}T^{-(1-\beta)/2})$.

- In Section 5, we provide numerical results from experiments on a classification problem with fairness constraints. Specifically, we take the logistic regression formulation proposed in Zafar et al. (2019). The numerical results on random problem instances demonstrate the efficacy of our proposed algorithmic frameworks for solving SCSO with Markovian data.

The main component of our analysis is to provide bounds on the terms

$$\mathbb{E}\left[\sum_{t=1}^{T} Q_t g_t(\boldsymbol{x})\right], \quad \mathbb{E}\left[\sum_{t=1}^{T} \frac{Q_t}{V_t} g_t(\boldsymbol{x})\right], \quad \mathbb{E}\left[Q_t\right], \quad \mathbb{E}\left[\frac{Q_t}{V_t}\right]$$

under our Markovian data regime. All these terms involve the virtual queue size Q_t, and therefore, controlling the virtual queue size is crucial to guarantee a fast convergence rate. We include our proofs of the main theorems in Sections 6 and 7. We include some of the known lemmas and tools for analyzing the drift-plus-penalty method due to Yu et al. (2017) in Appendices A and D. In fact, Appendix A state these results for any adaptive version of the drift-plus-penalty algorithm, which uses time-varying parameters V_t and α_t.

1.2 Related Work

Although we have listed and explained important lines of previous work that motivate this paper, we supplement the list by mentioning a few more relevant results in stochastic approximation algorithms and Markov chain stochastic gradient methods.

Stochastic Approximation for Stochastic Optimization Starting from the seminar paper by Robbins and Monro (1951), stochastic approximation algorithms, also known as stochastic gradient methods, have been a central topic of research in the domain of machine learning, operations research, and optimization (Ermoliev, 1983; Pflug, 1996; Ruszczyński and Syski, 1986; Bottou et al., 2018). There already exist numerous works on stochastic approximation algorithms for stochastic optimization. In particular, for SCSO or stochastic optimization with expectation constraints, various stochastic approximation methods were proposed and studied by Lan and Zhou (2020); Xiao (2019); Lin et al. (2020); Akhtar et al. (2021); Yan and Xu (2022); Zhang et al. (2023, 2022), as discussed earlier.

We note that online convex optimization with stochastic constraints is a superclass of SCSO under the i.i.d. data regime. Mahdavi et al. (2012); Jenatton et al. (2016) developed augmented Lagrangian-based primal-dual algorithms for online convex optimization with deterministic long-term constraints. It turns out that these algorithms and their analysis can be adapted to the setting of stochastic constraints (Akhtar et al., 2021). Later, Yu et al. (2017); Wei et al. (2020) proposed the drift-plus-penalty algorithm that achieves better regret and constraint violation guarantees under Slater's condition. Yuan and Lamperski (2018); Yi et al. (2021); Guo et al. (2022) also provided algorithms for online convex optimization with stochastic constraints.

Markov Chain Stochastic Gradient Descent   The asymptotic convergence of Markov chain SGD was studied in Bertsekas and Tsitsiklis (1996); Borkar (2008); Benveniste et al. (1990) based on ordinary differential equation methods. Lopes and Sayed (2007); Johansson et al. (2007, 2010); Ram et al. (2009a) developed incremental subgradient methods for the distributed optimization setting where there is a network of data centers and an algorithm performs a random walk over the network to obtain data samples. These methods are also referred to as token algorithms. More recently, Mao et al. (2020) devised what they called the walkman algorithm, which is a token algorithm based on an augmented Lagrangian method. Sun et al. (2022); Ayache et al. (2023) considered adaptive variants of token algorithms. Hendrikx (2023) developed a more general framework for token algorithms that allows multiple tokens for improving communication efficiency and may adopt existing optimization tools such as variance reduction and acceleration.

In addition to random walk-based token algorithms, there exist general Markov chain sampling frameworks for stochastic gradient methods. As mentioned earlier, Duchi et al. (2012); Sun et al. (2018); Doan et al. (2020); Dorfman and Levy (2022); Roy et al. (2022) developed Markov chain SGD methods for convex and nonconvex stochastic optimization problems. Moreover, Sun et al. (2020) developed a Markov chain sampling-based block coordinate descent method. Sun et al. (2023) proposed a decentralized variant of Markov chain SGD. Wang et al. (2022) considered the stability of Markov chain SGD and deduced its generalization bounds. Doan (2023) derived convergence guarantees on Markov chain SGD without a smoothness assusmption, and Even (2023) studied convergence of Markov chain SGD without the bounded gradient assumption.

2 Preliminaries

In this section, we provide problem formulations for stochastic-constrained stochastic and online convex optimization under the Markovian data sampling regime. In addition, Section 2.1 gives the formal definition of the mixing time of a Markov chain. Section 2.4 describes a list of assumptions considered throughout the paper.

2.1 Markov Chain and Mixing Time

Given two probability distributions $\mathbb{P}, \mathbb{Q}$ over the probability space $(\mathcal{S}, \mathcal{F})$, the total variation distance between them is defined as

$$\|\mathbb{P} - \mathbb{Q}\|_{TV} := \sup_{A \in \mathcal{F}} |\mathbb{P}(A) - \mathbb{Q}(A)|.$$

Let $\{\boldsymbol{\xi_t}\}_{t=1}^{\infty}$ be a time-homogeneous ergodic Markov chain with a finite state space $\mathcal{S}$. For a distribution ν over $(\mathcal{S}, \mathcal{F})$, we denote by $\mathbb{P}^t(\nu, \cdot)$ the conditional probability distribution of $\boldsymbol{\xi_{t+1}}$ given $\boldsymbol{\xi_1} \sim \nu$. Since $\{\boldsymbol{\xi_t}\}_{t=1}^{\infty}$ is ergodic, it has a unique stationary distribution μ, i.e., $\mathbb{P}^t(\mu, \cdot) = \mu(\cdot)$. In fact, the long-term distribution of the ergodic Markov chain converges to μ regardless of the initial distribution, which can be demonstrated as follows. It is known that $\mathbb{P}^t(\nu, \cdot)$ for any t and ν satisfies

$$\|\mathbb{P}^t(\nu, \cdot) - \mu\|_{TV} \leq C\alpha^t$$

for some $\alpha \in (0, 1)$ and $C > 0$ (Levin and Peres, 2017). Then we define quantities d_{mix} and $\tau_{\mathrm{mix}}(\epsilon)$ as follows.

$$d_{\mathrm{mix}} := \sup_{\nu}\|P^t(\nu, \cdot) - \mu\|_{TV}, \quad \tau_{\mathrm{mix}}(\epsilon) := \inf\{t \in \mathbb{N} : d_{\mathrm{mix}}(t) \leq \epsilon\}.$$

Moreover, following the convention, we define τ_{mix} as

$$\tau_{\mathrm{mix}} := \tau_{\mathrm{mix}}(1/4),$$

and we refer to τ_{mix} as the *mixing time* of the underlying Markov chain. It is known that $d_{\mathrm{mix}}(l\tau_{\mathrm{mix}}) \leq 2^{-l}$ for every $l \in \mathbb{N}$ (Levin and Peres, 2017, Chapter 4), which implies that

$$\tau_{\mathrm{mix}}(\epsilon) \leq \lceil \log_2 \epsilon^{-1} \rceil \tau_{\mathrm{mix}}.$$

In particular, throughout the paper, we will use quantity τ defined as

$$\tau := \tau_{\mathrm{mix}}(1/T) = O(\tau_{\mathrm{mix}} \log T)$$

where T is the length of the time horizon.

2.2 Stochastic-Constrained Online Convex Optimization with Markovian Data

Let $\mathcal{X} \subset \mathbb{R}^d$ be a compact convex domain. Let $\{f_t : \mathcal{X} \to \mathbb{R}\}_{t=1}^{\infty}$ be a sequence of arbitrary convex functions. Another sequence of convex functions $\{g_t : \mathcal{X} \to \mathbb{R}\}_{t=1}^{\infty}$ is assumed to follow an ergodic Markov chain. More precisely, there exists a time-homogeneous ergodic Markov chain $\{\boldsymbol{\xi_t}\}_{t=1}^{\infty}$ such that $g_t(\boldsymbol{x}) = g(\boldsymbol{x}, \boldsymbol{\xi_t})$. Let

$$\bar{g}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\xi} \sim \mu}[g(\boldsymbol{x}, \boldsymbol{\xi})]$$

for $\boldsymbol{x} \in \mathcal{X}$ where the expectation is taken with respect to the stationary distribution μ of the Markov chain. Then the problem is to solve and compute

$$\boldsymbol{x}^* \quad \in \quad \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \quad \sum_{t=1}^{T} f_t(\boldsymbol{x}) \quad \text{s.t.} \quad \bar{g}(\boldsymbol{x}) \leq 0.$$

However, the information about the functions $\{f_t\}_{t=1}^{T}$ and $\{g_t\}_{t=1}^{T}$ is revealed *online*. Basically, at each step t, we choose our decision $\boldsymbol{x}_t \in \mathcal{X}$ before observing f_t, g_t, after which we receive feedback about them. The stochastic setting studied in Yu et al. (2017) is that constraint functions $g_1, \ldots, g_T$ are i.i.d., which means that $\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_T}$ are i.i.d., while we consider the

case where constraint functions $g_1, \ldots, g_T$ follow an ergodic Markov chain and thus are dependent. That said, we may refer to the problem setting as *online convex optimization with ergodic constraints.*

To measure the performance of a learning algorithm for the online optimization problem, we adopt the regret and cumulative constraint violation definition due to Yu et al. (2017). The regret and cumulative constraint violation of an algorithm that generates solutions $\boldsymbol{x_1}, \ldots, \boldsymbol{x_T}$ over T time steps are given by

$$\text{Regret}(T) = \sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x}^*), \quad \text{Violation}(T) = \sum_{t=1}^{T} g_t(\boldsymbol{x}_t).$$

We want these quantities to have a sublinear growth in T. Note that the benchmark solution $\boldsymbol{x}^*$ satisfies the expectation constraint $\mathbb{E}_{\boldsymbol{\xi} \sim \mu}[g(\boldsymbol{x}^*, \boldsymbol{\xi})] \leq 0$.

2.3 Stochastic-Constrained Stochastic Optimization with Markovian Data

Next, we consider the setting where the objective functions $\{f_t\}_{t=1}^{\infty}$ as well as the constraint functions $\{g_t\}_{t=1}^{\infty}$ are given by an ergodic Markov chain. Without loss of generality, we may assume that $f_t(\boldsymbol{x}) = f(\boldsymbol{x}, \boldsymbol{\xi_t})$ and $g_t(\boldsymbol{x}) = g(\boldsymbol{x}, \boldsymbol{\xi_t})$ for some time-homogeneous ergodic Markov chain $\{\boldsymbol{\xi_t}\}_{t=1}^{\infty}$ with a stationary distribution μ. As before, let

$$\bar{f}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\xi} \sim \mu}[f(\boldsymbol{x}, \boldsymbol{\xi})]$$

and $\bar{g}(\boldsymbol{x}) = \mathbb{E}_{\boldsymbol{\xi} \sim \mu}[g(\boldsymbol{x}, \boldsymbol{\xi})]$ for $\boldsymbol{x} \in \mathcal{X}$. Then the problem is to solve and compute

$$\boldsymbol{x}^{\#} \quad \in \quad \underset{\boldsymbol{x} \in \mathcal{X}}{\operatorname{argmin}} \quad \bar{f}(\boldsymbol{x}) \quad \text{s.t.} \quad \bar{g}(\boldsymbol{x}) \leq 0.$$

Here, we do not have direct access to $(\bar{f}, \bar{g})$, but we receive samples (f_t, g_t) which converge to $(\bar{f}, \bar{g})$ in expectation.

For the performance measure of a learning algorithm for this problem, we consider the following standard notions of optimality gap and constraint violation. For a solution $\boldsymbol{x} \in \mathcal{X}$

$$\text{Gap}(\boldsymbol{x}) = \bar{f}(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}^{\#}), \quad \text{Infeasibility}(\boldsymbol{x}) = \bar{g}(\boldsymbol{x}).$$

We want these quantities to approach 0 as T grows. Hereinafter, we refer to $\text{Gap}(\boldsymbol{x})$ and $\text{Infeasibility}(\boldsymbol{x})$ as the optimality gap and the feasibility gap, respectively.

In the special case where $\{\boldsymbol{\xi_t}\}_{t=1}^{\infty}$ are i.i.d. samples drawn from μ, solving stochastic-constrained online convex optimization would provide a solution to stochastic-constrained stochastic optimization. One can argue that if $\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_T}$ are i.i.d., then

$$\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}}) \leq \frac{1}{T} \mathbb{E}\left[\text{Regret}(T)\right], \quad \text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}}) \leq \frac{1}{T} \mathbb{E}\left[\text{Violation}(T)\right]$$

where $\bar{\boldsymbol{x}}_{\boldsymbol{T}}$ denotes the simple average of $\boldsymbol{x_1}, \ldots, \boldsymbol{x_T}$:

$$\bar{\boldsymbol{x}}_{\boldsymbol{T}} = \frac{1}{T} \sum_{t=1}^{T} \boldsymbol{x}_t.$$

However, in contrast to the i.i.d. case, the above inequalities do not hold for the case of an arbitrary ergodic Markov chain. This is because the distribution of $\boldsymbol{\xi_{t+1}}$ conditioned on $\boldsymbol{\xi_t}$ is not equal to the stationary distribution μ.

On the other hand, we will use the fact that the long-term distribution of an ergodic Markov chain converges to its stationary distribution. Based on this observation, we first bound the expected regret and the expected constraint violation of the online problem, and then we use this to bound the optimality gap and the feasibility gap of the stochastic optimization problem.

2.4 Notations and Assumptions

We work over a norm $\|\cdot\|$ and its dual norm $\|\cdot\|_*$. We choose a convex mirror map $\Phi : \mathcal{C} \to \mathbb{R}$, where $\mathcal{C} \subseteq \mathbb{R}^d$ is a convex domain containing $\mathcal{X}$. We use the corresponding *Bregman divergence* defined as

$$D(\boldsymbol{x}, \boldsymbol{y}) = \Phi(\boldsymbol{x}) - \Phi(\boldsymbol{y}) - \nabla\Phi(\boldsymbol{y})^\top (\boldsymbol{x} - \boldsymbol{y}).$$

Assumption 1 *There is a constant $R > 0$ such that $D(\boldsymbol{x}, \boldsymbol{y}) \leq R^2$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$, and Φ is 2-strongly convex with respect to norm $\|\cdot\|$ i.e. $\|\boldsymbol{x} - \boldsymbol{y}\|^2 \leq D(\boldsymbol{x}, \boldsymbol{y})$ for any $\boldsymbol{x}, \boldsymbol{y} \in \mathcal{X}$. Moreover, $f(\cdot, \boldsymbol{\xi})$ and $g(\cdot, \boldsymbol{\xi})$ are differentiable for each $\boldsymbol{\xi} \in \mathcal{S}$.*

We use notations F_t, G_t, H_t for $t \in [T]$ given by

$$F_t := \|\nabla f_t(\boldsymbol{x}_t)\|_*, \ G_t := \|\nabla g_t(\boldsymbol{x}_t)\|_*, \ H_t := |g_t(\boldsymbol{x}_t)|$$

which are parameters used in our adaptive algorithm. Due to the stochasticity of f_t, g_t, these are random variables. We assume these quantities are bounded.

Assumption 2 *There exist constants $F, G, H > 0$ such that*

$$\|\nabla_{\boldsymbol{x}} f(\boldsymbol{x}, \boldsymbol{\xi})\|_* \leq F, \ \|\nabla_{\boldsymbol{x}} g(\boldsymbol{x}, \boldsymbol{\xi})\|_* \leq G, \ |g(\boldsymbol{x}, \boldsymbol{\xi})| \leq H$$

for any $\boldsymbol{x} \in \mathcal{X}$ and $\boldsymbol{\xi} \in \mathcal{S}$.

Assumptions 1 and 2 are common in online convex optimization, stochastic optimization, and the Markov chain SGD literature. The following assumption is referred to as *Slater's condition* or *Slater's constraint qualification* for constrained optimization.

Assumption 3 (Slater's condition) *There exists $\hat{\boldsymbol{x}} \in \mathcal{X}$ such that $\bar{g}(\hat{\boldsymbol{x}}) \leq -\epsilon$ for some $\epsilon > 0$.*

For online convex optimization with stochastic i.i.d. constraint functions, Slater's condition leads to improvement in regret and constraint violation (Yu et al., 2017). In Section 3, we will show that even for online convex optimization with ergodic constraint functions, assuming that Slater's condition holds results in an improvement in the regret and constraint violation of Algorithm 1.

Algorithm 1 Ergodic **D**rift-**P**lus-**P**enalty (EDPP)

Initialize: Initial iterates $\boldsymbol{x_1} \in \mathcal{X}$, $Q_1 = 0$, and $0 < \beta \leq 1/2$.

for $t = 1$ **to** T **do**

    Observe f_t and g_t.

    Set penalty parameter V_t and step size parameter α_t as

$$V_t = (\tau_{\mathrm{mix}}t)^{\beta}, \quad \alpha_t = \tau_{\mathrm{mix}}t.$$

  Primal update: Set $\boldsymbol{x}_{t+1}$ as

$$\boldsymbol{x}_{t+1} = \underset{\boldsymbol{x} \in \mathcal{X}}{\operatorname{argmin}} \left\{ (V_t \nabla f_t(\boldsymbol{x}_t) + Q_t \nabla g_t(\boldsymbol{x}_t))^{\top} \boldsymbol{x} + \alpha_t D(\boldsymbol{x}, \boldsymbol{x}_t) \right\}$$

  Dual update: Set Q_{t+1} as

$$Q_{t+1} = \left[Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^{\top}(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right]_{+}$$

end for

---

3 Known Mixing Time

We first focus on the setting where we have access to the mixing time τ_{mix} of the underlying Markov chain $\{\boldsymbol{\xi_t}\}_{t=1}^{\infty}$. Duchi et al. (2012) studied this case for stochastic convex minimization without a stochastic functional constraint, for which they modified the step size of the stochastic gradient descent method based on the mixing time parameter τ_{mix}. Inspired by this approach, we take and modify the drift-plus-penalty (DPP) algorithm due to Neely and Yu (2017); Yu et al. (2017) developed for stochastic-constrained online convex optimization. Based on DPP, we develop our algorithm by setting the algorithm parameters properly to adapt to the mixing time τ_{mix}.

3.1 Ergodic Drift-Plus-Penalty

The DPP algorithm has two parameters, V and α, where V is the penalty parameter and α determines the step size. Yu et al. (2017) set $V = \sqrt{T}$ and $\alpha = T$. In contrast to the vanilla DPP algorithm, our algorithm uses parameters

$$V_t = (\tau_{\mathrm{mix}}t)^{\beta}, \quad \alpha_t = \tau_{\mathrm{mix}}t$$

for iterations $t = 1, \ldots, T$, where T is the length of the horizon and β is another algorithm parameter that controls the balance between the regret and the constraint violation. Our algorithm, which we call *ergodic drift-plus-penalty (EDPP)*, is described in Algorithm 1.

    Note that the mixing time τ_{mix} is now part of parameters V_t and α_t, as in the ergodic mirror descent algorithm by Duchi et al. (2012). Second, V_t and α_t are time-varying, so our algorithm is *adaptive* (Jenatton et al., 2016) and oblivious to the length of the time horizon T. One may wonder why we do not use T instead of t, i.e., $V = (\tau_{\mathrm{mix}}T)^{\beta}$ and $\alpha = \tau_{\mathrm{mix}}T$. In fact, our numerical results, which will be presented in Section 5, demonstrate that Algorithm 1 with the adaptive parameters V_t and α_t outperforms the algorithm with

fixed parameters $V = (\tau_{\mathrm{mix}}T)^{\beta}$ and $\alpha = \tau_{\mathrm{mix}}T$. However, the performance analysis of the vanilla DPP algorithm by Yu et al. (2017) does not immediately extend to such adaptive parameters. Hence, we prove that the DPP framework with parameters of varying t still achieves the desired regret and constraint violation guarantees.

Let us also briefly explain how the DPP framework initially developed by Yu et al. (2017) as well as our Algorithm 1 works. We may regard Q_t as the size of a *virtual queue* at time t. Then we consider the associated *quadratic Lyapunov term* $L_t = Q_t^2/2$ and study the corresponding *drift* given by $\Delta_t = L_{t+1} - L_t = (Q_{t+1}^2 - Q_t^2)/2$. It is not difficult to see that

$$\Delta_t \leq Q_t \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^{\top}(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right) + \frac{1}{2}(H_t + G_t R)^2$$

holds (Lemma 19). Here, the upper bound on the drift Δ_t has term $Q_t \nabla g_t(\boldsymbol{x}_t)^{\top}(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)$ that depends on the next iterate $\boldsymbol{x}_{t+1}$. Hence, by choosing $\boldsymbol{x}_{t+1}$ that minimizes $Q_t \nabla g_t(\boldsymbol{x}_t)^{\top}(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)$, we may attempt to control the drift. In fact, the primal update sets $\boldsymbol{x}_{t+1}$ to be the minimizer of

$$\underbrace{Q_t \nabla g_t(\boldsymbol{x}_t)^{\top}(\boldsymbol{x} - \boldsymbol{x}_t)}_{\text{drift}} + \underbrace{V_t \nabla f_t(\boldsymbol{x}_t)^{\top}(\boldsymbol{x} - \boldsymbol{x}_t) + \alpha_t D(\boldsymbol{x}, \boldsymbol{x}_t)}_{\text{penalty}}$$

over $\mathcal{X}$. Consequently, at each iteration, we get to choose a solution that minimizes the drift term Δ_t and a penalty term for controlling the objective simultaneously.

3.2 Performance Guarantees of Ergodic Drift-Plus-Penalty

First, we analyze the regret and constraint violation of EDPP for the constrained online convex optimization setting under the Markovian sampling regime, formulated in Section 2.2. Recall that $\tau_{\mathrm{mix}} = \tau_{\mathrm{mix}}(1/4)$ and that parameter τ is defined as $\tau = \tau_{\mathrm{mix}}(T^{-1})$, which satisfies $\tau \leq \lceil \log_2 T \rceil \tau_{\mathrm{mix}}$.

Theorem 1 *Suppose that Assumptions 1 and 2 hold. Then for online convex optimization with ergodic constraints, Algorithm 1 achieves*

$$\mathbb{E}\left[\mathrm{Regret}(T)\right] = O\left(\tau_{mix}^{-\beta}\tau T^{1-\beta}\right),$$

$$\mathbb{E}\left[\mathrm{Violation}(T)\right] = O\left(\tau_{mix}^{\beta/2}T^{(\beta+1)/2} + \sqrt{(\tau-1)T}\right)$$

where the expectation is taken with respect to the randomness in running the algorithm.

One of the key components of the analysis for proving Theorem 1 is that we bound the expected size of the virtual queue Q_t at time t as follows.

$$\mathbb{E}[Q_t] = O\left(\tau_{\mathrm{mix}}^{\beta/2}T^{(\beta+1)/2} + \sqrt{(\tau-1)T}\right).$$

Another key part is to analyze the term

$$\sum_{t=1}^{T} \mathbb{E}\left[Q_t g_t(\boldsymbol{x})\right]$$

11

where function g_t is not independent of the virtual queue size Q_t under our Markovian sampling regime. In contrast, if $g_1, \ldots, g_T$ were i.i.d., then $\mathbb{E}\left[Q_t g_t(\boldsymbol{x})\right] = \mathbb{E}\left[Q_t \bar{g}(\boldsymbol{x})\right]$ would hold. Instead, we relate the term with

$$\sum_{t=1}^{T-\tau+1} \mathbb{E}\left[Q_t g_{t+\tau-1}(\boldsymbol{x})\right]$$

and use the intuition that the distribution of the ergodic Markov chain after τ steps is close to its stationary distribution. We provide a bound on the term in Lemma 11. Similarly, we also need to analyze the term

$$\sum_{t=1}^{T} \mathbb{E}\left[\frac{Q_t}{V_t} g_t(\boldsymbol{x})\right],$$

and an upper bound on the sum is given in Lemma 12. Based on these observations, we provide an upper bound on the expected virtual queue size $\mathbb{E}\left[Q_t\right]$, which is given in Lemma 13. Then we apply the performance analysis results of the general adaptive drift-plus-penalty method provided in Appendix A. The complete proof of Theorem 1 is given in Section 6.1.

Next we analyze the performance of EDPP on the stochastic-constrained stochastic optimization problem under Markovian data sampling, whose formulation is given by (SCSO).

Theorem 2 *Suppose that Assumptions 1 and 2 hold. Then for stochastic-constrained stochastic optimization* (SCSO), *Algorithm 1 guarantees that*

$$\mathbb{E}\left[\mathrm{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = O\left(\frac{\tau}{\tau_{mix}^{\beta} T^{\beta}} + \frac{\tau - 1}{\tau_{mix}^{1-\beta/2} T^{(1-\beta)/2}} + \frac{(\tau - 1)^{3/2}}{\tau_{mix} T^{1/2}} + \frac{\tau}{T}\right),$$

$$\mathbb{E}\left[\mathrm{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = O\left(\frac{\tau}{\tau_{mix}^{1-\beta/2} T^{(1-\beta)/2}} + \frac{(\tau - 1)^{3/2}}{\tau_{mix} T^{1/2}} + \frac{\tau}{T}\right)$$

where $\bar{\boldsymbol{x}}_{\boldsymbol{T}} = \sum_{t=1}^{T} \boldsymbol{x}_t / T$ and the expectation is taken with respect to the randomness in running the algorithm.

Note that Theorem 1 implies

$$\frac{1}{T}\mathbb{E}\left[\mathrm{Regret}(T)\right] = O\left(\frac{\tau}{\tau_{\mathrm{mix}}^{\beta} T^{\beta}}\right),$$

but the bound on the optimality gap given in Theorem 2 has additional terms due to the difference between the stationary distribution of the ergodic Markov chain and the distribution of f_{t+1} conditioned on f_t. In fact, under Markovian sampling, we have

$$\sum_{t=1}^{T} \mathbb{E}\left[f_t(\boldsymbol{x}_t)\right] \neq \sum_{t=1}^{T} \mathbb{E}\left[\bar{f}(\boldsymbol{x}_t)\right].$$

Here, to get around this issue, we also use the intuition that $\mathbb{E}\left[\bar{f}(\boldsymbol{x}_t)\right]$ is close to $\mathbb{E}\left[f_{t+\tau-1}(\boldsymbol{x}_t)\right]$ as the distribution of the Markov chain after τ steps is close to its stationary distribution.

The proof of Theorem 2 is given in Section 6.2. Moreover, since $\tau = \tilde{O}(\tau_{\text{mix}})$, it follows that

$$\mathbb{E}\left[\text{Regret}(T)\right] = \tilde{O}\left(\tau_{\text{mix}}^{1-\beta} T^{1-\beta}\right), \quad \mathbb{E}\left[\text{Violation}(T)\right] = \tilde{O}\left(\tau_{\text{mix}}^{\beta/2} T^{(\beta+1)/2} + \sqrt{\tau_{\text{mix}} T}\right),$$

$$\mathbb{E}\left[\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = \tilde{O}\left(\frac{\tau_{\text{mix}}^{1-\beta}}{T^{\beta}} + \frac{\tau_{\text{mix}}^{\beta/2}}{T^{(1-\beta)/2}} + \frac{\tau_{\text{mix}}}{T}\right),$$

$$\mathbb{E}\left[\text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = \tilde{O}\left(\frac{\tau_{\text{mix}}^{\beta/2}}{T^{(1-\beta)/2}} + \frac{\sqrt{\tau_{\text{mix}}}}{\sqrt{T}} + \frac{\tau_{\text{mix}}}{T}\right)$$

where the $\tilde{O}$ hides a $\log T$ factor. In particular, if we set $\beta = 1/3$, then we have $\mathbb{E}\left[\text{Regret}(T)\right] = \tilde{O}(\tau_{\text{mix}}^{2/3} T^{2/3})$, $\mathbb{E}\left[\text{Violation}(T)\right] = O(\tau_{\text{mix}}^{1/6} T^{2/3} + \tau_{\text{mix}}^{1/2} T^{1/2})$, $\mathbb{E}\left[\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = \tilde{O}(\tau_{\text{mix}}^{2/3} T^{-1/3} + \tau_{\text{mix}} T^{-1})$, and $\mathbb{E}\left[\text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = O(\tau_{\text{mix}}^{1/6} T^{-1/3} + \tau_{\text{mix}}^{1/2} T^{-1/2} + \tau_{\text{mix}} T^{-1})$. Furthermore, observe that if $\{\boldsymbol{\xi}_t\}_{t=1}^{\infty}$ is a sequence of i.i.d. random variables, then we have $\tau_{\text{mix}} = \tau = 1$. In this case, by Theorems 1 and 2, Algorithm 1 guarantees that $\mathbb{E}\left[\text{Regret}(T)\right] = O(T^{1-\beta})$, $\mathbb{E}\left[\text{Violation}(T)\right] = O(T^{(\beta+1)/2})$, $\mathbb{E}\left[\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = O(T^{-\beta})$, and $\mathbb{E}\left[\text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] = O(T^{-(1-\beta)/2})$ for any $\beta \in (0, 1/2]$, which recovers the result of Jenatton et al. (2016).

If we further assume that Slater's constraint qualification holds, we can argue that we get a better control on the size of $\mathbb{E}[Q_t]$. Note that the upper bound on the expected queue size given by Lemma 13 is $\mathbb{E}[Q_t] = O(\tau_{\text{mix}}^{\beta/2} t^{(\beta+1)/2})$ which holds regardless of whether Slater's condition holds or not. On the other hand, we will argue that under Slater's condition, we have

$$\mathbb{E}[Q_t] = O\left(\frac{\tau(\tau + t)}{\sqrt{\tau_{\text{mix}} t}}\right) = \tilde{O}(\sqrt{\tau_{\text{mix}} t}).$$

This is consistent with Yu et al. (2017) as they proved that $\mathbb{E}[Q_t] = O(\sqrt{T})$ for the i.i.d. setting. In fact, our proof for bounding $\mathbb{E}[Q_t]$ is more involved than the argument of Yu et al. (2017) because we use adaptive step sizes for Algorithm 1 and consider non-i.i.d. constraint functions. This leads to improvements as stated in the following result.

Theorem 3 *Suppose that Assumptions 1–3 hold. Then for online convex optimization with ergodic constraints and stochastic-constrained stochastic optimization, Algorithm 1 with $\beta = 1/2$ guarantees*

$$\mathbb{E}\left[\text{Regret}(T)\right] = O\left(\frac{\tau\sqrt{T}}{\sqrt{\tau_{mix}}}\right), \quad \mathbb{E}\left[\text{Violation}(T)\right] = O\left(\frac{\tau(\tau + T)}{\sqrt{\tau_{mix} T}}\right),$$

$$\mathbb{E}[\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = O\left(\frac{\tau^2}{\tau_{mix}^{3/2} \sqrt{T}} + \frac{\tau^{5/2}}{\tau_{mix}^{3/2} T}\right),$$

$$\mathbb{E}[\text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = O\left(\frac{\tau^2}{\tau_{mix}^{3/2} \sqrt{T}} + \frac{\tau^{5/2}}{\tau_{mix}^{3/2} T} + \frac{\tau^2}{\tau_{mix}^{1/2} T^{3/2}}\right),$$

where the expectation is taken with respect to the randomness in running the algorithm.

Our analysis takes into account the time-varying algorithm parameters V_t and α_t as well as the fact that the functions are correlated according to a Markov chain. To consider this, we

prove Lemmas 14 and 15 that lead to time-varying bounds on the expected virtual queue size. The complete proof of Theorem 3 is included in Section 6.3. Since $\tau = \tilde{O}(\tau_{\text{mix}})$,

$$\mathbb{E}\left[\text{Regret}(T)\right] = \tilde{O}\left(\sqrt{\tau_{\text{mix}}T}\right), \quad \mathbb{E}\left[\text{Violation}(T)\right] = \tilde{O}\left(\sqrt{\tau_{\text{mix}}T} + \frac{\tau_{\text{mix}}^{3/2}}{\sqrt{T}}\right),$$

$$\mathbb{E}[\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = \tilde{O}\left(\frac{\sqrt{\tau_{\text{mix}}}}{\sqrt{T}} + \frac{\tau_{\text{mix}}}{T}\right), \quad \mathbb{E}[\text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = \tilde{O}\left(\frac{\sqrt{\tau_{\text{mix}}}}{\sqrt{T}} + \frac{\tau_{\text{mix}}}{T} + \frac{\tau_{\text{mix}}^{3/2}}{T^{3/2}}\right).$$

4 Unknown Mixing Time

Next, we study the setting where the mixing time τ_{mix} is not observable. Even if we do not know the mixing time of the underlying Markov chain, we would still want to provide a learning algorithm that provides performance guarantees of similar orders. To achieve this goal, we develop yet another variant of the drift-plus-penalty algorithm which incorporates the multi-level Monte Carlo (MLMC) gradient estimation scheme (Giles, 2015; Blanchet and Glynn, 2015; Dorfman and Levy, 2022). Dorfman and Levy (2022) first introduced the approach of combining stochastic gradient descent with the MLMC gradient estimation framework for stochastic optimization with no stochastic functional constraint.

For the case of known mixing time, we may update the step size based on the mixing time τ_{mix} to achieve an optimal dependence on the parameter. When τ_{mix} is not known, Dorfman and Levy (2022) used AdaGrad-based adaptive step sizes (Duchi et al., 2011; Levy, 2017; Ward et al., 2019). We take this idea to develop an AdaGrad variant of the drift-plus-penalty algorithm for SCSO, described in Algorithm 2, that incorporates the MLMC gradient estimation framework. The AdaGrad version of DPP itself is of independent interest.

4.1 Multi-Level Monte Carlo Sampling

The idea behind the multi-level Monte Carlo estimation scheme is to obtain many consecutive samples from an ergodic Markov chain and take their average. At the same time, we may control the expected number of consecutive samples required for each time step by $O(\log T)$.

More precisely, for each time step t, we N_t sample $\boldsymbol{\xi}_{\boldsymbol{t}}^{\boldsymbol{(1)}}, \ldots, \boldsymbol{\xi}_{\boldsymbol{t}}^{\boldsymbol{(N_t)}}$ where N_t itself is a random variable given by

$$N_t = \begin{cases} \tilde{N}_t, & \text{if } \tilde{N}_t \leq T^2 \\ 1, & \text{otherwise} \end{cases}$$

and $\tilde{N}_t = 2^{J_t}$ with $J_t \sim \text{Geom}(1/2)$. Note that in our case, the condition is that $\tilde{N}_t \leq T^2$ where the bound on $\tilde{N}_t$ is T^2, while it was set to T in Dorfman and Levy (2022). With this sampling strategy, we define $\mathcal{F}_t$ as the σ-field

$$\mathcal{F}_t = \sigma\left(\{N_1, \ldots, N_t\} \cup \bigcup_{s=1}^{t} \left\{\boldsymbol{\xi}_{\boldsymbol{s}}^{\boldsymbol{(1)}}, \ldots, \boldsymbol{\xi}_{\boldsymbol{s}}^{\boldsymbol{(N_s)}}\right\}\right).$$

14

Let $\mathbb{E}_t[\cdot]$ denote the conditional expectation with respect to $\mathcal{F}_t$, i.e., $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$. Next, for $t \geq 1$ and $N \geq 1$, we define

$$f_t^N(\boldsymbol{x}) := \frac{1}{N} \sum_{i=1}^{N} f(\boldsymbol{x}, \boldsymbol{\xi_t^{(i)}}), \quad g_t^N(\boldsymbol{x}) := \frac{1}{N} \sum_{i=1}^{N} g(\boldsymbol{x}, \boldsymbol{\xi_t^{(i)}}).$$

Based on this, we define the MLMC estimators of f and g as follows.

$$(f_t, g_t) = (f_t^1, g_t^1) + \begin{cases} N_t \left((f_t^{N_t}, g_t^{N_t}) - (f_t^{N_t/2}, g_t^{N_t/2}) \right), & \text{if } N_t > 1 \\ 0, & \text{otherwise.} \end{cases}$$

Basically, functions f_t and g_t are obtained after applying the MLMC estimation scheme to the underlying ergodic Markov chain. One thing to note, however, is that f_t and g_t are not necessarily convex anymore (Dorfman and Levy, 2022). To remedy this issue, what we can argue instead is that $\mathbb{E}_{t-1}[f_t]$ and $\mathbb{E}_{t-1}[g_t]$ are convex. Based on (Dorfman and Levy, 2022, Lemma 3.1), we deduce the following lemma.

Lemma 4 *Let $j_{\max} := \max \left\{ j \in \mathbb{N} : 2^j \leq T^2 \right\} = \lfloor 2 \log_2 T \rfloor$. Then for each t,*

$$\mathbb{E}_{t-1}[f_t] = \mathbb{E}_{t-1}\left[f_t^{2^{j_{\max}}} \right], \quad \mathbb{E}_{t-1}[\nabla f_t] = \mathbb{E}_{t-1}\left[\nabla f_t^{2^{j_{\max}}} \right],$$

$$\mathbb{E}_{t-1}[g_t] = \mathbb{E}_{t-1}\left[g_t^{2^{j_{\max}}} \right], \quad \mathbb{E}_{t-1}[\nabla g_t] = \mathbb{E}_{t-1}\left[\nabla g_t^{2^{j_{\max}}} \right].$$

Moreover, we have

$$\mathbb{E}\left[\|\nabla f_t(\boldsymbol{x}_t)\|_*^2 \right] = \tilde{O}(F^2 \tau_{mix}), \quad \mathbb{E}\left[\|\nabla g_t(\boldsymbol{x}_t)\|_*^2 \right] = \tilde{O}(G^2 \tau_{mix}), \quad \mathbb{E}\left[|g_t(\boldsymbol{x}_t)|^2 \right] = \tilde{O}(H^2 \tau_{mix}).$$

Lastly, the expected number of samples for time step t satisfies $\mathbb{E}[N_t] \leq 2 \log_2 T + 1$.

The reason for setting the upper bound T^2 on $\tilde{N}_t$ instead of T is to achieve high accuracy of estimation for f_t, g_t, ∇f_t, and ∇g_t that leads to the desired performance guarantees of Algorithm 2. To be specific, we use the following estimation bounds based on (Dorfman and Levy, 2022, Lemma A.6).

Lemma 5 *There exists $C(T) > 0$ with*

$$C(T) = O \left(\left(\log(T) \log \left(\tau_{mix} T^2 \log(T) \right) \right)^{1/2} \right)$$

such that

$$\mathbb{E}_{t-1}\left[\left| f_t^{2^{j_{\max}}}(\boldsymbol{x}) - \bar{f}(\boldsymbol{x}) \right|^2 \right] \leq C(T)^2 \frac{\tau_{mix}}{T^2}, \quad \mathbb{E}_{t-1}\left[\|\nabla f_t^{2^{j_{\max}}}(\boldsymbol{x}) - \nabla \bar{f}(\boldsymbol{x})\|_*^2 \right] \leq C(T)^2 \frac{\tau_{mix}}{T^2},$$

$$\mathbb{E}_{t-1}\left[\left| g_t^{2^{j_{\max}}}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x}) \right|^2 \right] \leq C(T)^2 \frac{\tau_{mix}}{T^2}, \quad \mathbb{E}_{t-1}\left[\|\nabla g_t^{2^{j_{\max}}}(\boldsymbol{x}) - \nabla \bar{g}(\boldsymbol{x})\|_*^2 \right] \leq C(T)^2 \frac{\tau_{mix}}{T^2}.$$

hold for any $\boldsymbol{x} \in \mathcal{X}$ that is measurable with respect to $\mathcal{F}_{t-1}$ and any $t \in [T]$.

The complete proof of this lemma is given in Appendix E.

Algorithm 2 MLMC Adaptive **D**rift-**P**lus-**P**enalty (MDPP)

    Initialize: Initial iterates $\boldsymbol{x_1} \in \mathcal{X}$, $Q_1 = 0$ and parameters $0 < \beta \leq 1/2$, $\delta > 0$.
    for $t = 1$ **to** T **do**
        Observe f_t and g_t via MLMC method.
        Set penalty parameter V_t, step size parameter α_t as (1).
        Primal update: Set $\boldsymbol{x}_{t+1}$ as

$$\boldsymbol{x}_{t+1} = \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \left\{ (V_t \nabla f_t(\boldsymbol{x}_t) + Q_t \nabla g_t(\boldsymbol{x}_t))^\top \boldsymbol{x} + \alpha_t D(\boldsymbol{x}, \boldsymbol{x}_t) \right\}$$

        Dual update: Set Q_{t+1} as

$$Q_{t+1} = \left[Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right]_+$$

    end for

4.2 Adaptive Drift-Plus-Penalty

The second component of our algorithm for the unknown mixing time setting is the AdaGrad variant of the drift-plus-penalty algorithm. To develop AdaGrad-style step sizes, let us define the following sequence of parameters. For some positive constant $\delta > 0$, we set

$$a_0 = S_0 = \delta, \quad a_t := \frac{F_t^2}{4} + R^2 G_t^2 + H_t^2 + \delta, \quad S_t := \delta + \sum_{s=1}^{t} a_s$$

for $t \geq 1$. Here, we may choose any positive number for δ. Recall that $F_t = \|\nabla f_t(\boldsymbol{x}_t)\|_*$, $G_t = \|\nabla g_t(\boldsymbol{x}_t)\|_*$, and $H_t = |g_t(\boldsymbol{x}_t)|$. Based on these parameters, we set the algorithm parameters V_t and α_t as follows.

$$V_t = \frac{S_{t-1}^\beta}{R}, \quad \alpha_t = \frac{S_{t-1}}{R^2} \tag{1}$$

for some $0 < \beta \leq 1/2$. Here, the penalty parameter V_t and the step size parameter α_t can be chosen without knowledge of the global upper bounds F, G, H on F_t, G_t, H_t.

    Now we are ready to describe our algorithm, which we call the *MLMC adaptive drift-plus-penalty (MDPP)* algorithm. Here, MDPP is a combination of the AdaGrad-style adaptive drift-plus-penalty algorithm with the MLMC estimator presented in the previous subsection. More importantly, the algorithm is designed to solve constrained online convex optimization where the MLMC estimators $f_1, \ldots, f_T$ are the objective loss functions and the MLMC estimators $g_1, \ldots, g_T$ are the constraint functions. The following lemma provides an adaptive regret guarantee and an adaptive constraint violation bound for Algorithm 2 for the associated online convex optimization.

Lemma 6 *Suppose that Assumptions 1 and 2 hold. Then for the constrained online convex optimization problem where the MLMC estimators $f_1, \ldots, f_T$ are the objective loss functions and the MLMC estimators $g_1, \ldots, g_T$ are the constraint functions, Algorithm 2 achieves the*

following. For any $\boldsymbol{x} \in \mathcal{X}$ with $\bar{g}(\boldsymbol{x}) \leq 0$, we have

$$\mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x})\right]$$

$$= \tilde{O}\left(\mathbb{E}\left[S_T\right]^{1-\beta} + \tau_{mix}^{1/2}\mathbb{E}\left[S_T\right]^{1/2-\beta} + \tau_{mix}^{1/2}T^{1/4}\mathbb{E}\left[S_T\right]^{1/4-\beta/2} + \tau_{mix}^{1/2}T^{1-\beta}\right),$$

$$\mathbb{E}\left[\sum_{t=1}^{T} g_t(\boldsymbol{x}_t)\right]$$

$$= \tilde{O}\left(\mathbb{E}[S_T]^{1/2} + T^{1/4}\mathbb{E}[S_T]^{\beta/2+1/4} + \tau_{mix}^{1/2} + \frac{\mathbb{E}[S_T] + T^{1/2}\mathbb{E}[S_T]^{\beta+1/2}}{\tau_{mix}^{\beta/2+1/4}T^{\beta/2+1/2}}\right.$$

$$\left. + \frac{\tau_{mix}^{1/2} + \mathbb{E}[S_T]^{1/2} + \mathbb{E}[S_T]^{\beta/2+1/4}}{\tau_{mix}^{\beta/2+1/4}T^{\beta/2+1/2}} + (\log S_T)^2\tau_{mix}^{\beta/2-1/4}T^{\beta/2+1/2}\right).$$

Recall that the parameters V_t and α_t depend on the parameter δ. Here, we may decide any positive number for δ, its choice does affect the performance of Algorithm 2. Although the bounds given in Lemma 6 do not exhibit an explicit dependence on δ, our proof of Lemma 6 in Section 7.2 reveals that increasing δ increases the objective gap $\mathbb{E}[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x})]$ and decreases the constraint violation $\mathbb{E}[\sum_{t=1}^{T} g_t(\boldsymbol{x}_t)]$. Likewise, decreasing δ decreases the objective gap and increases the constraint violation.

Lemma 4 implies that $\mathbb{E}\left[S_T\right] = \tilde{O}(\tau_{\mathrm{mix}}T)$. Plugging in this bound on $\mathbb{E}\left[S_T\right]$ to the adaptive performance guarantees given in Lemma 6, we deduce the following result.

Proposition 7 *Suppose that Assumptions 1 and 2 hold. Then for the constrained online convex optimization problem where the MLMC estimators $f_1, \ldots, f_T$ are the objective loss functions and the MLMC estimators $g_1, \ldots, g_T$ are the constraint functions, Algorithm 2 achieves the following. For any $\boldsymbol{x} \in \mathcal{X}$ with $\bar{g}(\boldsymbol{x}) \leq 0$, we have*

$$\mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x})\right] = \tilde{O}\left(\tau_{mix}^{1-\beta}T^{1-\beta}\right),$$

$$\mathbb{E}\left[\sum_{t=1}^{T} g_t(\boldsymbol{x}_t)\right] = \tilde{O}\left(\tau_{mix}^{\beta/2+1/4}T^{\beta/2+1/2} + \tau_{mix}^{3/4-\beta/2}T^{1/2-\beta/2}\right).$$

for any $\boldsymbol{x} \in \mathcal{X}$ satisfying $\bar{g}(\boldsymbol{x}) \leq 0$.

We remark that the performance bounds given in Lemma 6 and Proposition 7 are not comparable to the regret and constraint violation bounds for stochastic-constrained stochastic optimization. When each pair of loss and constraint functions for stochastic-constrained stochastic optimization corresponds to a single data, the associated regret and constraint

violation measure are given by the following.

$$\text{Regret}(T) = \sum_{t=1}^{T} \sum_{j=1}^{N_t} f_t^{(j)}(\boldsymbol{x}_t) - \sum_{t=1}^{T} \sum_{j=1}^{N_t} f_t^{(j)}(\boldsymbol{x}^*),$$

$$\text{Violation}(T) = \sum_{t=1}^{T} \sum_{j=1}^{N_t} g_t^{(j)}(\boldsymbol{x}_t).$$

Here, $f_t^{(1)}, \ldots, f_t^{(N_t)}$ are the N_t sampled functions from which we derive the MLMC estimator f_t for $t \in [T]$. In contrast, Lemma 6 and Proposition 7 analyze the performance of Algorithm 2 on the sequence of the MLMC estimators. Despite this, it would be an interesting question to understand the performance of Algorithm 2 for the latter online convex optimization setting where each function pair corresponds to a single data.

The proof of Lemma 6 is given in Section 7. One of the main components of the analysis is to provide adaptive bounds on the terms $\mathbb{E}[Q_t]$ and $\mathbb{E}[Q_t/V_t]$. This is possible thanks to our subtle choice of parameters V_t and α_t. It turns out that the AdaGrad style analysis for drift-plus-penalty is more sophisticated than that for unconstrained stochastic gradient descent.

Finally, we state the following theorem providing upper bounds on the optimality gap and the feasibility gap under Algorithm 2 for SCSO. The argument is to use the results of Proposition 7 and the estimation error bounds due to Lemma 5. In contrast to the setting of Section 3 for which we had to rely on the mixing property of ergodic Markov chains directly, the MLMC estimators are already close to $\bar{f}$ and $\bar{g}$.

Theorem 8 *Suppose that Assumptions 1 and 2 hold. Then for stochastic-constrained stochastic optimization* (SCSO), *Algorithm 2 guarantees that*

$$\mathbb{E}[\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = \tilde{O}\left(\frac{\tau_{mix}^{1-\beta}}{T^\beta}\right),$$

$$\mathbb{E}[\text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = \tilde{O}\left(\frac{\tau_{mix}^{(2\beta+1)/4}}{T^{(1-\beta)/2}} + \frac{\tau_{mix}^{(3-2\beta)/4}}{T^{(\beta+1)/2}}\right).$$

5 Numerical Experiments

We examine the performance of the ergodic drift-plus-penalty algorithm (Algorithm 1) for the known mixing time case and the MLMC adaptive drift-plus-penalty algorithm (Algorithm 2) for the unknown mixing time case on a linear classification problem with fairness constraints using synthetic data. We follow the experimental setup of Zafar et al. (2019). We adopt Zafar et al. (2019) for creating data points and sensitive features (with $\phi = \pi/2$) and imposing fairness constraints. To make our Markov chain more meaningful and complex enough, we experiment with a 3-state chain instead of a 2-state chain in contrast to Dorfman and Levy (2022). The 3-state Markov chain is given with the following transition matrix

$$\begin{pmatrix} 1-2p & p & p \\ p & 1-2p & p \\ p & p & 1-2p \end{pmatrix}$$

18

which has stationary distribution $(1/3, 1/3, 1/3)$.

Theorem 9 *(Levin and Peres, 2017, Theorems 12.4 and 12.5) For an ergodic and reversible Markov chain with n states, whose transition matrix is P, let $1 = \lambda_1 > \lambda_2 \geq \ldots \geq \lambda_n$ be the eigenvalues of P and $\mu_{\min}$ be the minimum entry of the stationary distribution. Then, the mixing time of the chain satisfies*

$$\frac{|\lambda_2|}{1 - \max\{|\lambda_2|, |\lambda_n|\}} \log 2 \leq \tau_{mix} \leq \frac{1}{1 - \max\{|\lambda_2|, |\lambda_n|\}} \log\left(\frac{4}{\mu_{\min}}\right).$$

Then it follows that the mixing time of the 3-state Markov chain satisfies

$$\frac{1 - 3p}{3p} \log 2 \leq \tau_{\mathrm{mix}} \leq \frac{1}{3p} \log 12.$$

In our experiment, we used $1/3p$ as an approximation of the mixing time.

For each set of data points, we generate two clusters, each of which has 1,000 data points in $\mathbb{R}^2$ sampled from a multivariate normal distribution. All data points from a cluster have the same label in $\{-1, 1\}$. We denote the index set corresponding to each state $j \in \{1, 2, 3\}$ as D_j and the whole index set as $D = D_1 \cup D_2 \cup D_3$. The data clusters are drawn in Figure 1. The color corresponding to each state and label is summarized in Table 2. We then generate
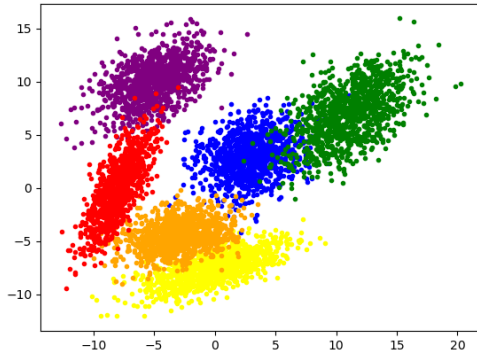
Figure 1: Data Points

label state	1	-1
1	purple	yellow
2	blue	orange
3	green	red

Table 2: Colors for (State, Label) Pairs

binary sensitive feature $z_i \in \{0, 1\}$ randomly for each data point $x_i \in \mathbb{R}^d$, i.e., gender. We want the binary-sensitive feature of the data points to have low covariance with the results

from our classifier. More details about how to create sensitive features are included in the supplement.

We use logistic regression classifiers with the following loss functions

$$f_j(w, b) = \frac{1}{|D_j|} \sum_{i \in D_j} \log(1 + e^{-y_i(w^\top x_i + b)})$$

and constraint functions

$$g_j(w, b) = \frac{1}{|D_j|} \sum_{i \in D_j} (z_i - \bar{z})(w^\top x_i + b) - c$$

$$h_j(w, b) = -\frac{1}{|D_j|} \sum_{i \in D_j} (z_i - \bar{z})(w^\top x_i + b) - c$$

for each state j, where $\bar{z} = \sum_{i \in D} z_i / |\mathcal{D}|$ and $c > 0$. Then the stochastic-constrained stochastic optimization problem is to minimize the usual logistic regression loss function under fairness constraints, which was proposed by Zafar et al. (2019), as follows.

$$\min_{(w,b) \in \mathcal{X}} \quad \frac{1}{|D|} \sum_{i \in D} \log(1 + e^{-y_i(w^\top x_i + b)})$$

$$\text{s.t.} \quad -c \leq \frac{1}{|D|} \sum_{i \in D} (z_i - \bar{z})(w^\top x_i + b) \leq c.$$

Solving this problem with our framework can be viewed as a distributed optimization scheme of Ram et al. (2009b) and Johansson et al. (2010). Basically, there are agents 1,2, and 3 sharing $\bar{z}$, and agent i has data D_i. After we update our parameters w_t and b_t, they are sent to agent i_t, and the agent sends us back the information $\nabla f_{i_t}(w_t, b_t), g_t(w_t, b_t), \nabla g_{i_t}(w_t, b_t)$, $h_t(w_t, b_t), \nabla h_{i_t}(w_t, b_t)$. Here, we may impose that the sequence of selected agents gives rise to an ergodic Markov chain.

The experimental setup involves two constraints. Although we consider the single constraint setting in the paper for simplicity, our results can be easily extended to the case with multiple constraint functions $\bar{g}_1, \ldots, \bar{g}_n$. For each $\bar{g}_i$, we obtain sampled functions $g_{t,i}$ for $t \in [T]$. Then for each t, we update $\boldsymbol{x}_t$ and $\{Q_{t,i}\}_{i=1}^n$ as

$$\boldsymbol{x}_{t+1} = \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \left\{ \left(V_t \nabla f_t(\boldsymbol{x}_t) + \sum_{i=1}^n Q_{t,i} g_{t,i}(\boldsymbol{x}_t) \right)^\top \boldsymbol{x} + \alpha_t D(x, \boldsymbol{x}_t) \right\},$$

$$Q_{t+1,i} = \left[Q_{t,i} + g_{t,i}(\boldsymbol{x}_t) + \nabla g_{t,i}(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right]_+, \quad i = 1, \ldots, n.$$

For Algorithm 1, we use the parameters $V_t = (\tau_{\text{mix}} t)^\beta$ and $\alpha_t = \tau_{\text{mix}} t$ as before. For Algorithm 2, we define the MLMC estimator $g_{t,i}$ using $\{g_{t,i}^{(j)}\}_{j=1}^{N_t}$ for each $i \in [n]$ and define

$$a_0 = S_0 = \delta, \quad a_t = \frac{F_t^2}{4} + \sum_{i=1}^n R^2 G_{t,i}^2 + \sum_{i=1}^n H_{t,i}^2, \quad S_t = \delta + \sum_{s=1}^t a_s,$$

where $G_{t,i} = \|\nabla g_{t,i}(\boldsymbol{x}_t)\|_*, H_{t,i} = |g_{t,i}(\boldsymbol{x}_t)|$. The parameters V_t and α_t are defined as in (1).

We compare our DPP-based algorithms with some existing algorithms developed for stochastic-constrained stochastic optimization with i.i.d. data. The list of algorithms that we tested is given as follows.

- PD : Primal-dual method by Mahdavi et al. (2012).

- PD2 : Primal-dual method by Jenatton et al. (2016).

- DPP : Drift-plus-penalty algorithm by Yu et al. (2017).

- EDPP-t : Ergodic drift-plus-penalty (Algorithm 1).

- EDPP-T : modification of Algorithm 1 with non-adaptive parameters $V_t = \sqrt{\tau_{\mathrm{mix}}T}$ and $\alpha_t = \tau_{\mathrm{mix}}T$.

- MDPP : MLMC adaptive drift-plus-penalty (Algorithm 2).

For MDPP, we observed that the MLMC estimator usually has a high variance in practice, making experimental results unstable. Hence, we truncated MLMC sampling so that the number of samples per iteration is at most 2^4. We chose $\delta = F^2/4 + 2R^2G^2 + 2H^2$, for which we computed the constants $F, G, H > 0$ such that $F_t \leq F, G_{t,i} \leq G, H_{t,i} \leq H$.

We set parameters to $p = 0.001$ and $c = 0.5$ with which we ran the list of algorithms with the same initial parameters and sequence of states. We first ran MDPP with 25,000 iterations which created 101,034 samples. The results on the optimality gap are summarized in Figure 2, and the results on the infeasibility are presented in Figure 3 and Table 3. The results on the regret and the cumulative constraint violations are shown in Figures 4 and 5, respectively.
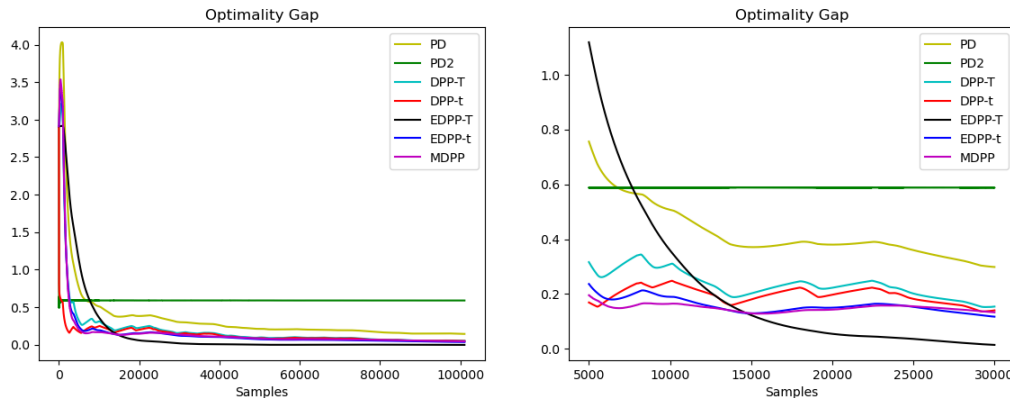

Figure 2: Optimality Gap (Left), Enlarged Figure around 5,000 - 30,000 Samples (Right)

As shown in the figures, Our algorithms (EDPP-T, EDPP-t, MDPP) outperform the other algorithms in terms of the optimality gap. DPP also shows a good optimality gap but it ends with a positive constraint 1 infeasibility. In contrast, EDPP-T, EDPP-t, and MDPP
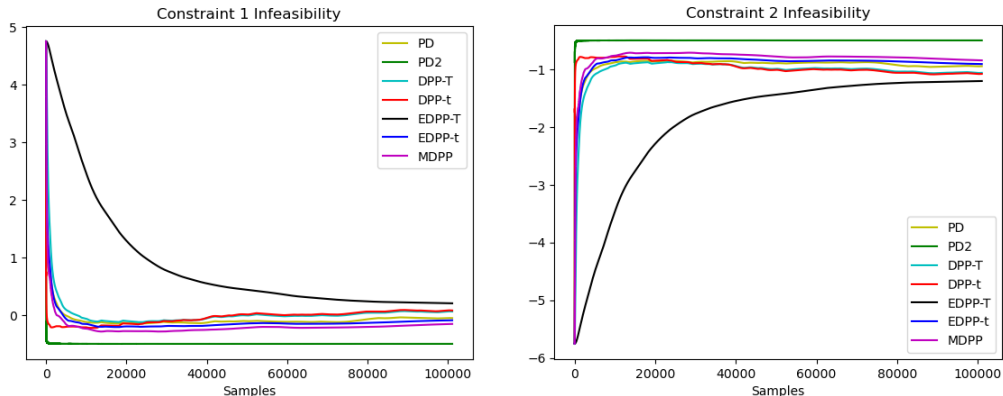
Figure 3: Constraint 1 Infeasibility (Left), Constraint 2 Infeasibility (Right)

Algorithm	Final values of constraint 1 infeasibility
PD	-0.0533
PD2	-0.4997
DPP-T	0.0622
DPP-t	0.0800
EDPP-T	0.2042
EDPP-t	-0.0904
MDPP	-0.1536

Table 3: Final Values of Constraint 1 Infeasibility

all end with a negative constraint infeasibility. Note that after 20,000 samples, EDPP-T achieves the smallest optimality gap, followed by EDPP-t and MDPP. However, Figure 3 shows that EDPP-T incurs a significantly higher infeasibility for constraint 1, given by

$$\frac{1}{|D|} \sum_{i \in D} (z_i - \bar{z})(w^\top x_i + b) \le c$$

than the other algorithms. In contrast, EDPP-t outperforms DPP-T and DPP-t in terms of both the optimality gap and the infeasibility measure. It is also interesting to see that DPP-T and DPP-t behave similarly, while DPP-t performs better than DPP-T.

Figure 4 shows the regret values under various algorithms for the online convex optimization setting where each pair (f_t, g_t) of functions corresponds to one data sample. That said, we excluded MDPP as it requires multiple samples in one round. In addition, we excluded PD2, which exhibits much higher regret values than the other algorithms, to focus on and better present the performance of the other algorithms. Figure 5 shows that EDPP-T incurs a positive cumulative constraint 1 violation, while the other algorithms result in a negative
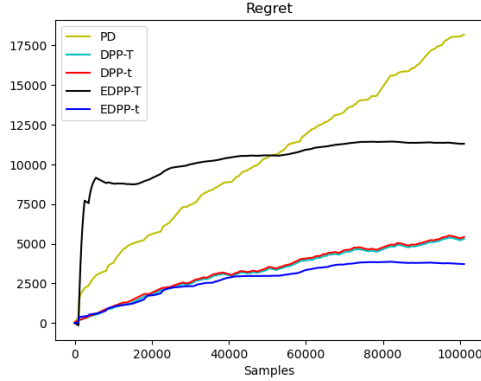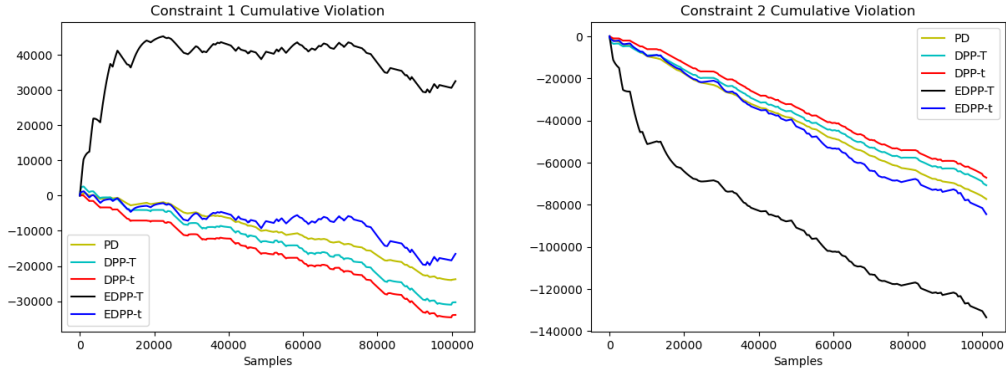
Figure 4: Regret

Figure 5: Constraint 1 Cumulative Violation (Left), Constraint 2 Cumulative Violation (Right)

cumulative constraint violation. We may check from Figures 4 and 5 that EDPP-t performs the best for online convex optimization with ergodic constraints.

6 Analysis of Ergodic Drift-Plus-Penalty for the Known Mixing Time Case

This section presents the proofs of Theorems 1 to 3 given in Section 3. Section 6.1 contains the proof of Theorem 1 which provides regret and constraint violation bounds on the formulation of online convex optimization with ergodic constraints. Then Section 6.2 presents the proof of Theorem 2 that gives bounds on the optimality gap and the feasibility gap for SCSO. In Section 6.3, we prove Theorem 3 for the case where Slater's condition is satisfied. As before, throughout this section, we denote by $\mathcal{F}_t$ the σ-field generated by the information

23

accumulated up to time step t. That is,

$$\mathcal{F}_t = \sigma\left(\{\boldsymbol{\xi_1}, \ldots, \boldsymbol{\xi_t}\}\right).$$

Moreover, $\mathbb{E}_t[\cdot]$ in this section refers to the conditional expectation with respect to $\mathcal{F}_t$, i.e., $\mathbb{E}_t[\cdot] = \mathbb{E}[\cdot \mid \mathcal{F}_t]$ (In Sections 4 and 7, $\mathcal{F}_t$ denotes the σ-field generated up to time step t under the MLMC estimation scheme). Furthermore, $\mathbb{P}_{[t]}^s$ for $t > s$ denotes the probability measure of $\boldsymbol{\xi_s}$ conditional on $\mathcal{F}_t$.

6.1 Ergodic Drift-Plus-Penalty for Online Convex Optimization with Ergodic Constraints

The three important components of our analysis are the one (Lemma 11) bounding the term $\mathbb{E}\left[\sum_{t=1}^T Q_t g_t(\boldsymbol{x})\right]$, the part (Lemma 12) providing an upper bound on the term $\mathbb{E}\left[\sum_{t=1}^T (Q_t/V_t)g_t(\boldsymbol{x})\right]$, and Lemma 13 that derives an upper bound on the expected queue size $\mathbb{E}[Q_t]$. Then, plugging in the deduced bounds to the lemmas in Appendix A analyzing the general template of adaptive drift-plus-penalty, we prove Theorem 1.

By the update rule and the convexity of g_t, we deduce the following straightforward bound on $|Q_{t+1} - Q_t|$.

Lemma 10 *For $t \geq 1$, we have $-H - GR \leq Q_{t+1} - Q_t \leq H$.*

Proof Note that we have

$$Q_{t+1} = \left[Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^T (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)\right]_+ \leq [Q_t + g_t(\boldsymbol{x}_{t+1})]_+ \leq Q_t + H.$$

For the lower bound,

$$\begin{aligned}
Q_{t+1} &\geq Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^T (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \\
&\geq Q_t - |g_t(\boldsymbol{x}_t)| - \|\nabla g_t(\boldsymbol{x}_t)\|_* \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\| \\
&\geq Q_t - H - GR,
\end{aligned}$$

as required. ■

As $Q_1 = 0$, it follows that $Q_t \leq (t-1)H$. Recall that we denote $\tau = \tau_{\text{mix}}(T^{-1})$. The next lemma provides a bound on the term $\sum_{t=1}^T \mathbb{E}[Q_t g_t(\boldsymbol{x})]$. As mentioned in Section 3.1, we need to take into account that $g_1, \ldots, g_T$ are not i.i.d. while they are generated by an ergodic Markov chain.

Lemma 11 *For any $\boldsymbol{x} \in \mathcal{X}$ such that $\bar{g}(\boldsymbol{x}) \leq 0$,*

$$\mathbb{E}\left[\sum_{t=1}^T Q_t g_t(\boldsymbol{x})\right] \leq H(H + GR)(\tau - 1)T + \frac{2H}{T}\sum_{t=1}^T \mathbb{E}[Q_t].$$

Proof If $T < \tau$, then

$$\sum_{t=1}^T Q_t g_t(\boldsymbol{x}) \leq \sum_{t=1}^T (t-1)H^2 = \frac{T(T-1)H^2}{2} < H^2(\tau - 1)T,$$

and the statement follows. If $T \geq \tau$, then

$$\mathbb{E}\left[\sum_{t=1}^{T} Q_t g_t(\boldsymbol{x})\right]$$

$$= \sum_{t=1}^{\tau-1} \mathbb{E}[Q_t g_t(\boldsymbol{x})] + \sum_{t=1}^{T-\tau+1} \mathbb{E}[Q_{t+\tau-1} g_{t+\tau-1}(\boldsymbol{x})]$$

$$\leq \sum_{t=1}^{\tau-1}(t-1)H^2 + \sum_{t=1}^{T-\tau+1} \mathbb{E}[(Q_{t+\tau-1} - Q_t) g_{t+\tau-1}(\boldsymbol{x})] + \sum_{t=1}^{T-\tau+1} \mathbb{E}[Q_t g_{t+\tau-1}(\boldsymbol{x})]$$

$$\leq \frac{(\tau-1)(\tau-2)H^2}{2} + \sum_{t=1}^{T-\tau+1}(\tau-1)H(H+GR) + \sum_{t=1}^{T-\tau+1} \mathbb{E}\left[Q_t \mathbb{E}_{t-1}[g_{t+\tau-1}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})]\right]$$

$$= \frac{(\tau-1)(2T-\tau)H(H+GR)}{2} + \sum_{t=1}^{T-\tau+1} \mathbb{E}\left[Q_t \mathbb{E}_{t-1}[g_{t+\tau-1}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})]\right]$$

where the second inequality holds because $(Q_{t+\tau-1} - Q_t)g_{t+\tau-1}(\boldsymbol{x}) \leq (H+GR)H$ due to Lemma 10, $\bar{g}(\boldsymbol{x}) \leq 0$, and Q_t is $\mathcal{F}_{t-1}$-measurable. Here, to bound $\mathbb{E}_{t-1}[g_{t+\tau-1}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})]$, we consider

$$\mathbb{E}_{t-1}[g_{t+\tau-1}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})] = \mathbb{E}_{t-1}[g(\boldsymbol{x}, \xi_{t+\tau-1}) - g(\boldsymbol{x}, \xi)]$$

$$\leq \int_A H(d\mathbb{P}_{[t-1]}^{t+\tau-1} - d\mu) + \int_{A^C} H(d\mathbb{P}_{[t-1]}^{t+\tau-1} - d\mu)$$

$$\leq 2H\|\mathbb{P}_{[t-1]}^{t+\tau-1} - \mu\|_{TV}$$

where $\mathbb{P}_{[t-1]}^{t+\tau-1}$ is the probability measure of $\boldsymbol{\xi_{t+\tau-1}}$ conditional on $\mathcal{F}_{t-1}$, $\boldsymbol{\xi} \sim \mu$, and A is any measurable set. By the definition of $\tau = \tau_{\mathrm{mix}}(T^{-1})$, it follows that $\|\mathbb{P}_{[t-1]}^{t+\tau-1} - \mu\|_{TV} \leq 1/T$, which implies that $\mathbb{E}_{t-1}[g_{t+\tau-1}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})] \leq 2H/T$. From this, we deduce the desired statement. ∎

Next we provide an upper bound on the term $\sum_{t=1}^{T} \mathbb{E}\left[Q_t g_t(\boldsymbol{x})/V_t\right]$, as we mentioned in Section 3.1.

Lemma 12 *For any $\boldsymbol{x} \in \mathcal{X}$ such that $\bar{g}(\boldsymbol{x}) \leq 0$,*

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{Q_t}{V_t} g_t(\boldsymbol{x})\right] \leq \frac{2H(H+GR)(\tau+1)}{\tau_{mix}^{\beta}(1-\beta)}(T+1)^{1-\beta}.$$

Proof Consider the case $T < \tau$ first. Note that

$$\sum_{t=1}^{T} \frac{Q_t}{V_t} g_t(\boldsymbol{x}) \leq \frac{H^2}{\tau_{\mathrm{mix}}^{\beta}} \sum_{t=1}^{T} t^{1-\beta} \leq \frac{H^2}{\tau_{\mathrm{mix}}^{\beta}(2-\beta)}(T+1)^{2-\beta} \leq \frac{2H^2(\tau+1)}{\tau_{\mathrm{mix}}^{\beta}(1-\beta)}(T+1)^{1-\beta}$$

where the second inequality is due to Corollary 29 and the third inequality is from $T < \tau$. Next we consider the case $T \geq \tau$. Note that

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{Q_t}{V_t} g_t(\boldsymbol{x})\right] = \sum_{t=1}^{\tau-1} \mathbb{E}\left[\frac{Q_t}{V_t} g_t(\boldsymbol{x})\right] + \sum_{t=1}^{T-\tau+1} \mathbb{E}\left[\frac{Q_{t+\tau-1}}{V_{t+\tau-1}} g_{t+\tau-1}(\boldsymbol{x})\right].$$

Here the first term on the right-hand side can be bounded as

$$\sum_{t=1}^{\tau-1} \mathbb{E}\left[\frac{Q_t}{V_t} g_t(\boldsymbol{x})\right] \le \frac{H^2}{\tau_{\text{mix}}^\beta} \sum_{t=1}^{\tau-1} t^{1-\beta} \le \frac{H^2}{\tau_{\text{mix}}^\beta (2-\beta)} \tau^{2-\beta} \le \frac{H^2 \tau}{\tau_{\text{mix}}^\beta (1-\beta)} (T+1)^{1-\beta}$$

as above. Moreover, the second term can be bounded as follows.

$$\sum_{t=1}^{T-\tau+1} \mathbb{E}\left[\frac{Q_{t+\tau-1}}{V_{t+\tau-1}} g_{t+\tau-1}(\boldsymbol{x})\right]$$

$$\le \sum_{t=1}^{T-\tau+1} \mathbb{E}\left[\frac{(Q_{t+\tau-1}-Q_t)}{V_{t+\tau-1}} g_{t+\tau-1}(\boldsymbol{x})\right] + \sum_{t=1}^{T-\tau+1} \mathbb{E}\left[\frac{Q_t}{V_{t+\tau-1}} g_{t+\tau-1}(\boldsymbol{x})\right]$$

$$\le \frac{(\tau-1)H(H+GR)}{\tau_{\text{mix}}^\beta} \sum_{t=1}^{T-\tau+1} \frac{1}{(t+\tau-1)^\beta} + \sum_{t=1}^{T-\tau+1} \frac{1}{V_{t+\tau-1}} \mathbb{E}[Q_t \mathbb{E}_{t-1}[g_{t+\tau-1}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})]]$$

$$\le \frac{(\tau-1)H(H+GR)}{\tau_{\text{mix}}^\beta (1-\beta)} (T^{1-\beta} - (\tau-1)^{1-\beta}) + \frac{2H}{T} \sum_{t=1}^{T-\tau+1} \mathbb{E}\left[\frac{Q_t}{V_{t+\tau-1}}\right]$$

$$\le \frac{H(H+GR)}{\tau_{\text{mix}}^\beta (1-\beta)} \tau (T+1)^{1-\beta} + \frac{2H}{T} \sum_{t=1}^{T} \mathbb{E}\left[\frac{Q_t}{V_t}\right]$$

$$\le \frac{H(H+GR)\tau}{\tau_{\text{mix}}^\beta (1-\beta)} (T+1)^{1-\beta} + \frac{2H^2}{\tau_{\text{mix}}^\beta T} \sum_{t=1}^{T} t^{1-\beta}$$

$$\le \frac{H(H+GR)\tau}{\tau_{\text{mix}}^\beta (1-\beta)} (T+1)^{1-\beta} + \frac{2H^2}{\tau_{\text{mix}}^\beta T(2-\beta)} (T+1)^{2-\beta}$$

$$\le \frac{H(H+GR)(\tau+1)}{\tau_{\text{mix}}^\beta (1-\beta)} (T+1)^{1-\beta},$$

where the second inequality is from Lemma 10 and $\bar{g}(\boldsymbol{x}) \le 0$, the third inequality holds due to Corollary 29 and the choice of τ, the sixth inequality comes from Corollary 29, and the last inequality holds because $T(2-\beta) > (T+1)(1-\beta)$. ∎

Based on Lemmas 11 and 12, we can now prove the first part of Theorem 1, which upper bounds the regret of Algorithm 1 for online convex optimization with ergodic constraints.

Proof [The first part of Theorem 1] Lemma 21 implies that

$$\mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x}^*)\right] \le \frac{\alpha_T}{V_T} R^2 + \frac{F^2}{4} \sum_{t=1}^{T} \frac{V_t}{\alpha_t} + \frac{(H+GR)^2}{2} \sum_{t=1}^{T} \frac{1}{V_t} + \sum_{t=1}^{T} \mathbb{E}\left[\frac{Q_t}{V_t} g_t(\boldsymbol{x}^*)\right].$$

26

Next using Lemma 12 with $\boldsymbol{x}^*$ satisfying $\bar{g}(\boldsymbol{x}^*) \leq 0$ and plugging in our choice of α_t and V_t, we deduce the following.

$$
\mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x})\right]
$$
$$
\leq R^2(\tau_{\mathrm{mix}}T)^{1-\beta} + \frac{F^2\tau_{\mathrm{mix}}^{\beta-1}}{4\beta}T^{\beta} + \frac{(H+GR)^2\tau_{\mathrm{mix}}^{-\beta}}{2(1-\beta)}T^{1-\beta} + \frac{2H(H+GR)(\tau+1)}{\tau_{\mathrm{mix}}^{\beta}(1-\beta)}(T+1)^{1-\beta}
$$
$$
= O(\tau_{\mathrm{mix}}^{-\beta}\tau T^{1-\beta}),
$$

as required. ∎

Next, we prove the second part of Theorem 1, which gives an upper bound on constraint violation under Algorithm 1. The following lemma provides a time-varying bound on the expected virtual queue size.

Lemma 13 *For $t \in [T+1]$, $\mathbb{E}[Q_t]$ is bounded above by*

$$
\frac{3}{2}\sqrt{2I(t-1)} + \frac{\beta+3}{\beta+1}\sqrt{\frac{2FR\tau_{mix}^{\beta}t^{\beta+1}}{1+\beta}} + \frac{3R}{2}\sqrt{2\tau_{mix}(t-1)} + \frac{3H}{2}\sqrt{2(\tau-1)(t-1)} + 4H
$$

where $I = H^2 + G^2R^2 + F^2/4$.

Proof To prove the lemma, we argue by induction. Note that the statement of the lemma trivially holds when $t = 1$ because $Q_1 = 0$. Suppose the statement holds for $t \leq s$ for some $s \geq 1$. What remains is to provide an upper bound on $\mathbb{E}[Q_{s+1}]$. Note that

$$
\mathbb{E}\left[Q_t\left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right)\right] = \mathbb{E}\left[Q_t\mathbb{E}_{t-1}\left[g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right]\right]
$$
$$
\leq \mathbb{E}[Q_t\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]]
$$
$$
= \mathbb{E}[Q_tg_t(\boldsymbol{x})].
$$

Then Lemma 24 together with Jensen's inequality implies the following.

$$
\mathbb{E}[Q_{s+1}] \leq \sqrt{2s\left(H^2 + R^2G^2\right) + 2RF\sum_{t=1}^{s}V_t + 2R^2\alpha_s + 2\sum_{t=1}^{s}\mathbb{E}[Q_tg_t(\boldsymbol{x})]}.
$$

Moreover, $\sum_{t=1}^{s}\mathbb{E}[Q_tg_t(\boldsymbol{x})]$ can be upper bounded based on Lemma 11. Then it follows that

$$\mathbb{E}[Q_{s+1}]$$

$$\leq \sqrt{2Is + \frac{2FR\tau_{\mathrm{mix}}^{\beta}(s+1)^{\beta+1}}{1+\beta} + 2R^2\tau_{\mathrm{mix}}s + 2H(H+GR)(\tau-1)s + \frac{4H}{s}\sum_{t=1}^{s}\mathbb{E}[Q_t]}$$

$$\leq \sqrt{2Is} + \sqrt{\frac{2FR\tau_{\mathrm{mix}}^{\beta}(s+1)^{\beta+1}}{1+\beta}} + \sqrt{2R^2\tau_{\mathrm{mix}}s} + \sqrt{2H^2(\tau-1)s} + 2\sqrt{\frac{H}{s}\sum_{t=1}^{s}\mathbb{E}[Q_t]}$$

$$\leq \sqrt{2Is} + \sqrt{\frac{2FR\tau_{\mathrm{mix}}^{\beta}(s+1)^{\beta+1}}{1+\beta}} + \sqrt{2R^2\tau_{\mathrm{mix}}s} + \sqrt{2H^2(\tau-1)s} + 2H + \frac{1}{2s}\sum_{t=1}^{s}\mathbb{E}[Q_t]$$

for $s \leq T$. By the induction hypothesis, it follows that $\mathbb{E}[Q_t]$ for any $1 \leq t \leq s$ is upper bounded by

$$\frac{3}{2}\sqrt{2I(t-1)} + \frac{\beta+3}{\beta+1}\sqrt{\frac{2FR\tau_{\mathrm{mix}}^{\beta}t^{\beta+1}}{1+\beta}} + \frac{3R}{2}\sqrt{2\tau_{\mathrm{mix}}(t-1)} + \frac{3H}{2}\sqrt{2(\tau-1)(t-1)} + 4H$$

$$\leq \frac{3}{2}\sqrt{2Is} + \frac{\beta+3}{\beta+1}\sqrt{\frac{2FR\tau_{\mathrm{mix}}^{\beta}(s+1)^{\beta+1}}{1+\beta}} + \frac{3R}{2}\sqrt{2\tau_{\mathrm{mix}}s} + \frac{3H}{2}\sqrt{2(\tau-1)s} + 4H.$$

This leads to the desired upper bound on $\mathbb{E}[Q_{s+1}]$. ∎

We are now ready to prove the second part of Theorem 1, which proves the constraint violation bound of Algorithm 1.

Proof [The second part of Theorem 1] Combining Lemmas 22 and 13, we deduce that

$$\mathbb{E}\left[\sum_{t=1}^{T}g_t(\boldsymbol{x}_t)\right] \leq \mathbb{E}[Q_{T+1}] + \frac{FG}{2}\sum_{t=1}^{T}\frac{V_t}{\alpha_t} + \frac{G^2}{2}\sum_{t=1}^{T}\frac{\mathbb{E}[Q_t]}{\alpha_t} = O\left(\tau_{\mathrm{mix}}^{\beta/2}T^{\beta/2+1/2} + \sqrt{(\tau-1)T}\right),$$

as required. ∎

6.2 Ergodic Drift-Plus-Penalty for Stochastic-Constrained Stochastic Optimization

In this section, we provide our formal proof of Theorem 2. To better organize and present the result, we divide the analysis into two, one of which is for bounding the optimality gap while the other is for bounding the feasibility gap. The main idea is to use the performance bounds given in Theorem 1 and relate them to the optimality gap and the feasibility gap. The relation between them is not as clear as in the i.i.d. setting, but we use the mixing property of ergodic Markov chains.

Proof [The first part of Theorem 2] For $T \geq \tau = \tau_{\mathrm{mix}}(T^{-1})$, we have the following

$$
\sum_{t=1}^{T} \left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}) \right) = \sum_{t=1}^{T-\tau+1} \left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}) - f_{t+\tau-1}(\boldsymbol{x}_t) + f_{t+\tau-1}(\boldsymbol{x}) \right)
$$

$$
+ \sum_{t=1}^{T-\tau+1} \left(f_{t+\tau-1}(\boldsymbol{x}_t) - f_{t+\tau-1}(\boldsymbol{x_{t+\tau-1}}) \right)
$$

$$
+ \sum_{t=\tau}^{T} \left(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}) \right) + \sum_{t=T-\tau+2}^{T} \left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}) \right).
$$

We consider the four parts of the right-hand side separately. Here, the first part satisfies

$$
\mathbb{E}_{t-1} \left[\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}) - f_{t+\tau-1}(\boldsymbol{x}_t) + f_{t+\tau-1}(\boldsymbol{x}) \right]
$$

$$
= \int \left(f(\boldsymbol{x}_t, \xi) - f(\boldsymbol{x}, \xi) \right) \left(d\mu(\xi) - d\mathbb{P}_{[t-1]}^{t+\tau-1}(\xi) \right)
$$

$$
\leq FR \int |d\mu(\xi) - d\mathbb{P}_{[t-1]}^{t+\tau-1}(\xi)|
$$

$$
\leq 2FR \|\mu - \mathbb{P}_{[t-1]}^{t+\tau-1}\|_{TV}
$$

$$
\leq \frac{2FR}{T}.
$$

Hence, it follows that $\sum_{t=1}^{T-\tau+1} \mathbb{E}[\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}) - f_{t+\tau-1}(\boldsymbol{x}_t) + f_{t+\tau-1}(\boldsymbol{x})] \leq 2FR$. To bound the second part, we consider the following.

$$
\mathbb{E}[f_{t+\tau-1}(\boldsymbol{x}_t) - f_{t+\tau-1}(\boldsymbol{x_{t+\tau-1}})] = \sum_{s=t}^{t+\tau-2} \mathbb{E}[f_{t+\tau-1}(\boldsymbol{x_s}) - f_{t+\tau-1}(\boldsymbol{x_{s+1}})]
$$

$$
\leq \sum_{s=t}^{t+\tau-2} \mathbb{E}[F\|\boldsymbol{x_s} - \boldsymbol{x_{s+1}}\|]
$$

$$
\leq \sum_{s=t}^{t+\tau-2} \frac{F}{2\alpha_s} \mathbb{E}[V_s F + Q_s G]
$$

$$
= \frac{F^2 \tau_{\mathrm{mix}}^{\beta-1}}{2} \sum_{s=t}^{t+\tau-2} s^{\beta-1} + \frac{FG}{2\tau_{\mathrm{mix}}} \sum_{s=t}^{t+\tau-2} \frac{\mathbb{E}[Q_s]}{s}
$$

$$
\leq \frac{F^2 \tau_{\mathrm{mix}}^{\beta-1}}{2\beta} \left((t+\tau-2)^{\beta} - (t-1)^{\beta} \right) + \frac{FG}{2\tau_{\mathrm{mix}}} \sum_{s=t}^{t+\tau-2} \frac{\mathbb{E}[Q_s]}{s}
$$

where the last inequality holds due to Corollary 29. Here, we consider the following to provide an upper bound on the second term on the right-most side.

$$\sum_{s=t}^{t+\tau-2} \frac{\mathbb{E}[Q_s]}{s} \le 3\left(\sqrt{2I(t+\tau-2)} - \sqrt{2I(t-1)}\right)$$

$$+ \frac{2(\beta+3)}{(\beta+1)^2}\sqrt{\frac{2FR\tau_{\text{mix}}^{\beta}}{\beta+1}}\left(\sqrt{(t+\tau-2)^{\beta+1}} - \sqrt{(t-1)^{\beta+1}}\right)$$

$$+ 3R\left(\sqrt{2\tau_{\text{mix}}(t+\tau-2)} - \sqrt{2\tau_{\text{mix}}(t-1)}\right)$$

$$+ 3H\left(\sqrt{2(\tau-1)(t+\tau-2)} - \sqrt{2(\tau-1)(t-1)}\right)$$

$$+ 4H\log\left[\frac{t+\tau-2}{\sigma(t-1)}\right].$$

Here, we handle the above terms in the following manner. Note that

$$\sum_{t=1}^{T-\tau+1}\left(\sqrt{(t+\tau-2)^{\beta+1}} - \sqrt{(t-1)^{\beta+1}}\right)$$

$$= (T-1)^{\frac{\beta+1}{2}} + \cdots + (T-\tau+1)^{\frac{\beta+1}{2}} - (\tau-2)^{\frac{\beta+1}{2}} - \cdots - 1$$

$$\le (\tau-1)T^{\beta/2+1/2}$$

holds and that

$$\sum_{t=1}^{T-\tau+1}\left(\sqrt{(t+\tau-2)} - \sqrt{(t-1)}\right) = (T-1)^{\frac{1}{2}} + \cdots + (T-\tau+1)^{\frac{1}{2}} - (\tau-2)^{\frac{1}{2}} - \cdots - 1$$

$$\le (\tau-1)T^{1/2}.$$

The resulting terms form dominant terms, which means that

$$\sum_{t=1}^{T-\tau+1} \mathbb{E}[f_{t+\tau-1}(\boldsymbol{x}_t) - f_{t+\tau-1}(\boldsymbol{x_{t+\tau-1}})]$$

$$= O\left(\tau_{\text{mix}}^{\beta/2-1}(\tau-1)T^{\beta/2+1/2} + \tau_{\text{mix}}^{-1}(\tau-1)^{3/2}T^{1/2}\right).$$

Moreover, by Theorem 1 and the triangular inequality, the third part is bounded as

$$\mathbb{E}\left[\sum_{t=\tau}^{T}(f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}))\right] = O\left(\tau_{\text{mix}}^{-\beta}\tau T^{1-\beta} + (\tau-1)\right).$$

The fourth part can be bounded as follows.

$$\mathbb{E}\left[\sum_{t=T-\tau+2}^{T}\left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x})\right)\right] \le RF(\tau-1).$$

Combining the bounds on the four parts, we can conclude that for any $\boldsymbol{x} \in \mathcal{X}$ such that $\bar{g}(\boldsymbol{x}) \leq 0$,

$$\mathbb{E}\left[\sum_{t=1}^{T} \left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x})\right)\right]$$
$$= O\left(\tau_{\mathrm{mix}}^{-\beta}\tau T^{1-\beta} + \tau_{\mathrm{mix}}^{\beta/2-1}(\tau-1)T^{\beta/2+1/2} + \tau_{\mathrm{mix}}^{-1}(\tau-1)^{3/2}T^{1/2} + \tau\right),$$

which implies

$$\mathbb{E}\left[\bar{f}(\bar{\boldsymbol{x}}_T) - \bar{f}(\boldsymbol{x})\right] \leq \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T}\left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x})\right)\right]$$
$$= O\left(\frac{\tau}{\tau_{\mathrm{mix}}^{\beta}T^{\beta}} + \frac{\tau-1}{\tau_{\mathrm{mix}}}\frac{\tau_{\mathrm{mix}}^{\beta/2}}{T^{(1-\beta)/2}} + \frac{(\tau-1)^{3/2}}{\tau_{\mathrm{mix}}T^{1/2}} + \frac{\tau}{T}\right),$$

as required. ∎

Next we prove the feasibility gap bound of Algorithm 1, which is given as the second part of Theorem 2.

Proof [Proof of the second part of Theorem 2] For $T \geq \tau = \tau_{\mathrm{mix}}(T^{-1})$, we have the following,

$$\sum_{t=1}^{T} \bar{g}(\boldsymbol{x}_t) = \sum_{t=1}^{T-\tau+1}\left(\bar{g}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x}_t)\right) + \sum_{t=1}^{T-\tau+1}\left(g_{t+\tau-1}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x}_{t+\tau-1})\right)$$
$$+ \sum_{t=\tau}^{T} g_t(\boldsymbol{x}_t) + \sum_{t=T-\tau+2}^{T} \bar{g}(\boldsymbol{x}_t)$$

where the right-hand side consists of four terms. We separately upper bound the four parts of the right-hand side. As before, the first part satisfies

$$\mathbb{E}_{t-1}[\bar{g}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x}_t)] = \int g(\boldsymbol{x}_t, \xi)\left(d\mu(\xi) - d\mathbb{P}_{[t-1]}^{t+\tau-1}(\xi)\right)$$
$$\leq H\int|d\mu(\xi) - d\mathbb{P}_{[t-1]}^{t+\tau-1}(\xi)|$$
$$\leq 2H\|\mu - \mathbb{P}_{[t-1]}^{t+\tau-1}\|_{TV}$$
$$\leq \frac{2H}{T}.$$

This implies that $\sum_{t=1}^{T-\tau+1}\mathbb{E}[\bar{g}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x}_t)] \leq 2H$. Next, the second part satisfies

$$\mathbb{E}[g_{t+\tau-1}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x}_{t+\tau-1})] = \sum_{s=t}^{t+\tau-2}\mathbb{E}[g_{t+\tau-1}(\boldsymbol{x}_s) - g_{t+\tau-1}(\boldsymbol{x}_{s+1})]$$
$$\leq \sum_{s=t}^{t+\tau-2}\mathbb{E}[G\|\boldsymbol{x}_s - \boldsymbol{x}_{s+1}\|]$$

As in the proof of the first part of Theorem 2, we deduce that

$$\sum_{t=1}^{T-\tau+1} \mathbb{E}[g_{t+\tau-1}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x}_t)] \leq O\left(\tau_{\mathrm{mix}}^{\beta/2-1}(\tau-1)T^{\beta/2+1/2} + \tau_{\mathrm{mix}}^{-1}(\tau-1)^{3/2}T^{1/2}\right).$$

Moreover, it follows from Theorem 1 and the triangular inequality that the third part can be bounded as

$$\mathbb{E}\left[\sum_{t=\tau+2}^{T} g_t(\boldsymbol{x}_t)\right] = O\left(\tau_{\mathrm{mix}}^{\beta/2}T^{\beta/2+1/2} + \sqrt{(\tau-1)T} + \tau - 1\right).$$

Lastly, the following provides an upper bound on the fourth part.

$$\mathbb{E}\left[\sum_{t=T-\tau+2}^{T} \bar{g}(\boldsymbol{x}_t)\right] \leq H(\tau-1).$$

Combining the bounds on the four parts, we have that

$$\mathbb{E}\left[\sum_{t=1}^{T} \bar{g}(\boldsymbol{x}_t)\right] \leq O\left(\tau_{\mathrm{mix}}^{\beta/2}T^{\beta/2+1/2} + \tau_{\mathrm{mix}}^{\beta/2-1}(\tau-1)T^{\beta/2+1/2} + \tau_{\mathrm{mix}}^{-1}(\tau-1)^{3/2}T^{1/2} + \tau\right),$$

implying in turn that

$$\mathbb{E}\left[\bar{g}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})\right] \leq \frac{1}{T}\mathbb{E}\left[\sum_{t=1}^{T} \bar{g}(\boldsymbol{x}_t)\right] = O\left(\frac{\tau_{\mathrm{mix}}^{\beta/2-1}\tau}{T^{(1-\beta)/2}} + \frac{(\tau-1)^{3/2}}{\tau_{\mathrm{mix}}T^{1/2}} + \frac{\tau}{T}\right),$$

as required. ■

6.3 Ergodic Drift-Plus-Penalty under Slater's Condition

In this section, we prove Theorem 3, which analyzes the performance of Theorem 3 under Slater's constraint qualification stated in Assumption 3. We will see that Slater's condition does lead to improvement. Basically, we deduce a reduction on $\mathbb{E}[Q_t]$. To argue this, we follow the proof path of Yu et al. (2017). We first present a lemma, which is analogous to (Yu et al., 2017, Lemma 5). The difference is that we use a time-varying parameter $\theta(t)$ which allows time-varying algorithm parameters V_t and α_t. In fact, its proof is similar to that of (Yu et al., 2017, Lemma 5), so we defer it to the appendix (Appendix D).

Lemma 14 *Let $\{Z_t, t \geq 0\}$ be a discrete-time stochastic process adapted to a filtration $\{\mathcal{W}_t, t \geq 0\}$ with $Z_0 = 0$ and $\mathcal{W}_0 = \{\emptyset, \mathcal{S}\}$. Suppose there exists a positive integer t_0, real numbers $0 < \zeta \leq \delta_{\max}$, and a non-decreasing function $\theta(t) > 0$ such that*

$$|Z_{t+1} - Z_t| \leq \delta_{\max},$$

$$\mathbb{E}[Z_{t+t_0} - Z_t|\mathcal{W}_t] \leq \begin{cases} t_0\delta_{\max}, & \text{if } Z_t < \theta(t) \\ -t_0\zeta, & \text{if } Z_t \geq \theta(t) \end{cases}$$

for all positive integer t. Then,

$$\mathbb{E}[Z_t] \leq \theta(t) + t_0 \delta_{\max} + t_0 \frac{4\delta_{\max}^2}{\zeta} \log\left(\frac{8\delta_{\max}^2}{\zeta^2}\right)$$

for all positive integer t.

Here, the important part is that although $\theta(t)$ is time-varying, the parameter t_0 is some fixed value that does not depend on t. That is why the condition in Lemma 14 can be recursively applied.

Lemma 14 implies that if the stochastic process given by $\{Q_t, t \geq 0\}$ where $Q_0 = 0$ satisfies the drift condition as in Lemma 14 with appropriate parameters $\delta_{\max}$, t_0, and ζ, then the expected queue size $\mathbb{E}[Q_t]$ can be bounded. Moreover, by properly setting $\theta(t)$ as well as the parameters, we may control the size of $\mathbb{E}[Q_t]$. The next lemma shows that the stochastic process given by $\{Q_t, t \geq 0\}$ indeed satisfies the desired drift condition.

While proving the lemma, we need to consider the term

$$\mathbb{E}_{t-1}[Q_i g_i(\hat{\boldsymbol{x}})]$$

for $i \geq t$ where $\hat{\boldsymbol{x}}$ is the solution satisfying Slater's constraint qualification, i.e. $\bar{g}(\hat{\boldsymbol{x}}) \leq -\epsilon$. Here, we need to relate $\bar{g}(\hat{\boldsymbol{x}}) \leq -\epsilon$ and the term $\mathbb{E}_{t-1}[Q_i g_i(\hat{\boldsymbol{x}})]$, from which we deduce the drift condition. Although $g_i(\hat{\boldsymbol{x}})$ is not necessarily negative, we again use the property of ergodic Markov chains that the distribution of $g_i(\hat{\boldsymbol{x}})$ conditional on $\mathcal{F}_{t-1}$ gets close to the stationary distribution for a sufficiently large i.

Lemma 15 *Let $T \geq 4H/\epsilon$, $t_0 > 2\tau = 2\tau_{mix}(T^{-1})$, and*

$$\theta_{t_0}(t) = \frac{2}{\epsilon(t_0 - 2\tau)}\left(\epsilon(t_0 - \tau)^2(H + GR) + 4\tau(t_0 + t - 1)(H + GR)^2\right)$$

$$+ \frac{2}{\epsilon(t_0 - 2\tau)}\left(2t_0(H^2 + G^2R^2) + 2RF\tau_{mix}^{\beta}t_0(t + t_0 - 1)^{\beta}\right).$$

Then Algorithm 1 under Assumption 3 satisfies

$$|Q_{t+1} - Q_t| \leq H + GR,$$

$$\mathbb{E}_{t-1}[Q_{t+t_0} - Q_t] \leq \begin{cases} (H + GR)t_0, & \text{if } Q_t < \theta_{t_0}(t) \\ -\epsilon t_0/4, & \text{if } Q_t \geq \theta_{t_0}(t). \end{cases}$$

Proof $|Q_{t+1} - Q_t| \leq H + GR$ follows directly from Lemma 10.

For the second part, note that by Jensen's inequality, we have

$$\mathbb{E}_{t-1}[Q_{t+t_0}]^2 \leq \mathbb{E}_{t-1}[Q_{t+t_0}^2].$$

Moreover, observe that

$$\theta_{t_0}(t) \geq \frac{2\epsilon(t_0 - \tau)^2(H + GR)}{\epsilon(t_0 - 2\tau)} \geq (t_0 - \tau)(H + GR) \geq \frac{\epsilon t_0}{2},$$

where the last inequality holds because $t_0 > 2\tau$ and $H > \epsilon$. This means that $Q_t > \epsilon t_0/4$ if $Q_t > \theta(t_0)$. Therefore, it is sufficient to show that if $Q_t \geq \theta(t_0)$,

$$\mathbb{E}_{t-1}[Q_{t+t_0}^2] \leq \left(Q_t - \frac{\epsilon t_0}{4} \right)^2 .$$

Recall that By Lemma 23 and the convexity of g_t, we have for any $i \geq 1$,

$$\Delta_i \leq H^2 + G^2 R^2 + Q_i g_i(\hat{\boldsymbol{x}}) + V_i RF + \alpha_i \left(D(\hat{\boldsymbol{x}}, \boldsymbol{x_i}) - D(\hat{\boldsymbol{x}}, \boldsymbol{x_{i+1}}) \right) .$$

Thus, we have

$$\sum_{i=t}^{t+t_0-1} \mathbb{E}_{t-1}[\Delta_i]$$

$$\leq t_0(H^2 + G^2 R^2) + \sum_{i=t}^{t+t_0-1} \mathbb{E}_{t-1}[Q_i g_i(\hat{\boldsymbol{x}})] + RF\tau_{\text{mix}}^{\beta} \sum_{i=t}^{t+t_0-1} i^{\beta}$$

$$+ \alpha_t D(\hat{\boldsymbol{x}}, \boldsymbol{x_t}) + \sum_{i=t}^{t+t_0-2} (\alpha_{i+1} - \alpha_i) D(\hat{\boldsymbol{x}}, \boldsymbol{x_{i+1}}) - \alpha_{t+t_0-1} D(\hat{\boldsymbol{x}}, x_{t+t_0})$$

$$\leq t_0(H^2 + G^2 R^2) + RF\tau_{\text{mix}}^{\beta} t_0(t + t_0 - 1)^{\beta} + \alpha_{t+t_0-1} R^2 + \sum_{i=t}^{t+t_0-1} \mathbb{E}_{t-1}[Q_i g_i(\hat{\boldsymbol{x}})].$$

We factor the last term as follows.

$$\sum_{i=t}^{t+t_0-1} \mathbb{E}_{t-1}[Q_i g_i(\hat{\boldsymbol{x}})]$$

$$= \sum_{i=t}^{t+\tau-2} \mathbb{E}_{t-1}[Q_i g_i(\hat{\boldsymbol{x}})] + \sum_{i=t}^{t+t_0-\tau} \mathbb{E}_{t-1}[(Q_{i+\tau-1} - Q_i)g_{i+\tau-1}(\hat{\boldsymbol{x}})]$$

$$+ \sum_{i=t}^{t+t_0-\tau} \mathbb{E}_{t-1}[Q_i(g_{i+\tau-1}(\hat{\boldsymbol{x}}) - \bar{g}(\hat{\boldsymbol{x}}))] + \sum_{i=t}^{t+t_0-\tau} \mathbb{E}_{t-1}[Q_i \bar{g}(\hat{\boldsymbol{x}})]$$

$$\leq H^2 \sum_{i=t-1}^{t+\tau-3} i + (t_0 - \tau + 1)(\tau - 1)H(H + GR) + \left(\frac{2H}{T} - \epsilon \right) \sum_{i=t}^{t+t_0-\tau} \mathbb{E}_{t-1}[Q_i]$$

where the inequality is due to

$$\mathbb{E}_{t-1}[\bar{g}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x}_t)] \leq H \int |d\mu(\xi) - d\mathbb{P}_{[t-1]}^{t+\tau-1}(\xi)| \leq 2H\|\mu - \mathbb{P}_{[t-1]}^{t+\tau-1}\|_{TV} \leq \frac{2H}{T}.$$

Then, since $T \geq 4H/\epsilon$ and thus $2H/T \leq \epsilon/2$, it follows that

$$\sum_{i=t}^{t+t_0-1} \mathbb{E}_{t-1}[Q_i g_i(\hat{\boldsymbol{x}})]$$

$$\leq (t_0 + t - 2)(\tau - 1)H(H + GR) - \frac{\epsilon}{2}\sum_{i=t}^{t+t_0-\tau} \mathbb{E}_{t-1}[Q_i]$$

$$\leq (t_0 + t - 2)(\tau - 1)H(H + GR) - \frac{\epsilon}{2}\sum_{i=0}^{t_0-\tau} (Q_t - i(H + GR))$$

$$\leq (t_0 + t - 2)(\tau - 1)H(H + GR) - \frac{\epsilon}{2}\left((t_0 - \tau)Q_t - (t_0 - \tau)^2(H + GR)\right).$$

As a result, we obtain

$$\mathbb{E}_{t-1}[Q_{t+t_0}^2] = Q_t^2 + 2\sum_{i=t}^{t+t_0-1} \mathbb{E}_{t-1}[\Delta_i]$$

$$\leq Q_t^2 - \epsilon(t_0 - \tau)Q_t + \epsilon(t_0 - \tau)^2(H + GR) + 2(t_0 + t - 2)(\tau - 1)H(H + GR)$$

$$+ 2t_0(H^2 + G^2R^2) + 2RF\tau_{\mathrm{mix}}^{\beta}t_0(t + t_0 - 1)^{\beta} + 2\tau_{\mathrm{mix}}(t + t_0 - 1)R^2$$

$$\leq Q_t^2 - \epsilon(t_0 - \tau)Q_t + \epsilon(t_0 - \tau)^2(H + GR) + 4\tau(t_0 + t - 1)(H + GR)^2$$

$$+ 2t_0(H^2 + G^2R^2) + 2RF\tau_{\mathrm{mix}}^{\beta}t_0(t + t_0 - 1)^{\beta}$$

If Q_t satisfies

$$\frac{\epsilon}{2}(t_0 - 2\tau)Q_t \geq \epsilon(t_0 - \tau)^2(H + GR) + 4\tau(t_0 + t - 1)(H + GR)^2$$

$$+ 2t_0(H^2 + G^2R^2) + 2RF\tau_{\mathrm{mix}}^{\beta}t_0(t + t_0 - 1)^{\beta},$$

which is equivalent to $Q_t \geq \theta_{t_0}(t)$, then

$$\mathbb{E}_{t-1}[Q_{t+t_0}^2] \leq Q_t^2 - \frac{\epsilon}{2}t_0Q_t \leq \left(Q_t - \frac{\epsilon t_0}{4}\right)^2$$

as required. ∎

Having prepared Lemmas 14 and 15, we can derive refined bounds on the expected virtual queue size. Based on this, we are ready to prove Theorem 3.

Proof [Proof of Theorem 3] As the result is trivial when $T \leq 4H/\epsilon$ where H and ϵ are constants, we may assume without loss of generality that $T \geq 4H/\epsilon$. Moreover, we will use Lemma 14 for many values of t_0 greater than 2τ. Setting $\delta_{\max} = H + GR$, $\zeta = \epsilon/4$, and $\theta = \theta_{t_0}$ where θ_{t_0} is defined as in Lemma 15, Lemma 14 gives us that

$$\mathbb{E}[Q_{t+1}] \leq \theta_{t_0}(t) + t_0(H + GR) + t_0\frac{16(H + GR)^2}{\epsilon}\log\left(\frac{128(H + GR)^2}{\epsilon^2}\right).$$

In particular, as this inequality holds for any $t_0 > 2\tau$, the inequality with $t_0 = 2\tau + \lceil\sqrt{\tau_{\mathrm{mix}}t}\rceil$ holds. Recall that in Lemma 15, we set

$$\theta_{t_0}(t) = \frac{2}{\epsilon(t_0 - 2\tau)}\left(\epsilon(t_0 - \tau)^2(H + GR) + 4\tau(t_0 + t - 1)(H + GR)^2\right)$$
$$+ \frac{2}{\epsilon(t_0 - 2\tau)}\left(2t_0(H^2 + G^2R^2) + 2RF\tau_{\mathrm{mix}}^\beta t_0(t + t_0 - 1)^\beta\right).$$

Moreover,

$$\frac{\epsilon(t_0 - 2\tau)}{2}\theta_{t_0}(t) = \frac{\epsilon\lceil\sqrt{\tau_{\mathrm{mix}}t}\rceil}{2}\theta_{t_0}(t) \geq \frac{\epsilon\sqrt{\tau_{\mathrm{mix}}t}}{2}\theta_{t_0}(t).$$

This implies the following.

$$\frac{\epsilon\sqrt{\tau_{\mathrm{mix}}t}}{2}\theta_{t_0}(t)$$
$$\leq \epsilon(\tau + \lceil\sqrt{\tau_{\mathrm{mix}}t}\rceil)^2(H + GR) + 4\tau(2\tau + \lceil\sqrt{\tau_{\mathrm{mix}}t}\rceil + t - 1)(H + GR)^2$$
$$+ 2(2\tau + \lceil\sqrt{\tau_{\mathrm{mix}}t}\rceil)(H^2 + G^2R^2) + 2RF\tau_{\mathrm{mix}}^\beta(2\tau + \lceil\sqrt{\tau_{\mathrm{mix}}t}\rceil)(t + 2\tau + \lceil\sqrt{\tau_{\mathrm{mix}}t}\rceil - 1)^\beta$$
$$\leq 2\epsilon(H + GR)(\tau^2 + 4\tau_{\mathrm{mix}}t) + 4\tau(2\tau + 2\sqrt{\tau_{\mathrm{mix}}t} + t)(H + GR)^2$$
$$+ 4(\tau + \sqrt{\tau_{\mathrm{mix}}t})(H^2 + G^2R^2) + 4RF\tau_{\mathrm{mix}}^\beta(\tau + \sqrt{\tau_{\mathrm{mix}}t})(t + 2\tau + 2\sqrt{\tau_{\mathrm{mix}}t})^\beta$$
$$\leq \underbrace{2\epsilon(H + GR)(\tau^2 + 4\tau_{\mathrm{mix}}t)}_{(a)} + \underbrace{4\tau(2\tau + \tau_{\mathrm{mix}} + 2t)(H + GR)^2}_{(b)}$$
$$+ \underbrace{4(\tau + \sqrt{\tau_{\mathrm{mix}}t})(H^2 + G^2R^2)}_{(c)} + \underbrace{4RF\tau_{\mathrm{mix}}^\beta(\tau + \sqrt{\tau_{\mathrm{mix}}t})(2t + 2\tau + \tau_{\mathrm{mix}})^\beta}_{(d)}$$

where the second inequality follows from $(a + b)^2 \leq 2(a^2 + b^2)$, $\lceil x\rceil \leq 2x$ which holds for $x \geq 1$ and the third inequality follows from AM-GM inequality. We see that term (c) grows more slowly than term (d) which is of order

$$O\left(\tau_{\mathrm{mix}}^\beta(\tau + t)^\beta(\tau + \sqrt{\tau_{\mathrm{mix}}t})\right) = O\left(\tau_{\mathrm{mix}}^\beta(\tau^\beta + t^\beta)(\tau + \sqrt{\tau_{\mathrm{mix}}t})\right).$$

Here, we used the known fact that $(a + b)^\beta \leq a^\beta + b^\beta$ for $a, b \geq 0$ and $\beta < 1$. Likewise, term (a) grows more slowly than term (b) which is of order

$$O\left(\tau(\tau + t)\right).$$

Then it follows that

$$\theta_{t_0}(t) = O\left(\tau_{\mathrm{mix}}^\beta(\tau^\beta + t^\beta)\left(1 + \frac{\tau}{\sqrt{\tau_{\mathrm{mix}}t}}\right) + \frac{\tau(\tau + t)}{\sqrt{\tau_{\mathrm{mix}}t}}\right)$$

$$= O\left(\underbrace{\tau_{\mathrm{mix}}^\beta\tau^\beta}_{(a')} + \underbrace{\tau_{\mathrm{mix}}^\beta t^\beta}_{(b')} + \underbrace{\frac{\tau^{1+\beta}\tau_{\mathrm{mix}}^\beta}{\sqrt{\tau_{\mathrm{mix}}t}}}_{(c')} + \underbrace{\frac{\tau\tau_{\mathrm{mix}}^\beta t^\beta}{\sqrt{\tau_{\mathrm{mix}}t}}}_{(d')} + \underbrace{\frac{\tau^2}{\sqrt{\tau_{\mathrm{mix}}t}}}_{(e')} + \underbrace{\frac{\tau t}{\sqrt{\tau_{\mathrm{mix}}t}}}_{(f')}\right).$$

Here, term (c') grows more slowly than term (e'), and term (f') dominates term (b'). Also, terms (a') and (d') are dominated by

$$\tau = O\left(\tau\sqrt{\frac{\tau}{\tau_{\text{mix}}}}\right) = O\left((e') + (f')\right),$$

which is due to the AM-GM inequality. Therefore,

$$\theta_{t_0}(t) = O\left(\frac{\tau(\tau + t)}{\sqrt{\tau_{\text{mix}}t}}\right).$$

Furthermore, since

$$t_0 = 2\tau + \lceil\sqrt{\tau_{\text{mix}}t}\rceil = O\left(\frac{\tau(\tau + t)}{\sqrt{\tau_{\text{mix}}t}}\right)$$

which holds because $\tau(\tau + t) \geq 2\tau\sqrt{\tau t}$, we have

$$\mathbb{E}[Q_t] \leq O\left(\frac{\tau(\tau + t)}{\sqrt{\tau_{\text{mix}}t}}\right), \tag{2}$$

which essentially gives rise to the desired refined bound on the expected virtual queue size.

Having provided the bound on the expected queue size, we now proceed to provide bounds on the performance guarantees of Algorithm 1. First of all, note that the regret bound given by Theorem 1 still holds with $\beta = 1/2$, as we use the same algorithm parameters. Then we deduce that

$$\mathbb{E}\left[\text{Regret}(T)\right] = O\left(\frac{\tau\sqrt{T}}{\sqrt{\tau_{\text{mix}}}}\right).$$

For constraint violation, we deduce from Lemma 22 and (2) that

$$\mathbb{E}\left[\text{Violation}(T)\right] \leq \mathbb{E}[Q_{T+1}] + \frac{FG}{2}\sum_{t=1}^{T}\frac{V_t}{\alpha_t} + \frac{G^2}{2}\sum_{t=1}^{T}\frac{\mathbb{E}[Q_t]}{\alpha_t}$$

$$= O\left(\frac{\tau(\tau + T + 1)}{\sqrt{\tau_{\text{mix}}(T+1)}} + \tau_{\text{mix}}^{\beta-1}\sum_{t=1}^{T}t^{\beta-1} + \frac{1}{\tau_{\text{mix}}}\sum_{t=1}^{T}\frac{\tau(\tau + t)}{t\sqrt{\tau_{\text{mix}}t}}\right)$$

$$= O\left(\frac{\tau(\tau + T)}{\sqrt{\tau_{\text{mix}}T}}\right).$$

Next, we consider the optimality gap. In the proof of Theorem 2, we argued that

$$\mathbb{E}[f_{t+\tau-1}(\boldsymbol{x}_t) - f_{t+\tau-1}(\boldsymbol{x}_{t+\tau-1})] \leq \frac{F^2\tau_{\text{mix}}^{\beta-1}}{2\beta}\left((t + \tau - 2)^{\beta} - (t-1)^{\beta}\right) + \frac{FG}{2\tau_{\text{mix}}}\sum_{s=t}^{t+\tau-2}\frac{\mathbb{E}[Q_s]}{s}.$$

Here, using (2) again, the second term on the right-hand side can be bounded as follows.

$$\frac{1}{\tau_{\text{mix}}}\sum_{s=t}^{t+\tau-2}\frac{\mathbb{E}[Q_s]}{s}$$

$$= O\left(\frac{\tau^2}{\tau_{\text{mix}}^{3/2}}\left(\underbrace{\lambda((t-1)^{-1/2}) - \lambda((t+\tau-2)^{-1/2})}_{(a'')}\right) + \frac{\tau}{\tau_{\text{mix}}^{3/2}}(\sqrt{t+\tau-2} - \sqrt{t-1})\right)$$

where
$$\lambda(x) = \begin{cases} x, & \text{if } x > 0 \\ 3/2, & \text{if } x = 0 \end{cases}.$$

Therefore, it follows that

$$\sum_{t=1}^{T-\tau+1} \mathbb{E}[f_{t+\tau-1}(\boldsymbol{x}_t) - f_{t+\tau-1}(\boldsymbol{x_{t+\tau-1}})] = O\left(\frac{\tau_{\text{mix}}^{\beta-1}}{2\beta} \left((T-1)^\beta + \cdots + (T-\tau+1)^\beta \right) \right.$$
$$+ \frac{\tau^2}{\tau_{\text{mix}}^{3/2}} \left(\frac{3}{2} + 1^{-1/2} + \cdots + (\tau-2)^{-1/2} \right) \qquad (3)$$
$$\left. + \frac{\tau}{\tau_{\text{mix}}^{3/2}} \left(\sqrt{T-1} + \cdots + \sqrt{T-\tau+1} \right) \right)$$

if $\tau \geq 2$. Thus, we deduce

$$\sum_{t=1}^{T-\tau+1} \mathbb{E}[f_{t+\tau-1}(\boldsymbol{x}_t) - f_{t+\tau-1}(\boldsymbol{x_{t+\tau-1}})]$$
$$= O\left(\tau_{\text{mix}}^{\beta-1}(\tau-1)T^\beta + \tau^2\sqrt{\tau-1}\tau_{\text{mix}}^{-3/2} + \tau(\tau-1)\tau_{\text{mix}}^{-3/2}\sqrt{T} \right)$$
$$= O\left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau^2\sqrt{T}}{\tau_{\text{mix}}^{3/2}} \right).$$

We combine this result with the other parts of the proof of the first part of Theorem 2 to get

$$\mathbb{E}\left[\sum_{t=1}^{T} \left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}) \right) \right] = O\left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau^2\sqrt{T}}{\tau_{\text{mix}}^{3/2}} + \tau_{\text{mix}}^{-\beta}\tau T^{1-\beta} + \tau \right) = O\left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau^2\sqrt{T}}{\tau_{\text{mix}}^{3/2}} \right)$$

for any $\boldsymbol{x} \in \mathcal{X}$ such that $\bar{g}(\boldsymbol{x}) \leq 0$ since $\beta = 1/2$. Similarly, we get

$$\sum_{t=1}^{T-\tau+1} \mathbb{E}[g_{t+\tau-1}(\boldsymbol{x}_t) - g_{t+\tau-1}(\boldsymbol{x_{t+\tau-1}})] = O\left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau^2\sqrt{T}}{\tau_{\text{mix}}^{3/2}} \right).$$

As before, we combine this result with the other parts of the proof of the second part of Theorem 2. Then we get

$$\mathbb{E}\left[\sum_{t=1}^{T} \bar{g}(\boldsymbol{x}_t) \right] = O\left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau^2\sqrt{T}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau(\tau+T)}{\sqrt{\tau_{\text{mix}}T}} + \tau \right) = O\left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau^2\sqrt{T}}{\tau_{\text{mix}}^{3/2}} + \frac{\tau^2}{\sqrt{\tau_{\text{mix}}T}} \right).$$

Finally, we obtain that

$$\mathbb{E}[\text{Gap}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = O\left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}T} + \frac{\tau^2}{\tau_{\text{mix}}^{3/2}\sqrt{T}} \right),$$

$$\mathbb{E}[\text{Infeasibility}(\bar{\boldsymbol{x}}_{\boldsymbol{T}})] = \left(\frac{\tau^{5/2}}{\tau_{\text{mix}}^{3/2}T} + \frac{\tau^2}{\tau_{\text{mix}}^{3/2}\sqrt{T}} + \frac{\tau^2}{\sqrt{\tau_{\text{mix}}}T^{3/2}} \right),$$

as required. $\blacksquare$

7 Analysis of MLMC Adaptive Drift-Plus-Penalty for the Unknown Mixing Time Case

This section presents a complete proof of Theorem 8. Section 7.1 provides bounds on several terms that involve the virtual queue size. In Section 7.2, we state the proof of Lemma 6, and lastly, we prove Theorem 8.

7.1 Controlling the Expected Virtual Queue Size

In this section we provide upper bounds on terms $\mathbb{E}[Q_t/V_t]$ and $\mathbb{E}[Q_t]$ under the MLMC Adaptive Drift-Plus-Penalty algorithm (Algorithm 2). Lemma 16 gives a refined bound on the term $\mathbb{E}[g_t(\boldsymbol{x})]$. Using this, Lemma 17 derives an adaptive upper bound on the term $\mathbb{E}[Q_t/V_t]$, and Lemma 18 deduces an adaptive upper bound on the term $\mathbb{E}[Q_t]$.

Note that if $\bar{x} \in \mathcal{X}$ satisfies $\bar{g}(\boldsymbol{x}) \leq 0$, then $g_t(\boldsymbol{x}) \leq g_t(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})$. Then we deduce

$$\mathbb{E}_{t-1}[g_t(\boldsymbol{x})] \leq \mathbb{E}_{t-1}[g_t(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})] = \mathbb{E}_{t-1}\left[g_t^{2^{j\max}}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})\right] \leq \mathbb{E}_{t-1}\left[|g_t^{2^{j\max}}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})|\right].$$

where the equality is due to Lemma 4. Applying Lemma 5 together with Jensen's inequality on the right-hand side, we obtain

$$\mathbb{E}_{t-1}[g_t(\boldsymbol{x})] \leq C(T)\tau_{\mathrm{mix}}^{1/2}T^{-1} = \tilde{O}\left(\tau_{\mathrm{mix}}^{1/2}T^{-1}\right). \tag{4}$$

In fact, we may derive a tighter bound on the term $\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]$ as follows.

Lemma 16 *For any $\boldsymbol{x} \in \mathcal{X}$ such that $\bar{g}(\boldsymbol{x}) \leq 0$,*

$$\mathbb{E}_{t-1}[g_t(\boldsymbol{x})] \leq D(T)\tau_{mix}^{1/4}T^{-1/2},$$

where $D(T) = \sqrt{C(T)H(4\log_2 T + 3)}$

Proof Note that $g_t(\boldsymbol{x}) \leq |g_t(\boldsymbol{x})| \leq (2N_t + 1)H$ by the definition of g_t. This implies that

$$\mathbb{E}_{t-1}[g_t(\boldsymbol{x})] \leq \mathbb{E}_{t-1}[(2N_t + 1)H] = (4\log_2 T + 3)H,$$

where the equality follows from Lemma 4 and the fact that N_t is independent of $\mathcal{F}_{t-1}$. Moreover, as we argued above, we have

$$\mathbb{E}_{t-1}[g_t(\boldsymbol{x})] \leq \mathbb{E}_{t-1}\left[|g_t^{2^{j\max}}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})|\right] \leq \sqrt{\mathbb{E}_{t-1}\left[|g_t^{2^{j\max}}(\boldsymbol{x}) - \bar{g}(\boldsymbol{x})|^2\right]} \leq C(T)\tau_{\mathrm{mix}}^{1/2}T^{-1},$$

where the last inequality follows from Jensen's inequality and the equality follows from Lemma 5. By taking the geometric mean of these two upper bounds, we get that

$$\mathbb{E}_{t-1}[g_t(\boldsymbol{x})] \leq \sqrt{C(T)H(4\log_2 T + 3)}\tau_{\mathrm{mix}}^{1/4}T^{-1/2},$$

as required. ∎

Next, Lemma 24 implies that Q_{s+1} is at most

$$\underbrace{\sqrt{2\sum_{t=1}^{s} RV_t F_t}}_{(a)} + \underbrace{\sqrt{2\sum_{t=1}^{s}(R^2 G_t^2 + H_t^2)}}_{(b)} + \underbrace{\sqrt{2R^2\alpha_s}}_{(c)} + \sqrt{2\sum_{t=1}^{s} Q_t(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t))}.$$

Here, term (c) is equal to $\sqrt{2S_{s-1}}$ which is at most $\sqrt{2S_s}$, and term (b) is at most $\sqrt{2S_s}$. For term (a), we have

$$(a)^2 = 2\sum_{t=1}^{s} S_{t-1}^\beta F_t \leq 2S_s^\beta \sum_{t=1}^{s} F_t \leq 2S_s^\beta \sqrt{s\sum_{t=1}^{s} F_t^2} \leq 2\sqrt{s}S_s^{\beta+1/2},$$

where the second inequality holds by the power mean inequality. Thus,

$$Q_{s+1}^2 \leq 2\sqrt{s}S_s^{\beta+1/2} + 4S_s + 2\sum_{t=1}^{s} Q_t(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)), \tag{5}$$

which in turn implies that

$$Q_{s+1} \leq 2\sqrt{2}S_s^{1/2} + \sqrt{2}s^{1/4}S_s^{\beta/2+1/4} + \sqrt{2\sum_{t=1}^{s} Q_t\left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right)} \tag{6}$$

Moreover, (6) implies that

$$\frac{Q_{s+1}}{V_{s+1}} \leq 2\sqrt{2}RS_s^{1/2-\beta} + \sqrt{2}Rs^{1/4}S_s^{1/4-\beta/2} + \sqrt{\frac{2}{V_{s+1}}\sum_{t=1}^{s}\frac{Q_t}{V_t}\left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right)}$$

as V_t is non-decreasing. Taking the expectation, applying Jensen's inequality, and using the fact that $V_{s+1} \geq s\delta$, it follows that

$$\mathbb{E}\left[\frac{Q_{s+1}}{V_{s+1}}\right] \leq 2\sqrt{2}R\mathbb{E}[S_s]^{1/2-\beta} + \sqrt{2}Rs^{1/4}\mathbb{E}[S_s]^{1/4-\beta/2}$$

$$+ \sqrt{\frac{2R}{(s\delta)^\beta}\sum_{t=1}^{s}\mathbb{E}\left[\frac{Q_t}{V_t}\mathbb{E}_{t-1}[g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)]\right]}.$$

Here, the last term on the right-hand side can be further upper bounded as follows.

$$\sqrt{\frac{2R}{(s\delta)^\beta}\sum_{t=1}^{s}\mathbb{E}\left[\frac{Q_t}{V_t}\mathbb{E}_{t-1}[g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)]\right]} \leq \sqrt{\frac{2R}{(s\delta)^\beta}\sum_{t=1}^{s}\mathbb{E}\left[\frac{Q_t}{V_t}\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]\right]}$$

where we used the fact that

$$\begin{aligned}
\mathbb{E}_{t-1}[g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)] &= \mathbb{E}_{t-1}\left[g^{j_{\max}}(\boldsymbol{x}_t) + \nabla g^{j_{\max}}(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right] \\
&\leq \mathbb{E}_{t-1}\left[g^{j_{\max}}(\boldsymbol{x})\right] \\
&= \mathbb{E}_{t-1}[g_t(\boldsymbol{x})]
\end{aligned} \tag{7}$$

40

holds due to Lemma 4. Furthermore,

$$
\sqrt{\frac{2R}{(s\delta)^\beta} \sum_{t=1}^{s} \mathbb{E}\left[\frac{Q_t}{V_t}\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]\right]} \leq \sqrt{\frac{2R}{(s\delta)^\beta} D(T)\tau_{\text{mix}}^{1/4}T^{-1/2} \sum_{t=1}^{s} \mathbb{E}\left[\frac{Q_t}{V_t}\right]}
$$

$$
\leq \frac{2Rs}{(s\delta)^\beta} D(T)\tau_{\text{mix}}^{1/4}T^{-1/2} + \frac{1}{2s}\sum_{t=1}^{s}\mathbb{E}\left[\frac{Q_t}{V_t}\right]
$$

$$
\leq \frac{2RD(T)}{\delta^\beta}\tau_{\text{mix}}^{1/4}T^{1/2-\beta} + \frac{1}{2s}\sum_{t=1}^{s}\mathbb{E}\left[\frac{Q_t}{V_t}\right].
$$

where the second inequality follows from the AM-GM inequality and the last equality follows from $s \leq T$. Thus, we have

$$
\mathbb{E}\left[\frac{Q_{s+1}}{V_{s+1}}\right]
$$

$$
\leq 2\sqrt{2}R\mathbb{E}[S_s]^{1/2-\beta} + \sqrt{2}Rs^{1/4}\mathbb{E}[S_s]^{1/4-\beta/2} + \frac{2RD(T)}{\delta^\beta}\tau_{\text{mix}}^{1/4}T^{1/2-\beta} + \frac{1}{2s}\sum_{t=1}^{s}\mathbb{E}\left[\frac{Q_t}{V_t}\right] \quad (8)
$$

$$
\leq 2\sqrt{2}R\mathbb{E}[S_T]^{1/2-\beta} + \sqrt{2}Rs^{1/4}\mathbb{E}[S_T]^{1/4-\beta/2} + \frac{2RD(T)}{\delta^\beta}\tau_{\text{mix}}^{1/4}T^{1/2-\beta} + \frac{1}{2s}\sum_{t=1}^{s}\mathbb{E}\left[\frac{Q_t}{V_t}\right].
$$

Now we can bound $\mathbb{E}\left[Q_t/V_t\right]$ as follow.

Lemma 17 *For $t \in [T+1]$,*

$$
\mathbb{E}\left[\frac{Q_t}{V_t}\right] \leq 4\sqrt{2}R\mathbb{E}[S_T]^{1/2-\beta} + 2\sqrt{2}RT^{1/4}\mathbb{E}[S_T]^{1/4-\beta/2} + \frac{4RD(T)}{\delta^\beta}\tau_{mix}^{1/4}T^{1/2-\beta}.
$$

Proof For $t = 1$, $Q_1 = 0$ and the inequality of the lemma holds. Assume that it holds for $t = 1, \ldots, s$ with $s \leq T$. Substituting the inequalities for $t = 1, \ldots, s$ into (8), we derive the inequality for $t = s + 1$, as required. ∎

Next, we deduce an upper bound on $\mathbb{E}\left[Q_t\right]$. First, we observe that (6) and Jensen's inequality imply the following. For $s \in [T]$,

$$
\mathbb{E}\left[Q_{s+1}\right] \leq 2\sqrt{2}\mathbb{E}[S_s]^{1/2} + \sqrt{2}s^{1/4}\mathbb{E}[S_s]^{\beta/2+1/4} + \sqrt{2C(T)\frac{\sqrt{\tau_{\text{mix}}}}{T}\sum_{t=1}^{s}\mathbb{E}[Q_t]}
$$

$$
\leq 2\sqrt{2}\mathbb{E}[S_s]^{1/2} + \sqrt{2}s^{1/4}\mathbb{E}[S_s]^{\beta/2+1/4} + sC(T)\frac{\sqrt{\tau_{\text{mix}}}}{T} + \frac{1}{2s}\sum_{t=1}^{s}\mathbb{E}[Q_t] \quad (9)
$$

$$
\leq 2\sqrt{2}\mathbb{E}[S_s]^{1/2} + \sqrt{2}s^{1/4}\mathbb{E}[S_s]^{\beta/2+1/4} + C(T)\tau_{\text{mix}}^{1/2} + \frac{1}{2s}\sum_{t=1}^{s}\mathbb{E}[Q_t],
$$

where the first inequality is implied by combining (6) and (7) and applying Lemma 16. Next we provide an adaptive upper bound on $\mathbb{E}\left[Q_t\right]$

Lemma 18 *For any $t \in [T+1]$,*

$$\mathbb{E}[Q_t] \leq 4\sqrt{2}\mathbb{E}[S_{t-1}]^{1/2} + \frac{5\sqrt{2}}{3}(t-1)^{1/4}\mathbb{E}[S_{t-1}]^{\beta/2+1/4} + 2C(T)\tau_{mix}^{1/2}.$$

Proof We argue by induction. The inequality trivially holds when $t = 1$ as $Q_1 = 0$. Suppose that the inequality holds for $t = 1, 2, \ldots, s$. Then

$$\mathbb{E}[Q_{s+1}] \leq 2\sqrt{2}\mathbb{E}[S_s]^{1/2} + \sqrt{2}s^{1/4}\mathbb{E}[S_s]^{\beta/2+1/4} + C(T)\tau_{\text{mix}}^{1/2}$$
$$+ \frac{1}{2s}\sum_{t=1}^{s}\left(4\sqrt{2}\mathbb{E}[S_{t-1}]^{1/2} + \frac{10}{3}(t-1)^{1/4}\mathbb{E}[S_{t-1}]^{\beta/2+1/4} + 2C(T)\tau_{\text{mix}}^{1/2}\right)$$
$$\leq 2\sqrt{2}\mathbb{E}[S_s]^{1/2} + \sqrt{2}s^{1/4}\mathbb{E}[S_s]^{\beta/2+1/4} + C(T)\tau_{\text{mix}}^{1/2}$$
$$+ \frac{1}{2s}\left(4\sqrt{2}s\mathbb{E}[S_s]^{1/2} + \frac{4\sqrt{2}}{3}s^{5/4}\mathbb{E}[S_s]^{\beta/2+1/4} + 2sC(T)\tau_{\text{mix}}^{1/2}\right)$$
$$= 4\sqrt{2}\mathbb{E}[S_s]^{1/2} + \frac{5\sqrt{2}}{3}s^{1/4}\mathbb{E}[S_s]^{\beta/2+1/4} + 2C(T)\tau_{\text{mix}}^{1/2},$$

where the first inequality is from (9) and the second inequality is by Corollary 29. ∎

7.2 MLMC Adaptive Drift-Plus-Penalty for Stochastic-Constrained Stochastic Optimization

Based on the bounds on $\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]$, $\mathbb{E}[Q_t/V_t]$, and $\mathbb{E}[Q_t]$, in this section, we prove Lemma 6 and Theorem 8. To better present the proof of Lemma 6, we divide the analysis into two parts, one for the regret and the other for the constraint violation. Then, using Lemma 6, we prove Theorem 8.

Proof [Proof of the first part of lemma 6] We first bound

$$\mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x})\right].$$

By Lemma 21, this is less than or equal to

$$\mathbb{E}\left[\frac{\alpha_T}{V_T}R^2\right] + \mathbb{E}\left[\sum_{t=1}^{T}\frac{V_t F_t^2}{4\alpha_t}\right] + \mathbb{E}\left[\frac{1}{2}\sum_{t=1}^{T}\frac{(H_t + G_t R)^2}{V_t}\right] + \mathbb{E}\left[\sum_{t=1}^{T}\frac{Q_t}{V_t}g_t(\boldsymbol{x})\right].$$

Here, the first term is equal to $RS_{T-1}^{1-\beta}$ which is less or equal to $RS_T^{1-\beta}$, and the second term satisfies

$$\mathbb{E}\left[\sum_{t=1}^{T}\frac{V_t F_t^2}{4\alpha_t}\right] = \frac{R}{4}\mathbb{E}\left[\sum_{t=1}^{T}S_{t-1}^{\beta-1}F_t^2\right] \leq R\left[\sum_{t=1}^{T}S_{t-1}^{\beta-1}a_t\right] = O\left(\mathbb{E}\left[S_T^\beta\right]\right) = O\left(\mathbb{E}\left[S_T^\beta\right]\right)$$

where the inequality holds by Corollary 31. The third term satisfies

$$\mathbb{E}\left[\sum_{t=1}^{T}\frac{(H_t + G_t R)^2}{2V_t}\right] \leq \mathbb{E}\left[R\sum_{t=1}^{T}\frac{R^2 G_t^2 + H_t^2}{S_{t-1}^\beta}\right] \leq \mathbb{E}\left[R\sum_{t=1}^{T}\frac{a_t}{S_{t-1}^\beta}\right] = O\left(\mathbb{E}\left[S_T^{1-\beta}\right]\right).$$

42

By Jensen's inequality, we have $\mathbb{E}\left[S_T^{1-\beta}\right] \leq \mathbb{E}\left[S_T\right]^{1-\beta}$ and $\mathbb{E}\left[S_T^{\beta}\right] \leq \mathbb{E}\left[S_T\right]^{\beta}$. The fourth term can be bounded as follows. Note that

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{Q_t}{V_t} g_t(\boldsymbol{x})\right] = \sum_{t=1}^{T} \mathbb{E}\left[\frac{Q_t}{V_t} \mathbb{E}_{t-1}[g_t(\boldsymbol{x})]\right].$$

We have two upper bounds on $\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]$, one given in (4) and the other one due to Lemma 16. Let $m(T)$ be the minimum of the two upper bounds, i.e.,

$$m(T) = \min\left\{C(T)\tau_{\text{mix}}^{1/2}T^{-1}, \ D(T)\tau_{\text{mix}}^{1/4}T^{-1/2}\right\}.$$

Then it follows that

$$\sum_{t=1}^{T} \mathbb{E}\left[\frac{Q_t}{V_t} \mathbb{E}_{t-1}[g_t(\boldsymbol{x})]\right]$$

$$\leq \sum_{t=1}^{T} m(T)\left(4\sqrt{2}R\mathbb{E}[S_T]^{1/2-\beta} + 2\sqrt{2}RT^{1/4}\mathbb{E}[S_T]^{1/4-\beta/2} + \frac{4RD(T)}{\delta^\beta}\tau_{\text{mix}}^{1/4}T^{1/2-\beta}\right)$$

$$\leq \sum_{t=1}^{T}\left(C(T)\tau_{\text{mix}}^{1/2}T^{-1}\left(4\sqrt{2}R\mathbb{E}[S_T]^{1/2-\beta} + 2\sqrt{2}RT^{1/4}\mathbb{E}[S_T]^{1/4-\beta/2}\right) + \frac{4RD(T)^2}{\delta^\beta}\tau_{\text{mix}}^{1/2}T^{-\beta}\right)$$

$$= \tilde{O}\left(\tau_{\text{mix}}^{1/2}\mathbb{E}[S_T]^{1/2-\beta} + \tau_{\text{mix}}^{1/2}T^{1/4}\mathbb{E}[S_T]^{1/4-\beta/2} + \tau_{\text{mix}}^{1/2}T^{1-\beta}\right)$$

where the first inequality follows from (4) and Lemmas 16 and 17. Combining the bounds on the four terms, we have proved the desired bound on $\mathbb{E}\left[\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x})\right]$. ∎

Proof [Proof of the second part of Lemma 6] By Lemma 22, we have

$$\sum_{t=1}^{T} g_t(\boldsymbol{x}_t) \leq Q_{T+1} + \sum_{t=1}^{T} \frac{G_t}{2\alpha_t}(V_t F_t + Q_t G_t).$$

Moreover, by Lemma 18, we have

$$\mathbb{E}[Q_{T+1}] \leq \tilde{O}\left(\mathbb{E}[S_T]^{1/2} + T^{1/4}\mathbb{E}[S_T]^{\beta/2+1/4} + \tau_{\text{mix}}^{1/2}\right).$$

Next it follows from Corollary 31 that

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{V_t F_t G_t}{2\alpha_t}\right] \leq \mathbb{E}\left[\sum_{t=1}^{T} \frac{S_{t-1}^{\beta-1} a_t}{2}\right] = O\left(\mathbb{E}\left[S_T^{\beta}\right]\right) = O\left(\mathbb{E}[S_T]^{\beta}\right).$$

Furthermore,

$$\mathbb{E}\left[\sum_{t=1}^{T} \frac{G_t^2}{2\alpha_t}Q_t\right] \leq \mathbb{E}\left[\max_{t\in[T]} Q_t \cdot \sum_{t=1}^{T} \frac{R^2 G_t^2}{2S_{t-1}}\right] \leq \mathbb{E}\left[\left(\frac{I+\delta}{2\delta} + \frac{1}{2}\log\frac{S_T - I - \delta}{\delta}\right)\max_{t\in[T]} Q_t\right]$$

where the second inequality is implied by Lemma 30. Here, the last term can be upper bounded using the AM-GM inequality as follows.

$$\left(\frac{I+\delta}{2\delta} + \frac{1}{2}\log\frac{S_T - I - \delta}{\delta}\right)\max_{t\in[T]} Q_t$$

$$\leq \frac{\max_{t\in[T]} Q_t^2}{2(I\tau_{\text{mix}})^{\beta/2+1/4}T^{\beta/2+1/2}} + \frac{1}{2}\left(\frac{I+\delta}{2\delta} + \frac{1}{2}\log\frac{S_T}{\delta}\right)^2 (I\tau_{\text{mix}})^{\beta/2+1/4}T^{\beta/2+1/2}.$$

Let $s = \text{argmax}_{t\in[T]} Q_t$. Then by (5) and (7),

$$\mathbb{E}\left[Q_s^2\right] \leq 4\mathbb{E}[S_{s-1}] + 2\sqrt{s}\mathbb{E}[S_{s-1}]^{\beta+1/2} + 2\sum_{t=1}^{s-1}\mathbb{E}\left[Q_t\mathbb{E}_{t-1}\left[g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right]\right]$$

$$\leq 4\mathbb{E}[S_T] + 2\sqrt{T}\mathbb{E}[S_T]^{\beta+1/2} + 2\sum_{t=1}^{s-1}\mathbb{E}[Q_t\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]].$$

Moreover,

$$\sum_{t=1}^{s-1}\mathbb{E}[Q_t\mathbb{E}_{t-1}[g_t(\boldsymbol{x})]]$$

$$\leq C(T)\tau_{\text{mix}}^{1/2}T^{-1}\sum_{t=1}^{s-1}\mathbb{E}[Q_t]$$

$$\leq C(T)\tau_{\text{mix}}^{1/2}T^{-1}\sum_{t=1}^{T}\left(4\sqrt{2}\mathbb{E}[S_T]^{1/2} + \frac{5\sqrt{2}}{3}(t-1)^{1/4}\mathbb{E}[S_T]^{\beta/2+1/4} + 2C(T)\tau_{\text{mix}}^{1/2}\right)$$

$$\leq C(T)\tau_{\text{mix}}^{1/2}T^{-1}\left(4\sqrt{2}T\mathbb{E}[S_T]^{1/2} + \frac{4\sqrt{2}}{3}T^{5/4}\mathbb{E}[S_T]^{\beta/2+1/4} + 2C(T)\tau_{\text{mix}}^{1/2}T\right)$$

$$= \tilde{O}\left(\tau_{\text{mix}}^{1/2}\mathbb{E}[S_T]^{1/2} + \tau_{\text{mix}}^{1/2}T^{1/4}\mathbb{E}[S_T]^{\beta/2+1/4} + \tau_{\text{mix}}\right),$$

where the first inequality follows from (4), the second inequality follows from Lemma 18, and the third inequality follows from Corollary 29. Thus, we get

$$\mathbb{E}\left[\sum_{t=1}^{T}g_t(\boldsymbol{x}_t)\right] = \tilde{O}\left(\mathbb{E}[S_T]^{1/2} + T^{1/4}\mathbb{E}[S_T]^{\beta/2+1/4} + \tau_{\text{mix}}^{1/2} + \frac{\mathbb{E}[S_T] + T^{1/2}\mathbb{E}[S_T]^{\beta+1/2}}{\tau_{\text{mix}}^{\beta/2+1/4}T^{\beta/2+1/2}}\right.$$

$$\left. + \frac{\tau_{\text{mix}}^{1/2} + \mathbb{E}[S_T]^{1/2} + \mathbb{E}[S_T]^{\beta/2+1/4}}{\tau_{\text{mix}}^{\beta/2+1/4}T^{\beta/2+1/2}} + (\log S_T)^2\tau_{\text{mix}}^{\beta/2-1/4}T^{\beta/2+1/2}\right),$$

as required. ∎

Proof [Proof of Theorem 8] As $\bar{f}$ is convex,

$$\bar{f}(\bar{\boldsymbol{x}}_T) - \bar{f}(\boldsymbol{x}^{\#}) \leq \frac{1}{T}\sum_{t=1}^{T}\left(\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}^{\#})\right).$$

44

We can decompose the right-hand side as follows

$$\frac{\bar{f}(\boldsymbol{x}_t) - \bar{f}(\boldsymbol{x}^{\#})}{T} = \frac{\bar{f}(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t)}{T} + \frac{f_t(\boldsymbol{x}_t) - f_t(\boldsymbol{x}^{\#})}{T} + \frac{f_t(\boldsymbol{x}^{\#}) - \bar{f}(\boldsymbol{x}^{\#})}{T}.$$

Here, Jensen's inequality and Lemma 5 imply that

$$\frac{1}{T}\mathbb{E}\left[|\bar{f}(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t)|\right] \leq \frac{1}{T}\sqrt{\mathbb{E}\left[|\bar{f}(\boldsymbol{x}_t) - f_t(\boldsymbol{x}_t)|^2\right]} = \tilde{O}(\tau_{\text{mix}}^{1/2}T^{-2}),$$

$$\frac{1}{T}\mathbb{E}\left[|\bar{f}(\boldsymbol{x}^{\#}) - f_t(\boldsymbol{x}^{\#})|\right] \leq \frac{1}{T}\sqrt{\mathbb{E}\left[|\bar{f}(\boldsymbol{x}^{\#}) - f_t(\boldsymbol{x}^{\#})|^2\right]} = \tilde{O}(\tau_{\text{mix}}^{1/2}T^{-2}).$$

Moreover, we have derived Proposition 7 which provides an upper bound on the expectation of $\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x}^{\#})$. Consequently,

$$\mathbb{E}\left[\bar{f}(\bar{\boldsymbol{x}}_T) - \bar{f}(\boldsymbol{x}^{\#})\right] \leq \tilde{O}\left((\tau_{\text{mix}})^{1-\beta}T^{-\beta}\right).$$

For the second part, Jensen's inequality and Lemma 5 imply that

$$\frac{1}{T}\mathbb{E}\left[|\bar{g}(\boldsymbol{x}_t) - g_t(\boldsymbol{x}_t)|\right] \leq \frac{1}{T}\sqrt{\mathbb{E}\left[|\bar{g}(\boldsymbol{x}_t) - g_t(\boldsymbol{x}_t)|^2\right]} = \tilde{O}(\tau_{\text{mix}}^{1/2}T^{-2})$$

As a result,

$$\begin{aligned}
\mathbb{E}[\bar{g}(\bar{\boldsymbol{x}}_T)] &\leq \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\bar{g}(\boldsymbol{x}_t)] \\
&= \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[\bar{g}(\boldsymbol{x}_t) - g_t(\boldsymbol{x}_t)] + \frac{1}{T}\sum_{t=1}^{T}\mathbb{E}[g_t(\boldsymbol{x}_t)] \\
&= \tilde{O}\left(\tau_{\text{mix}}^{1/2}T^{-2} + \tau_{\text{mix}}^{\beta/2+1/4}T^{\beta/2-1/2} + \tau_{\text{mix}}^{3/4-\beta/2}T^{-\beta/2-1/2}\right) \\
&= \tilde{O}\left(\tau_{\text{mix}}^{\beta/2+1/4}T^{\beta/2-1/2} + \tau_{\text{mix}}^{3/4-\beta/2}T^{-\beta/2-1/2}\right),
\end{aligned}$$

where the first inequality holds because $\bar{g}$ is convex and the second equality follows from Proposition 7. ∎

Acknowledgments and Disclosure of Funding

Algorithm 3 Adaptive Drift-Plus-Penalty

    Initialize: Initial iterates $\boldsymbol{x_1} \in \mathcal{X}$, $Q_1 = 0$.

    for $t = 1$ **to** T **do**

        Observe f_t and g_t.

        Set penalty parameter V_t and step size parameter α_t such that $0 \leq V_{t-1} \leq V_t$, $0 \leq \alpha_{t-1} \leq \alpha_t$, and $0 \leq \alpha_{t-1}/V_{t-1} \leq \alpha_t/V_t$.

        Primal update: Set $\boldsymbol{x}_{t+1}$ as

$$\boldsymbol{x}_{t+1} = \operatorname*{argmin}_{\boldsymbol{x} \in \mathcal{X}} \left\{ (V_t \nabla f_t(\boldsymbol{x}_t) + Q_t \nabla g_t(\boldsymbol{x}_t))^\top \boldsymbol{x} + \alpha_t D(\boldsymbol{x}, \boldsymbol{x}_t) \right\}$$

        Dual update: Set Q_{t+1} as

$$Q_{t+1} = \left[Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right]_+$$

    end for

Appendix A. Analysis of Adaptive Drift-Plus-Penalty

Algorithm 3 is a general template for adaptive variants of the drift-plus-penalty algorithm whose parameters V_t and α_t satisfy that $\{V_t\}_{t=1}^T$, $\{\alpha_t\}_{t=1}^T$, and $\{\alpha_t/V_t\}_{t=1}^T$ are non-decreasing sequences of non-negative numbers. In this section, we analyze the general template of DPP given by Algorithm 3, based on which we deduce performance guarantees on Algorithms 1 and 2.

Recall that $\Delta_t = (Q_{t+1}^2 - Q_t^2)/2$ is the Lyapunov drift term. The following lemma provides a bound on the drift term.

Lemma 19 *For $t \geq 1$,*

$$\Delta_t \leq Q_t \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right) + \frac{1}{2}(H_t + G_t R)^2.$$

Proof As $Q_{t+1} = \max \left\{ Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t), 0 \right\}$, we have that $Q_{t+1}^2 \leq \left(Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right)^2$. Hence, it follows that

$$
\begin{aligned}
\Delta_t &= \frac{Q_{t+1}^2}{2} - \frac{Q_t^2}{2} \\
&\leq Q_t \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right) + \frac{1}{2} \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right)^2 \\
&\leq Q_t \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \right) + \frac{1}{2}(H_t + G_t R)^2
\end{aligned}
$$

where the last inequality is from Assumption 2. ∎

Since our drift-plus-penalty algorithm is a mirror descent version, we need the following lemma, which is obtained by substituting $\boldsymbol{y} = \boldsymbol{x}_t$, $\boldsymbol{x}^* = \boldsymbol{x}_{t+1}$, and $f(\boldsymbol{x}) = (V_t \nabla f_t(\boldsymbol{x}_t) + Q_t \nabla g_t(\boldsymbol{x}_t))^\top \boldsymbol{x}$ into (Wei et al., 2020, Lemma 2.1).

Lemma 20 *(Wei et al., 2020, Equation (22)) For any $\boldsymbol{x} \in \mathcal{X}$ and $t \geq 1$,*

$$(V_t \nabla f_t(\boldsymbol{x}_t) + Q_t \nabla g_t(\boldsymbol{x}_t))^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) + \alpha_t D(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t)$$
$$\leq (V_t \nabla f_t(\boldsymbol{x}_t) + Q_t \nabla g_t(\boldsymbol{x}_t))^\top (\boldsymbol{x} - \boldsymbol{x}_t) + \alpha_t D(\boldsymbol{x}, \boldsymbol{x}_t) - \alpha_t D(\boldsymbol{x}, \boldsymbol{x}_{t+1}).$$

Recall that f_t and g_t for the known mixing time case in Section 3 correspond to one sample and are assumed to be convex. In contrast, f_t and g_t for the unknown mixing time setting in Section 4 come from the MLMC estimation scheme with multiple data samples and thus are not necessarily convex. Nevertheless, we use the fact that $\mathbb{E}_{t-1}[f_t]$ and $\mathbb{E}_{t-1}[g_t]$ are convex. Based on this, we deduce the following lemma.

Lemma 21 *Suppose that $\mathbb{E}_{t-1}[f_t] = \mathbb{E}_{t-1}[\hat{f}_t]$ and $\mathbb{E}_{t-1}[g_t] = \mathbb{E}_{t-1}[\hat{g}_t]$ where $\hat{f}_t$ and $\hat{g}_t$ are convex functions and that $\mathbb{E}_{t-1}[\nabla f_t] = \mathbb{E}_{t-1}[\nabla \hat{f}_t]$ and $\mathbb{E}_{t-1}[\nabla g_t] = \mathbb{E}_{t-1}[\nabla \hat{g}_t]$. Then for any $\boldsymbol{x} \in \mathcal{X}$,*

$$\mathbb{E}\left[\sum_{t=1}^T f_t(\boldsymbol{x}_t) - \sum_{t=1}^T f_t(\boldsymbol{x})\right] \leq \mathbb{E}\left[\sum_{t=1}^T \frac{Q_t}{V_t} g_t(\boldsymbol{x})\right] + \mathbb{E}\left[\frac{\alpha_T}{V_T} R^2\right]$$
$$+ \mathbb{E}\left[\sum_{t=1}^T \frac{V_t F_t^2}{4\alpha_t}\right] + \frac{1}{2}\mathbb{E}\left[\sum_{t=1}^T \frac{(H_t + G_t R)^2}{V_t}\right].$$

Proof Dividing both sides of the inequality given in Theorem 20 by V_t and addding $f_t(\boldsymbol{x}_t) + Q_t g_t(\boldsymbol{x}_t)/V_t$ to both sides, we get

$$f_t(\boldsymbol{x}_t) + \frac{Q_t}{V_t} g_t(\boldsymbol{x}_t) + \left(\nabla f_t(\boldsymbol{x}_t) + \frac{Q_t}{V_t} \nabla g_t(\boldsymbol{x}_t)\right)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) + \frac{\alpha_t}{V_t} D(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t)$$
$$\leq f_t(\boldsymbol{x}_t) + \frac{Q_t}{V_t} g_t(\boldsymbol{x}_t) + \left(\nabla f_t(\boldsymbol{x}_t) + \frac{Q_t}{V_t} \nabla g_t(\boldsymbol{x}_t)\right)^\top (\boldsymbol{x} - \boldsymbol{x}_t) + \frac{\alpha_t}{V_t} D(\boldsymbol{x}, \boldsymbol{x}_t) - \frac{\alpha_t}{V_t} D(\boldsymbol{x}, \boldsymbol{x}_{t+1}).$$

Here, the left-hand side is bounded below by

$$f_t(\boldsymbol{x}_t) + \nabla f_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) + \frac{\alpha_t}{V_t} D(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t) + \frac{\Delta_t}{V_t} - \frac{1}{2V_t}(H_t + G_t R)^2$$

by Lemma 19. Moreover, we have

$$\nabla f_t(\boldsymbol{x}_t)^\top (\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) + \frac{\alpha_t}{V_t} D(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t) \geq -F_t \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\| + \frac{\alpha_t}{V_t} \|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|^2$$
$$= -\frac{V_t F_t^2}{4\alpha_t} + \frac{\alpha_t}{V_t}\left(\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\| - \frac{V_t F_t}{2\alpha_t}\right)^2$$
$$\geq -\frac{V_t F_t^2}{4\alpha_t}$$

where the first inequality holds by 2-strong convexity of Φ with respect to $\|\cdot\|$. Hence, we deduce that

$$f_t(\boldsymbol{x}_t) - \frac{V_t F_t^2}{4\alpha_t} + \frac{\Delta_t}{V_t} - \frac{1}{2V_t}(H_t + G_t R)^2$$
$$\leq f_t(\boldsymbol{x}_t) + \frac{Q_t}{V_t} g_t(\boldsymbol{x}_t) + \left(\nabla f_t(\boldsymbol{x}_t) + \frac{Q_t}{V_t} \nabla g_t(\boldsymbol{x}_t)\right)^\top (\boldsymbol{x} - \boldsymbol{x}_t) + \frac{\alpha_t}{V_t} D(\boldsymbol{x}, \boldsymbol{x}_t) - \frac{\alpha_t}{V_t} D(\boldsymbol{x}, \boldsymbol{x}_{t+1}).$$

For the right-hand side of this inequality, consider

$$\sum_{t=1}^{T} \frac{\alpha_t}{V_t} \left(D(\boldsymbol{x}, \boldsymbol{x}_t) - D(\boldsymbol{x}, \boldsymbol{x}_{t+1}) \right) = \frac{\alpha_1}{V_1} D(\boldsymbol{x}, \boldsymbol{x}_t) + \sum_{t=2}^{T} D(\boldsymbol{x}, \boldsymbol{x}_t) \left(\frac{\alpha_t}{V_t} - \frac{\alpha_{t-1}}{V_{t-1}} \right) - \frac{\alpha_T}{V_T} D(\boldsymbol{x}, \boldsymbol{x}_{t+1})$$

$$\leq \frac{\alpha_1}{V_1} R^2 + \sum_{t=2}^{T} R^2 \left(\frac{\alpha_t}{V_t} - \frac{\alpha_{t-1}}{V_{t-1}} \right)$$

$$= \frac{\alpha_T}{V_T} R^2$$

where the inequality holds since $\{\alpha_t/V_t\}_{t=1}^{T}$ is a non-decreasing sequence. Furthermore,

$$\sum_{t=1}^{T} \frac{\Delta_t}{V_t} = \frac{1}{2} \sum_{t=1}^{T} \frac{1}{V_t} (Q_{t+1}^2 - Q_t^2) = -\frac{Q_1^2}{2V_1} + \frac{Q_{T+1}^2}{2V_T} + \frac{1}{2} \sum_{t=2}^{T} Q_t^2 \left(\frac{1}{V_{t-1}} - \frac{1}{V_t} \right) \geq -\frac{Q_1^2}{2V_1} = 0,$$

where the inequality holds since the sequence $\{V_t\}_{t=1}^{T}$ is non-negative and non-decreasing. Combining these inequalities, we deduce that

$$\sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} \left(f_t(\boldsymbol{x}_t) + \nabla f_t(\boldsymbol{x}_t)^{\top} (\boldsymbol{x} - \boldsymbol{x}_t) \right)$$

$$\leq \sum_{t=1}^{T} \frac{Q_t}{V_t} \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^{\top} (\boldsymbol{x} - \boldsymbol{x}_t) \right) + \frac{\alpha_T}{V_T} R^2 + \sum_{t=1}^{T} \frac{V_t F_t^2}{4\alpha_t} + \frac{1}{2} \sum_{t=1}^{T} \frac{(H_t + G_t R)^2}{V_t}. \tag{10}$$

Recall that $\mathbb{E}_{t-1}[f_t] = \mathbb{E}_{t-1}[\hat{f}_t]$ and $\mathbb{E}_{t-1}[g_t] = \mathbb{E}_{t-1}[\hat{g}_t]$ where $\hat{f}_t$ and $\hat{g}_t$ are convex and that $\mathbb{E}_{t-1}[\nabla f_t] = \mathbb{E}_{t-1}[\nabla \hat{f}_t]$ and $\mathbb{E}_{t-1}[\nabla g_t] = \mathbb{E}_{t-1}[\nabla \hat{g}_t]$. Then it follows that

$$\mathbb{E}_{t-1}[f_t(\boldsymbol{x}_t)] + \mathbb{E}_{t-1}[\nabla f_t(\boldsymbol{x}_t)]^{\top} (\boldsymbol{x} - \boldsymbol{x}_t) = \mathbb{E}_{t-1}[\hat{f}_t(\boldsymbol{x}_t)] + \mathbb{E}_{t-1}[\nabla \hat{f}_t(\boldsymbol{x}_t)]^{\top} (\boldsymbol{x} - \boldsymbol{x}_t)$$

$$\leq \mathbb{E}_{t-1}[\hat{f}_t(\boldsymbol{x})]$$

$$= \mathbb{E}_{t-1}[f_t(\boldsymbol{x})]$$

where the second inequality holds because $\hat{f}_t$ is convex and $\boldsymbol{x}_t$ is $\mathcal{F}_{t-1}$-measurable. Likewise, we deduce that

$$\mathbb{E}_{t-1}[g_t(\boldsymbol{x}_t)] + \mathbb{E}_{t-1}[\nabla g_t(\boldsymbol{x}_t)]^{\top} (\boldsymbol{x} - \boldsymbol{x}_t) \leq \mathbb{E}_{t-1}[g_t(\boldsymbol{x})].$$

Taking the expectations of both sides of (10), we obtain the inequality of this lemma, as required. ∎

Next, we state a lemma that will be useful to provide an upper bound on the constraint violation.

Lemma 22 *Algorithm 3 achieves*

$$\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\| \leq \frac{1}{2\alpha_t} (V_t F_t + Q_t G_t), \quad \sum_{t=1}^{T} g_t(\boldsymbol{x}_t) \leq Q_{T+1} + \sum_{t=1}^{T} \frac{G_t}{2\alpha_t} (V_t F_t + Q_t G_t).$$

Proof As $Q_{t+1} \geq Q_t + g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)$, it follows that

$$g_t(\boldsymbol{x}_t) \leq Q_{t+1} - Q_t - \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \leq Q_{t+1} - Q_t + G_t\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|.$$

On the other hand, if we set $\boldsymbol{x} = \boldsymbol{x}_t$ for the inequaity of Lemma 20, we get

$$\alpha_t D(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t) + \alpha_t D(\boldsymbol{x}_t, \boldsymbol{x}_{t+1}) \leq \left(V_t\nabla f_t(\boldsymbol{x}_t) + Q_t\nabla g_t(\boldsymbol{x}_t)\right)^\top(\boldsymbol{x}_t - \boldsymbol{x}_{t+1}).$$

Here, the left-hand side is greater or equal to $2\alpha_t\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|^2$ while the right-hand side is less or equal to $(V_tF_t + Q_tG_t)\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\|$. Therefore, it follows that

$$\|\boldsymbol{x}_{t+1} - \boldsymbol{x}_t\| \leq \frac{1}{2\alpha_t}(V_tF_t + Q_tG_t), \tag{11}$$

which implies

$$\sum_{t=1}^T g_t(\boldsymbol{x}_t) \leq Q_{T+1} + \sum_{t=1}^T \frac{G_t}{2\alpha_t}(V_tF_t + Q_tG_t),$$

as required. ∎

To bound the constraint violation, we still need to bound the virtual queue size Q_{T+1}. We also need the following lemma.

Lemma 23 *For any $\boldsymbol{x} \in \mathcal{X}$,*

$$\Delta_t \leq H_t^2 + R^2G_t^2 + Q_t\left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right) + V_tRF_t + \alpha_t(D(\boldsymbol{x}, \boldsymbol{x}_t) - D(\boldsymbol{x}, \boldsymbol{x}_{t+1})).$$

Proof By Lemma 19,

$$\begin{aligned}
\Delta_t &\leq Q_t\left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)\right) + \frac{1}{2}(H_t + RG_t)^2 \\
&\leq Q_t\left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)\right) + H_t^2 + R^2G_t^2
\end{aligned}$$

where the last inequality comes from the fact that $(A + B)^2 \leq 2(A^2 + B^2)$. Here, using Lemma 20, the right-hand side can be further bounded above as follows.

$$\begin{aligned}
&Q_t(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t)) + H_t^2 + R^2G_t^2 \\
&\leq Q_t\left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t)\right) + V_t\nabla f_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t) - V_t\nabla f_t(\boldsymbol{x}_t)^\top(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) \\
&\quad - \alpha_t D(\boldsymbol{x}_{t+1}, \boldsymbol{x}_t) + \alpha_t\left(D(\boldsymbol{x}, \boldsymbol{x}_t) - D(\boldsymbol{x}, \boldsymbol{x}_{t+1})\right) + H_t^2 + R^2G_t^2.
\end{aligned}$$

Moreover, it follows from the Cauchy-Schwarz inequality that

$$V_t\nabla f_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_t) - V_t\nabla f_t(\boldsymbol{x}_t)^\top(\boldsymbol{x}_{t+1} - \boldsymbol{x}_t) = V_t\nabla f_t(\boldsymbol{x}_t)^\top(\boldsymbol{x} - \boldsymbol{x}_{t+1}) \leq V_tRF_t$$

by Cauchy-Schwarz inequality. Then we have proved the lemma, as desired. ∎

Based on Lemma 23, we may provide the following bound on the virtual queue size.

Lemma 24 *For any $\boldsymbol{x} \in \mathcal{X}, s \in [T]$,*

$$Q_{s+1} \leq \sqrt{2 \sum_{t=1}^{s} \left(H_t^2 + R^2 G_t^2 + V_t R F_t\right) + 2 \sum_{t=1}^{s} Q_t \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x} - \boldsymbol{x}_t)\right) + 2R^2 \alpha_s}.$$

Proof From Lemma 23, we get

$$\frac{Q_{s+1}^2}{2} = \sum_{t=1}^{s} \Delta_t$$

$$\leq \sum_{t=1}^{s} \left(H_t^2 + R^2 G_t^2 + Q_t \left(g_t(\boldsymbol{x}_t) + \nabla g_t(\boldsymbol{x}_t)^\top (\boldsymbol{x} - \boldsymbol{x}_t)\right) + V_t R F_t\right)$$

$$+ \alpha_1 D(\boldsymbol{x}, \boldsymbol{x_1}) + \sum_{t=2}^{s} D(\boldsymbol{x}, \boldsymbol{x}_t)(\alpha_t - \alpha_{t-1}) - \alpha_s D(\boldsymbol{x}, \boldsymbol{x}_{t+1}).$$

Since α_t is non-decreasing and $D(\boldsymbol{x}, \boldsymbol{x}_t) \leq R^2$, it follows that

$$\alpha_1 D(\boldsymbol{x}, \boldsymbol{x_1}) + \sum_{t=2}^{s} D(\boldsymbol{x}, \boldsymbol{x}_t)(\alpha_t - \alpha_{t-1}) - \alpha_s D(\boldsymbol{x}, \boldsymbol{x}_{t+1}) \leq R^2 \alpha_s.$$

This implies the desired bound on Q_{s+1}. ∎

Appendix B. Sum of Sequences

In this section, we consider some series of numbers and provide bounds on their partial sums to make our paper self-contained. Given a sequence $\{x_t\}_{t=1}^{\infty}$ of numbers, we use notation $X_s := \sum_{t=1}^{s} x_t$ to denote its partial sums.

Lemma 25 *If $f : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and non-increasing, then*

$$\sum_{t=1}^{T} f(X_t) x_t \leq x_1 f(X_1) + \int_{X_1}^{X_T} f(x) dx \leq \int_0^{X_T} f(x) dx$$

for any nonnegative $x_1, \ldots, x_T$.

Proof By considering the area between $f(x)$ and the x-axis over the interval $[X_{t-1}, X_t]$ of length x_t, we have $f(X_t) x_t \leq \int_{X_{t-1}}^{X_t} f(x) dx$ since f is non-increasing. Then it follows that

$$\sum_{t=1}^{T} f(X_t) x_t = x_1 f(X_1) + \sum_{t=2}^{T} f(X_t) x_t \leq x_1 f(X_1) + \int_{X_1}^{X_T} f(x) dx \leq \int_0^{X_T} f(x) dx,$$

as required. ∎

As a consequence of Lemma 25, we deduce the following list of bounds on series.

Corollary 26 *(Auer et al., 2002)*

$$\sum_{t=1}^{T} \frac{x_t}{\sqrt{X_t}} \leq 2\sqrt{X_T}.$$

Corollary 27

$$\sum_{t=1}^{T} \frac{x_t}{X_t} \leq 1 + \log_{X_1}(X_T).$$

Next we consider the following.

Lemma 28 *If $f : \mathbb{R}_+ \to \mathbb{R}_+$ is continuous and non-decreasing, then*

$$\sum_{t=1}^{T} f(X_t)x_{t+1} \leq \int_{X_1}^{X_{T+1}} f(x)dx$$

for any nonnegative $x_1, \ldots, x_T$.

Proof By considering the area between $f(x)$ and the x-axis over the interval $[X_t, X_{t+1}]$ of length x_{t+1}, we have $f(X_t)x_{t+1} \leq \int_{X_t}^{X_{t+1}} f(x)dx$ since f is non-decreasing. Then it follows that

$$\sum_{t=1}^{T} f(X_t)x_{t+1} = x_2 f(X_1) + \sum_{t=2}^{T} f(X_t)x_{t+1} \leq x_2 f(X_1) + \int_{X_2}^{X_{T+1}} f(x)dx \leq \int_{X_1}^{X_{T+1}} f(x)dx,$$

as required. ∎

Lemma 28 implies the following.

Corollary 29 *For $q > 0$,*

$$\sum_{t=1}^{T} t^q \leq \frac{1}{q+1}\big((T+1)^{q+1} - 1\big).$$

Moreover, when each x_t is bounded by some fixed constant, we can deduce the following result. For ease of notation, we start a sequence with x_0 and, with abuse of notation, define partial sum $X_s = \sum_{t=0}^{s} x_t$ with $x_0 = X_0 = \delta$.

Lemma 30 *If $0 \leq x_t \leq C$ for $t = 0, \ldots, T$ for some fixed constant C and $f : \mathbb{R}_{++} \to \mathbb{R}_+$ is continuous and non-increasing, then*

$$\sum_{t=1}^{T} f(X_{t-1})x_t \leq Cf(\delta) + \int_{\delta}^{\max\{\delta, X_T - C\}} f(x)dx.$$

Proof If we consider the area between $f(x)$ and the x-axis over the interval $[X_{t-1}, X_t]$, we have $f(X_{t-1})x_t \geq \int_{X_{t-1}}^{X_t} f(x)dx$ in which the inequality direction is the opposite of what we

51

want. However, since $x_t \le C$, we can use the idea of translation by C in the x-axis direction in the following way. Let

$$\tilde{f}(x) = \begin{cases} f(\delta), & x \in (-\infty, \delta], \\ f(x), & x \in (\delta, \infty). \end{cases}$$

Then the graph of the translation $\tilde{f}(x - C)$ is above the squares of height $f(X_{t-1})$ on the interval $[X_{t-1}, X_t]$. Thus,

$$\sum_{t=1}^{T} f(X_{t-1})x_t \le \int_{X_0}^{X_T} \tilde{f}(x - C)dx \le \begin{cases} (X_T - \delta)f(\delta) \le Cf(\delta), & \text{if } X_T \le \delta + C, \\ Cf(\delta) + \int_{\delta}^{X_T - C} f(x)dx, & \text{if } X_T > \delta + C \end{cases}$$

which implies the desired statement of this lemma. ∎

As a corollary of Lemma 30 with $f(x) = x^{-\gamma}$, we deduce the following inequality.

Corollary 31 *If $x_t \le C, 0 < \gamma \ne 1$, then*

$$\sum_{t=1}^{T} X_{t-1}^{-\gamma} x_t \le C\delta^{-\gamma} + \frac{1}{1-\gamma}\left(\max\{\delta, X_T - C\}^{1-\gamma} - \delta^{1-\gamma}\right).$$

Appendix C. Online Convex Optimization with Adversarial Losses and Constraints

In this section, we show the performance of Algorithm 2, which is an AdaGrad-style variant of the drift-plus-penalty algorithm for online convex optimization with adversarial loss and constraint functions. To deal with adversarial constraint functions, we set the parameters V_t and α_t differently as follows.

$$a_t := \frac{F_t^2}{4} + R^2 G_t^2 + H_t^2, \quad S_t := \sum_{s=1}^{t} a_s$$

and then set parameters as

$$V_t = \frac{S_t^\beta}{R}, \quad \alpha_t = \frac{S_t}{R^2}, \tag{12}$$

for some $0 < \beta \le 1/2$. One distinction from (4) is the presence of additional parameter δ, and another difference is that V_t and α_t are defined with S_t, not S_{t-1}. We now assume that the convex constraint functions $g_1, \ldots, g_T$ as well as the convex loss functions $f_1, \ldots, f_T$ are chosen adversarially. Following Neely and Yu (2017), we set the benchmark $\boldsymbol{x}^\circ$ as an optimal solution to

$$\min \sum_{t=1}^{T} f_t(\boldsymbol{x}) \quad \text{subject to} \quad g_t(\boldsymbol{x}) \le 0 \quad \text{for } t = 1, \ldots, T.$$

Then the goal is to obtain upper bounds on

$$\text{Regret}(T) = \sum_{t=1}^{T} f_t(\boldsymbol{x}_t) - \sum_{t=1}^{T} f_t(\boldsymbol{x}^\circ), \quad \text{Violation}(T) = \sum_{t=1}^{T} g_t(\boldsymbol{x}_t)$$

in sublinear orders of T by properly choosing our inputs $\boldsymbol{x}_t$. For this, we need the following theorem.

Theorem 32 *Algorithm 2 with V_t and α set as in (12) guarantees that*

$$\text{Regret}(T) = O\left(S_T^{1-\beta}\right), \quad \text{Violation}(T) = O\left(S_T^{1/2} + T^{1/4}S_T^{\beta/2+1/4}\right).$$

Proof Applying Lemma 21 with $\hat{f}_t = f_t$, $\hat{g}_t = g_t$, and $\boldsymbol{x} = \boldsymbol{x}^\circ$, we obtain

$$\text{Regret}(T) \leq \frac{\alpha_T}{V_T}R^2 + \sum_{t=1}^{T}\frac{V_tF_t^2}{4\alpha_t} + \frac{1}{2}\sum_{t=1}^{T}\frac{(H_t+G_tR)^2}{V_t}.$$

The first term of the right hand side is $RS_T^{1-\beta} = O(S_T^{1-\beta})$, and the second term satisfies

$$\sum_{t=1}^{T}\frac{V_tF_t^2}{4\alpha_t} = \frac{R}{4}\sum_{t=1}^{T}S_t^{\beta-1}F_t^2 \leq R\sum_{t=1}^{T}S_t^{\beta-1}a_t \leq \frac{R}{\beta}S_T^{\beta} = O\left(S_T^{\beta}\right),$$

where the last inequality follows from Lemma 25. The third term satisfies

$$\sum_{t=1}^{T}\frac{(H_t+G_tR)^2}{2V_t} \leq R\sum_{t=1}^{T}\frac{R^2G_t^2+H_t^2}{S_t^{\beta}} \leq R\sum_{t=1}^{T}\frac{a_t}{S_t^{\beta}} \leq \frac{R}{1-\beta}S_T^{1-\beta} = O\left(S_T^{1-\beta}\right).$$

Combining these two inequalities, we get

$$\text{Regret}(T) = O\left(S_T^{1-\beta}\right).$$

Next, we prove the second part of the theorem. By Lemma 22, we have

$$\sum_{t=1}^{T}g_t(\boldsymbol{x}_t) \leq Q_{T+1} + \sum_{t=1}^{T}\frac{G_t}{2\alpha_t}(V_tF_t + Q_tG_t).$$

If we apply Lemma 24 with $\boldsymbol{x} = \boldsymbol{x}^\circ$, we obtain

$$Q_{T+1} \leq \sqrt{2\sum_{t=1}^{T}\left(H_t^2 + R^2G_t^2 + V_tRF_t\right) + 2R^2\alpha_T}$$

$$\leq \underbrace{\sqrt{2\sum_{t=1}^{T}RV_tF_t}}_{(a)} + \underbrace{\sqrt{2\sum_{t=1}^{T}(R^2G_t^2 + H_t^2)}}_{(b)} + \underbrace{\sqrt{2R^2\alpha_T}}_{(c)}.$$

where the first inequality follows from the convexity of g_t and $g_t(\boldsymbol{x}^\circ) \leq 0$. Here, term (c) is equal to $\sqrt{2S_T}$, and term (b) is less than or equal to $\sqrt{2S_T}$. For term (a), we have

$$(a)^2 = 2\sum_{t=1}^{T}S_t^{\beta}F_t \leq 2S_T^{\beta}\sum_{t=1}^{T}F_t \leq 2S_T^{\beta}\sqrt{T\sum_{t=1}^{T}F_t^2} \leq 4\sqrt{T}S_T^{\beta+1/2},$$

where the second inequality holds by the power mean inequality. Thus,

$$Q_{T+1} \le (a) + (b) + (c) \le 2\sqrt{2}S_T^{1/2} + 2T^{1/4}S_T^{\beta/2+1/4} = O\left(S_T^{1/2} + T^{1/4}S_T^{\beta/2+1/4}\right). \quad (13)$$

We also have that

$$
\begin{aligned}
\sum_{t=1}^{T} \frac{V_t F_t G_t}{2\alpha_t} &= \sum_{t=1}^{T} S_t^{\beta-1} R F_t G_t / 2 \\
&\le \sum_{t=1}^{T} S_t^{\beta-1}(F_t^2/4 + R^2 G_t^2)/2 \\
&\le \sum_{t=1}^{T} S_t^{\beta-1} a_t / 2 \le \frac{S_T^\beta}{2\beta} \\
&= O\left(S_T^\beta\right),
\end{aligned}
$$

where the last inequality follows from Lemma 25. Lastly,

$$
\begin{aligned}
\sum_{t=1}^{T} \frac{G_t^2}{2\alpha_t} Q_t &\le \sum_{t=1}^{T} \sqrt{2}\frac{G_t^2}{\alpha_t} S_t^{1/2} + \sum_{t=1}^{T} \frac{G_t^2}{\alpha_t} T^{1/4} S_t^{\beta/2+1/4} \\
&\le \sqrt{2}\sum_{t=1}^{T} \frac{a_t}{S_t^{1/2}} + T^{1/4}\sum_{t=1}^{T} \frac{a_t}{S_t^{3/4-\beta/2}} \\
&\le 2\sqrt{2S_T} + \frac{T^{1/4}}{\beta/2+1/4} S_T^{\beta/2+1/4} \\
&= O\left(S_T^{1/2} + T^{1/4}S_T^{\beta/2+1/4}\right),
\end{aligned}
$$

where the first inequality follows from (13) and the last inequality follows from Lemma 25. Combining the results, we get

$$\text{Violation}(T) = O\left(S_T^{1/2} + T^{1/4}S_T^{\beta/2+1/4}\right),$$

as required. ∎

Appendix D. Proof of the Time-Varying Drift Lemma

In this section, we prove Lemma 14 for the case of time-varying parameter $\theta_{t_0}(t)$. We closely follow the proof of (Yu et al., 2017, Lemma 5).

Lemma 33 *Let $r = \zeta/(4t_0\delta_{\max}^2)$ and $\rho = 1 - \zeta^2/(8\delta_{\max}^2) = 1 - rt_0\zeta/2$. Then*

$$\mathbb{E}\left[e^{rZ(t)}\right] \le \frac{e^{rt_0\delta_{\max}}}{1-\rho} e^{r\theta(t)}$$

for all $t \ge 0$.

Proof Since $0 < \zeta < \delta_{\max}$, we have $0 < \rho < 1 < e^{r\delta_{\max}}$. Define $\eta(t) = Z(t + t_0) - Z(t)$. Note that $|\eta(t)| \leq t_0 \delta_{\max}$ for all $t \geq 0$ which implies that $|r\eta(t)| \leq \zeta/(4\delta_{\max}) \leq 1$. Then,

$$e^{rZ(t+t_0)} = e^{rZ(t)}e^{r\eta(t)} \leq e^{rZ(t)}\left[1 + r\eta(t) + 2r^2 t_0^2 \delta_{\max}^2\right] = e^{rZ(t)}\left[1 + r\eta(t) + \frac{1}{2}rt_0\zeta\right] \quad (14)$$

where the inequality follows from the fact that $e^x \leq 1 + x + 2x^2$ for $|x| \leq 1$, $|r\eta(t)| \leq 1$, and $|\eta(t)| \leq t_0 \delta_{\max}$ while the equality follows by substituting $r = \zeta/(4t_0\delta_{\max}^2)$.

Next, we consider the cases $Z(t) \geq \theta(t)$ and $Z(t) < \theta(t)$, separately. First, consider the case $Z(t) \geq \theta(t)$. Taking the conditional expectation of each side of (14) gives us the following.

$$\mathbb{E}\left[e^{rZ(t+t_0)} \mid Z(t)\right] \leq \mathbb{E}\left[e^{rZ(t)}(1 + r\eta(t) + \frac{1}{2}rt_0\zeta) \mid Z(t)\right]$$

$$\leq e^{rZ(t)}\left[1 - rt_0\zeta + \frac{1}{2}rt_0\zeta\right]$$

$$= e^{rZ(t)}\left[1 - \frac{rt_0\zeta}{2}\right]$$

$$= \rho e^{rZ(t)}$$

where the inequality follows from the fact that $\mathbb{E}[Z(t + t_0) - Z(t)|\mathcal{F}(t)] \leq -t_0\zeta$ when $Z(t) \geq \theta(t)$ while the second equality follows from the fact that $\rho = 1 - rt_0\zeta/2$. Likewise, for the case $Z(t) < \theta(t)$, we deduce that

$$\mathbb{E}\left[e^{rZ(t+t_0)} \mid Z(t)\right] = \mathbb{E}\left[e^{rZ(t)}e^{r\eta(t)} \mid Z(t)\right] = e^{rZ(t)}\mathbb{E}\left[e^{r\eta(t)} \mid Z(t)\right] \leq e^{rt_0\delta_{\max}}e^{rZ(t)},$$

where the inequality follows from the fact that $\eta(t) \leq t_0\delta_{\max}$.

Putting the two cases together, we deduce that

$$\mathbb{E}\left[e^{rZ(t+t_0)}\right]$$

$$= \mathbb{P}(Z(t) \geq \theta(t))\mathbb{E}\left[e^{rZ(t+t_0)} \mid Z(t) \geq \theta(t)\right] + \mathbb{P}(Z(t) < \theta(t))\mathbb{E}\left[e^{rZ(t+t_0)} \mid Z(t) < \theta(t)\right]$$

$$\leq \rho\mathbb{E}\left[e^{rZ(t)} \mid Z(t) \geq \theta(t)\right]\mathbb{P}(Z(t) \geq \theta(t)) + e^{rt_0\delta_{\max}}\mathbb{E}\left[e^{rZ(t)} \mid Z(t) < \theta(t)\right]\mathbb{P}(Z(t) < \theta(t))$$

$$= \rho\mathbb{E}\left[e^{rZ(t)}\right] + \left(e^{rt_0\delta_{\max}} - \rho\right)\mathbb{E}\left[e^{rZ(t)} \mid Z(t) < \theta(t)\right]\mathbb{P}(Z(t) < \theta(t))$$

$$\leq \rho\mathbb{E}\left[e^{rZ(t)}\right] + \left(e^{rt_0\delta_{\max}} - \rho\right)e^{r\theta(t)}$$

$$\leq \rho\mathbb{E}\left[e^{rZ(t)}\right] + e^{rt_0\delta_{\max}}e^{r\theta(t)}$$

where the first inequality follows from the analysis of the two separate cases and the second inequality follows from the fact that $e^{rt_0\delta_{\max}} > \rho$.

Then we argue by induction to prove the statement of this lemma. We first consider the base case $t \in \{0, 1, \ldots, t_0\}$. Since $Z(t) \leq t\delta_{\max}$ for all $t \geq 0$, it follows that

$$\mathbb{E}[e^{rZ(t)}] \leq e^{rt\delta_{\max}} \leq e^{rt_0\delta_{\max}} \leq \frac{e^{rt_0\delta_{\max}}}{1 - \rho}e^{r\theta(t)}$$

for all $t \in \{1, \ldots, t_0\}$, where the last inequality follows because $e^{r\theta(t)}/(1 - \rho) \geq 1$. Next we assume that the inequality holds for all $t \in \{0, 1, \ldots, s\}$ with some $s \geq t_0$ and consider iteration $t = s + 1$. Note that

$$
\begin{aligned}
\mathbb{E}\left[e^{rZ(s+1)}\right] &\leq \rho \mathbb{E}\left[e^{rZ(s+1-t_0)}\right] + e^{rt_0\delta_{\max}}e^{r\theta(s+1-t_0)} \\
&\leq \rho \frac{e^{rt_0\delta_{\max}}}{1 - \rho}e^{r\theta(s+1-t_0)} + e^{rt_0\delta_{\max}}e^{r\theta(s+1-t_0)} \\
&\leq \frac{e^{rt_0\delta_{\max}}}{1 - \rho}e^{r\theta(s+1-t_0)} \\
&\leq \frac{e^{rt_0\delta_{\max}}}{1 - \rho}e^{r\theta(s+1)}
\end{aligned}
$$

where the second inequality comes from the induction hypothesis by noting that $0 \leq \tau + 1 - t_0 \leq \tau$ while the last inequality follows from the fact that $\theta(t)$ is non-decreasing. ∎

Based on this lemma, we prove Lemma 14.

Proof [Proof of Lemma 14] Note that e^{rx} is convex in x when $r > 0$. By Jensen's inequality,

$$
e^{r\mathbb{E}[Z(t)]} \leq \mathbb{E}[e^{rZ(t)}] \leq \frac{e^{r(\theta(t)+t_0\delta_{\max})}}{1 - \rho}
$$

where the inequality is implied by Lemma 33. Taking logarithm on both sides and dividing by r yields that

$$
\mathbb{E}[Z(t)] \leq \theta(t) + t_0\delta_{\max} + \frac{1}{r}\log\left[\frac{1}{1 - \rho}\right] = \theta(t) + t_0\delta_{\max} + t_0\frac{4\delta_{\max}^2}{\zeta}\log\left[\frac{8\delta_{\max}^2}{\zeta^2}\right],
$$

where the equality holds because recalling that $r = \frac{\zeta}{4t_0\delta_{\max}^2}$ and $\rho = 1 - \frac{\zeta^2}{8\delta_{\max}^2}$. ∎

Appendix E. Properties of the MLMC Estimator

In this section, we prove Lemmas 4 and 5.

Lemma 34 *(Dorfman and Levy, 2022, Lemma A.6) Let $h : \mathcal{X} \times \mathcal{S} \to \mathbb{R}^k$ for some $k \geq 1$. Suppose that there exists some constant $L > 0$ such that $\|h(\boldsymbol{x}, \boldsymbol{\xi})\| \leq L$ for every $(\boldsymbol{x}, \boldsymbol{\xi}) \in \mathcal{X} \times \mathcal{S}$, where the norm $\|\cdot\|$ satisfies $\|\cdot\| \leq \eta\|\cdot\|_2$ for some $\eta > 0$. We denote by*

$$
\bar{h}(\boldsymbol{x}) := \mathbb{E}_{\boldsymbol{\xi} \sim \mu}[h(\boldsymbol{x}, \boldsymbol{\xi})], \quad h_t^N(\boldsymbol{x}) := \frac{1}{N}\sum_{i=1}^{N}h(\boldsymbol{x}, \boldsymbol{\xi}_t^{(i)}).
$$

Suppose that $\boldsymbol{x}$ is $\mathcal{F}_{t-1}$ measurable and $N \leq A$ for some $A \in \mathbb{N}$. If $2\tau_{mix}\lceil 2\log A\rceil \leq N$, then

$$
\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|\right] \leq 12L\eta\sqrt{\frac{\tau_{mix}\lceil 2\log A\rceil}{N}}\left(1 + \sqrt{\log(\tau_{mix}\lceil 2\log A\rceil N)}\right)
$$
$$
+ \frac{6L\eta\tau_{mix}\lceil 2\log A\rceil}{N} + \frac{4L\eta}{N},
$$
$$
\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|^2\right] \leq \frac{576L^2\eta^2\tau_{mix}\lceil 2\log A\rceil}{N}(1 + \log(\tau_{mix}\lceil 2\log A\rceil N))
$$
$$
+ \frac{72L^2\eta^2\tau_{mix}^2\lceil 2\log A\rceil^2}{N^2} + \frac{8L^2\eta^2}{N}.
$$

We point out that (Dorfman and Levy, 2022, Lemma A.6) was originally stated for the ℓ_2 norm, but the statement holds for any norm over a finite-dimensional vector space assuming that $\|\cdot\| \leq \eta\|\cdot\|_2$ for some fixed constant $\eta > 0$ and O hides the dependence on η. In the following lemma, we simplify the upper bounds of Lemma 34.

Lemma 35 *Let $h : \mathcal{X} \times \mathcal{S} \to \mathbb{R}^k$ for some $k \geq 1$. Suppose that there exists some constant $L > 0$ such that $\|h(\boldsymbol{x}, \boldsymbol{\xi})\| \leq L$ for every $(\boldsymbol{x}, \boldsymbol{\xi}) \in \mathcal{X} \times \mathcal{S}$, where the norm $\|\cdot\|$ satisfies $\|\cdot\| \leq \eta\|\cdot\|_2$ for some $\eta > 0$. We denote by*

$$
\bar{h}(\boldsymbol{x}) := \mathbb{E}_{\boldsymbol{\xi} \sim \mu}[h(\boldsymbol{x}, \boldsymbol{\xi})], \quad h_t^N(\boldsymbol{x}) := \frac{1}{N}\sum_{i=1}^{N} h(\boldsymbol{x}, \boldsymbol{\xi}_t^{(i)}).
$$

Suppose that $\boldsymbol{x}$ is $\mathcal{F}_{t-1}$ measurable and $N \leq A$ for some $A \in \mathbb{N}$. Then

$$
\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|\right] \leq 12L\max(1, \eta)\sqrt{\frac{\tau_{mix}\lceil 2\log A\rceil}{N}}\left(2 + \sqrt{\log(\tau_{mix}\lceil 2\log A\rceil N)}\right),
$$
$$
\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|^2\right] \leq \frac{576L^2\max(1, \eta)^2\tau_{mix}\lceil 2\log A\rceil}{N}(2 + \log(\tau_{mix}\lceil 2\log A\rceil N)).
$$

Proof First, we consider the case where $2\tau_{\text{mix}}\lceil 2\log A\rceil > N$. Then by the triangle inequality,

$$
\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|\right] \leq 2L < 2L\sqrt{\frac{2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}}
$$
$$
\leq 12L\max(1, \eta)\sqrt{\frac{\tau_{\text{mix}}\lceil 2\log A\rceil}{N}}\left(2 + \sqrt{\log(\tau_{\text{mix}}\lceil 2\log A\rceil N)}\right),
$$
$$
\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|^2\right] \leq 4L^2 < 4L^2\frac{2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}
$$
$$
\leq \frac{576L^2\max(1, \eta)^2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}(2 + \log(\tau_{\text{mix}}\lceil 2\log A\rceil N)).
$$

Next, we consider the case where $2\tau_{\text{mix}}\lceil 2\log A\rceil \leq N$. By Lemma 34,

$$\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|\right] \leq 12L\eta\sqrt{\frac{\tau_{\text{mix}}\lceil 2\log A\rceil}{N}}\left(1 + \sqrt{\log(\tau_{\text{mix}}\lceil 2\log A\rceil N)}\right)$$
$$+ \frac{6L\eta\tau_{\text{mix}}\lceil 2\log A\rceil}{N} + \frac{4L\eta}{N},$$
$$\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|^2\right] \leq \frac{576L^2\eta^2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}(1 + \log(\tau_{\text{mix}}\lceil 2\log A\rceil N))$$
$$+ \frac{72L^2\eta^2\tau_{\text{mix}}^2\lceil 2\log A\rceil^2}{N^2} + \frac{8L^2\eta^2}{N}.$$

Since $\tau_{\text{mix}}\lceil 2\log A\rceil/N \leq \frac{1}{2}$, we deduce that

$$\frac{6L\eta\tau_{\text{mix}}\lceil 2\log A\rceil}{N} + \frac{4L\eta}{N} \leq \frac{12L\eta\tau_{\text{mix}}\lceil 2\log A\rceil}{N} \leq 12L\eta\sqrt{\frac{\tau_{\text{mix}}\lceil 2\log A\rceil}{N}}$$

and that

$$\frac{72L^2\eta^2\tau_{\text{mix}}^2\lceil 2\log A\rceil^2}{N^2} + \frac{8L^2\eta^2}{N} \leq \frac{72L^2\eta^2\tau_{\text{mix}}\lceil 2\log A\rceil}{N} + \frac{8L^2\eta^2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}$$
$$\leq \frac{576L^2\eta^2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}.$$

As a result, we obtain

$$\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|\right] \leq 12L\eta\sqrt{\frac{\tau_{\text{mix}}\lceil 2\log A\rceil}{N}}\left(2 + \sqrt{\log(\tau_{\text{mix}}\lceil 2\log A\rceil N)}\right)$$
$$\leq 12L\max(1,\eta)\sqrt{\frac{\tau_{\text{mix}}\lceil 2\log A\rceil}{N}}\left(2 + \sqrt{\log(\tau_{\text{mix}}\lceil 2\log A\rceil N)}\right),$$
$$\mathbb{E}_{t-1}\left[\|h_t^N(\boldsymbol{x}) - \bar{h}(\boldsymbol{x})\|^2\right] \leq \frac{576L^2\eta^2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}(2 + \log(\tau_{\text{mix}}\lceil 2\log A\rceil N))$$
$$\leq \frac{576L^2\max(1,\eta)^2\tau_{\text{mix}}\lceil 2\log A\rceil}{N}(2 + \log(\tau_{\text{mix}}\lceil 2\log A\rceil N)),$$

as required. ∎

Proof [Proof of Lemma 4] We first argue that

$$\mathbb{E}_{t-1}[f_t(\boldsymbol{x})] = \mathbb{E}_{t-1}\left[f_t^{2^{j_{\max}}}(\boldsymbol{x})\right]$$

for any $\boldsymbol{x}$ and t. Note that

$$\mathbb{E}_{t-1}[f_t] = \mathbb{E}_{t-1}\left[f_t^1\right] + \sum_{j=1}^{j_{\max}}\mathbb{P}(J_t = j)\, 2^j\mathbb{E}_{t-1}\left[f_t^{2^j} - f_t^{2^{j-1}}\right] = \mathbb{E}_{t-1}\left[f_t^{2^{j_{\max}}}\right]$$

as $\mathbb{P}(J_t = j) = 1/2^j$ Similarly, we can show that

$$\mathbb{E}_{t-1}[g_t] = \mathbb{E}_{t-1}\left[g_t^{2^{j_{\max}}}\right], \quad \mathbb{E}_{t-1}[\nabla f_t] = \mathbb{E}_{t-1}\left[\nabla f_t^{2^{j_{\max}}}\right], \quad \mathbb{E}_{t-1}[\nabla g_t] = \mathbb{E}_{t-1}\left[\nabla g_t^{2^{j_{\max}}}\right]$$

holds for any $\boldsymbol{x}$ and t. For the second part, we have

$$\mathbb{E}\left[|g_t|^2\right] \leq 2\mathbb{E}\left[\left|g_t - g_t^1\right|^2\right] + 2H^2$$

since $\|\boldsymbol{x} + \boldsymbol{y}\|^2 \leq (\|\boldsymbol{x}\| + \|\boldsymbol{y}\|)^2 \leq 2\|\boldsymbol{x}\|^2 + 2\|\boldsymbol{y}\|^2$ and $\left|g_t^1\right| \leq H$. Note that

$$\mathbb{E}\left[\left|g_t - g_t^1\right|^2\right] = \sum_{j=1}^{j_{\max}} \mathbb{P}(J_t = j)2^{2j}\mathbb{E}\left[\left|g_t^{2^j} - g_t^{2^{j-1}}\right|^2\right] = \sum_{j=1}^{j_{\max}} 2^j\mathbb{E}\left[\left|g_t^{2^j} - g_t^{2^{j-1}}\right|^2\right]$$

because $\mathbb{P}\left(J_t = j\right) = 1/2^j$. Here we can bound the right-most side based on the following.

$$\mathbb{E}\left[\left|g_t^{2^j} - g_t^{2^{j-1}}\right|^2\right] \leq 2\mathbb{E}\left[\left|g_t^{2^j} - \bar{g}\right|^2\right] + 2\mathbb{E}\left[\left|g_t^{2^{j-1}} - \bar{g}\right|^2\right] = \tilde{\mathcal{O}}\left(\frac{\tau_{\mathrm{mix}}}{2^j}\right)$$

where the last inequality follows from Lemma 35. Then it follows that

$$\mathbb{E}\left[\left|g_t - g_t^1\right|^2\right] \leq \tilde{\mathcal{O}}(j_{\max}\tau_{\mathrm{mix}}) = \tilde{\mathcal{O}}(\tau_{\mathrm{mix}})$$

where the last equality holds because $j_{\max} = O(\log T)$. For the last part,

$$\mathbb{E}[N_t] = 1 + \sum_{j=1}^{j_{\max}} \mathbb{P}(J_t = j)(2^j - 1) \leq 1 + j_{\max} \leq 1 + 2\log_2 T,$$

as required. ∎

Proof [Proof of Lemma 5] By Assumption 2, we can apply Lemma 35 to $g_t, \nabla g_t, \nabla f_t$. Assumptions 1, 2 implies that $|f_t(\boldsymbol{x}_t)| \leq J$ for some $J > 0$. Thus, we can apply Lemma 35 to f_t as well. Let $L = \max(F, G, H, J)$, $A = T^2$, and $\|\cdot\|_* \leq \eta\|\cdot\|_2$ for some $\eta > 0$. By Lemma 35, the statement follows for

$$C(T) = \sqrt{576L^2\max(1,\eta)^2\lceil 4\log T\rceil(2 + \log(\tau_{\mathrm{mix}}\lceil 4\log T\rceil 2^{j_{\max}}))}.$$

Since $2^{j_{\max}} = \mathcal{O}(T^2)$, the order of $C(T)$ follows. ∎

References

Z. Akhtar, A. S. Bedi, and K. Rajawat. Conservative stochastic optimization with expectation constraints. *IEEE Transactions on Signal Processing*, 69:3190–3205, 2021. doi: 10.1109/TSP.2021.3082467.

C. Andrieu, N. de Freitas, A. Doucet, and M. I. Jordan. An introduction to mcmc for machine learning. *Machine Learning*, 50(1):5–43, 2003. doi: 10.1023/A:1020281327116. URL https://doi.org/10.1023/A:1020281327116.

P. Auer, N. Cesa-Bianchi, and C. Gentile. Adaptive and self-confident on-line learning algorithms. *Journal of Computer and System Sciences*, 64(1):48–75, 2002. ISSN 0022-0000. doi: https://doi.org/10.1006/jcss.2001.1795. URL `https://www.sciencedirect.com/science/article/pii/S0022000001917957`.

G. Ayache, V. Dassari, and S. E. Rouayheb. Walk for learning: A random walk approach for federated learning from heterogeneous data. *IEEE J.Sel. A. Commun.*, 41(4):929–940, apr 2023. ISSN 0733-8716. doi: 10.1109/JSAC.2023.3244250. URL `https://doi.org/10.1109/JSAC.2023.3244250`.

A. Benveniste, P. Priouret, and M. Métivier. *Adaptive Algorithms and Stochastic Approximations*. Springer-Verlag, Berlin, Heidelberg, 1990. ISBN 0387528946.

D. P. Bertsekas and J. N. Tsitsiklis. *Neuro-Dynamic Programming*. Athena Scientific, 1st edition, 1996. ISBN 1886529108.

J. H. Blanchet and P. W. Glynn. Unbiased monte carlo for optimization and functions of expectations via multi-level randomization. In *2015 Winter Simulation Conference (WSC)*, pages 3656–3667, 2015. doi: 10.1109/WSC.2015.7408524.

V. S. Borkar. *Stochastic Approximation A Dynamical Systems Viewpoint*. Texts and Readings in Mathematics ; 48. Hindustan Book Agency, Gurgaon, 1st ed. 2008. edition, 2008. ISBN 93-86279-38-X.

L. Bottou, F. E. Curtis, and J. Nocedal. Optimization methods for large-scale machine learning. *SIAM Review*, 60(2):223–311, 2018. doi: 10.1137/16M1080173. URL `https://doi.org/10.1137/16M1080173`.

J. Cao, R. Jiang, N. Abolfazli, E. Y. Hamedani, and A. Mokhtari. Projection-free methods for stochastic simple bilevel optimization with convex lower-level problem, 2023.

L. E. Celis, D. Straszak, and N. K. Vishnoi. Ranking with Fairness Constraints. In I. Chatzigiannakis, C. Kaklamanis, D. Marx, and D. Sannella, editors, *45th International Colloquium on Automata, Languages, and Programming (ICALP 2018)*, volume 107 of *Leibniz International Proceedings in Informatics (LIPIcs)*, pages 28:1–28:15, Dagstuhl, Germany, 2018. Schloss Dagstuhl–Leibniz-Zentrum fuer Informatik. ISBN 978-3-95977-076-7. doi: 10.4230/LIPIcs.ICALP.2018.28. URL `http://drops.dagstuhl.de/opus/volltexte/2018/9032`.

L. E. Celis, L. Huang, V. Keswani, and N. K. Vishnoi. Classification with fairness constraints: A meta-algorithm with provable guarantees. In *Proceedings of the Conference on Fairness, Accountability, and Transparency*, FAT* '19, page 319–328, New York, NY, USA, 2019. Association for Computing Machinery. ISBN 9781450361255. doi: 10.1145/3287560.3287586. URL `https://doi.org/10.1145/3287560.3287586`.

O. Chapelle, B. Schlkopf, and A. Zien. *Semi-Supervised Learning*. The MIT Press, 1st edition, 2010. ISBN 0262514125.

D. Dentcheva and A. Ruszczynski. Optimization with stochastic dominance constraints. *SIAM Journal on Optimization*, 14(2):548–566, 2003. doi: 10.1137/S1052623402420528. URL https://doi.org/10.1137/S1052623402420528.

T. T. Doan. Finite-time analysis of markov gradient descent. *IEEE Transactions on Automatic Control*, 68(4):2140–2153, 2023. doi: 10.1109/TAC.2022.3172593.

T. T. Doan, L. M. Nguyen, N. H. Pham, and J. Romberg. Convergence rates of accelerated markov gradient descent with applications in reinforcement learning, 2020.

M. Donini, L. Oneto, S. Ben-David, J. Shawe-Taylor, and M. Pontil. Empirical risk minimization under fairness constraints. In *Proceedings of the 32nd International Conference on Neural Information Processing Systems*, NIPS'18, page 2796–2806, Red Hook, NY, USA, 2018. Curran Associates Inc.

R. Dorfman and K. Y. Levy. Adapting to mixing time in stochastic optimization with Markovian data. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 5429–5446. PMLR, 17–23 Jul 2022. URL https://proceedings.mlr.press/v162/dorfman22a.html.

J. Duchi, E. Hazan, and Y. Singer. Adaptive subgradient methods for online learning and stochastic optimization. *Journal of Machine Learning Research*, 12(61):2121–2159, 2011. URL http://jmlr.org/papers/v12/duchi11a.html.

J. C. Duchi, A. Agarwal, M. Johansson, and M. I. Jordan. Ergodic mirror descent. *SIAM Journal on Optimization*, 22(4):1549–1578, 2012. doi: 10.1137/110836043. URL https://doi.org/10.1137/110836043.

Y. Ermoliev. stochastic quasigradient methods and their application to system optimization †. *Stochastics*, 9(1-2):1–36, 1983. doi: 10.1080/17442508308833246.

M. Even. Stochastic gradient descent under Markovian sampling schemes. In A. Krause, E. Brunskill, K. Cho, B. Engelhardt, S. Sabato, and J. Scarlett, editors, *Proceedings of the 40th International Conference on Machine Learning*, volume 202 of *Proceedings of Machine Learning Research*, pages 9412–9439. PMLR, 23–29 Jul 2023. URL https://proceedings.mlr.press/v202/even23a.html.

J. García, Fern, and o Fernández. A comprehensive survey on safe reinforcement learning. *Journal of Machine Learning Research*, 16(42):1437–1480, 2015. URL http://jmlr.org/papers/v16/garcia15a.html.

M. B. Giles. Multilevel monte carlo methods. *Acta Numerica*, 24:259–328, 2015. doi: 10.1017/S096249291500001X.

H. Guo, X. Liu, H. Wei, and L. Ying. Online convex optimization with hard constraints: Towards the best of two worlds and beyond. In A. H. Oh, A. Agarwal, D. Belgrave, and K. Cho, editors, *Advances in Neural Information Processing Systems*, 2022. URL https://openreview.net/forum?id=rwdpFgfVpvN.

H. Hendrikx. A principled framework for the design and analysis of token algorithms. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 470–489. PMLR, 25–27 Apr 2023. URL `https://proceedings.mlr.press/v206/hendrikx23a.html`.

A. Jalilzadeh, F. Yousefian, and M. Ebrahimi. Stochastic approximation for estimating the price of stability in stochastic nash games, 2023.

R. Jenatton, J. Huang, and C. Archambeau. Adaptive algorithms for online convex optimization with long-term constraints. In M. F. Balcan and K. Q. Weinberger, editors, *Proceedings of The 33rd International Conference on Machine Learning*, volume 48 of *Proceedings of Machine Learning Research*, pages 402–411, New York, New York, USA, 20–22 Jun 2016. PMLR. URL `https://proceedings.mlr.press/v48/jenatton16.html`.

B. Johansson, M. Rabi, and M. Johansson. A simple peer-to-peer algorithm for distributed optimization in sensor networks. In *2007 46th IEEE Conference on Decision and Control*, pages 4705–4710, 2007. doi: 10.1109/CDC.2007.4434888.

B. Johansson, A. Speranzon, M. Johansson, and K. H. Johansson. On decentralized negotiation of optimal consensus. *Automatica*, 44(4):1175–1179, 2008. ISSN 0005-1098. doi: https://doi.org/10.1016/j.automatica.2007.09.003. URL `https://www.sciencedirect.com/science/article/pii/S0005109807003962`.

B. Johansson, M. Rabi, and M. Johansson. A randomized incremental subgradient method for distributed optimization in networked systems. *SIAM Journal on Optimization*, 20(3):1157–1170, 2010. doi: 10.1137/08073038X. URL `https://doi.org/10.1137/08073038X`.

A. Juditsky, A. Nemirovski, and C. Tauvel. Solving variational inequalities with stochastic mirror-prox algorithm. *Stochastic Systems*, 1(1):17–58, 2011. doi: 10.1287/10-SSY011. URL `https://doi.org/10.1287/10-SSY011`.

S. Kowshik, D. Nagaraj, P. Jain, and P. Netrapalli. Streaming linear system identification with reverse experience replay. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 30140–30152. Curran Associates, Inc., 2021. URL `https://proceedings.neurips.cc/paper_files/paper/2021/file/fd2c5e4680d9a01dba3aada5ece22270-Paper.pdf`.

G. Lan and Z. Zhou. Algorithms for stochastic optimization with function or expectation constraints. *Computational Optimization and Applications*, 76(2):461–498, 2020. doi: 10.1007/s10589-020-00179-x. URL `https://doi.org/10.1007/s10589-020-00179-x`.

D. Lee, N. Ho-Nguyen, and D. Lee. Projection-free online convex optimization with stochastic constraints, 2023.

D. Levin and Y. Peres. *Markov Chains and Mixing Times*. MBK. American Mathematical Society, 2017. ISBN 9781470429621. URL `https://books.google.com/books?id=f208DwAAQBAJ`.

K. Levy. Online to offline conversions, universality and adaptive minibatch sizes. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL https://proceedings.neurips.cc/paper_files/paper/2017/file/ce5140df15d046a66883807d18d0264b-Paper.pdf.

Q. Lin, S. Nadarajah, N. Soheili, and T. Yang. A data efficient and feasible level set method for stochastic convex optimization with expectation constraints. *Journal of Machine Learning Research*, 21(143):1–45, 2020. URL http://jmlr.org/papers/v21/19-1022.html.

C. G. Lopes and A. H. Sayed. Incremental adaptive strategies over distributed networks. *IEEE Transactions on Signal Processing*, 55(8):4064–4077, 2007. doi: 10.1109/TSP.2007.896034.

M. Mahdavi, R. Jin, and T. Yang. Trading regret for efficiency: Online convex optimization with long term constraints. *Journal of Machine Learning Research*, 13(81):2503–2528, 2012. URL http://jmlr.org/papers/v13/mahdavi12a.html.

X. Mao, K. Yuan, Y. Hu, Y. Gu, A. H. Sayed, and W. Yin. Walkman: A communication-efficient random-walk algorithm for decentralized optimization. *IEEE Transactions on Signal Processing*, 68:2513–2528, 2020. doi: 10.1109/TSP.2020.2983167.

D. Nagaraj, X. Wu, G. Bresler, P. Jain, and P. Netrapalli. Least squares regression with markovian data: Fundamental limits and algorithms. In H. Larochelle, M. Ranzato, R. Hadsell, M. Balcan, and H. Lin, editors, *Advances in Neural Information Processing Systems*, volume 33, pages 16666–16676. Curran Associates, Inc., 2020. URL https://proceedings.neurips.cc/paper_files/paper/2020/file/c22abfa379f38b5b0411bc11fa9bf92f-Paper.pdf.

M. J. Neely and H. Yu. Online convex optimization with time-varying constraints, 2017. URL https://arxiv.org/abs/1702.04783.

A. Nemirovski and A. Shapiro. Convex approximations of chance constrained programs. *SIAM Journal on Optimization*, 17(4):969–996, 2007. doi: 10.1137/050622328. URL https://doi.org/10.1137/050622328.

A. Nemirovski, A. Juditsky, G. Lan, and A. Shapiro. Robust stochastic approximation approach to stochastic programming. *SIAM Journal on Optimization*, 19(4):1574–1609, 2009. doi: 10.1137/070704277. URL https://doi.org/10.1137/070704277.

G. Pflug. *The Interface Between Simulation and Optimization*. 1996.

B. Poljak and J. Tsypkin. Robust identification. *Automatica*, 16(1):53–63, 1980. ISSN 0005-1098. doi: https://doi.org/10.1016/0005-1098(80)90086-2. URL https://www.sciencedirect.com/science/article/pii/0005109880900862.

B. Polyak. A general method of solving extremum problems. *Doklady Akademii Nauk SSSR*, 174(1):33–36, 1967.

M. Rabbat and R. Nowak. Distributed optimization in sensor networks. In *Third International Symposium on Information Processing in Sensor Networks, 2004. IPSN 2004*, pages 20–27, 2004. doi: 10.1145/984622.984626.

S. S. Ram, A. Nedić, and V. V. Veeravalli. Incremental stochastic subgradient algorithms for convex optimization. *SIAM Journal on Optimization*, 20(2):691–717, 2009a. doi: 10.1137/080726380. URL https://doi.org/10.1137/080726380.

S. S. Ram, A. Nedić, and V. V. Veeravalli. Incremental stochastic subgradient algorithms for convex optimization. *SIAM Journal on Optimization*, 20(2):691–717, 2009b. doi: 10.1137/080726380. URL https://doi.org/10.1137/080726380.

P. Rigollet and X. Tong. Neyman-pearson classification, convexity and stochastic constraints. *Journal of Machine Learning Research*, 12(86):2831–2855, 2011. URL http://jmlr.org/papers/v12/rigollet11a.html.

H. Robbins and S. Monro. A Stochastic Approximation Method. *The Annals of Mathematical Statistics*, 22(3):400 – 407, 1951. doi: 10.1214/aoms/1177729586. URL https://doi.org/10.1214/aoms/1177729586.

R. Rockafellar and S. Uryasev. Optimization of conditional value-at-risk. *Journal of Risk*, 2: 21–41, 2000.

A. Roy, K. Balasubramanian, and S. Ghadimi. Constrained stochastic nonconvex optimization with state-dependent markov data. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 23256–23270. Curran Associates, Inc., 2022. URL https://proceedings.neurips.cc/paper_files/paper/2022/file/93b11b5128ced940120f41ce9b216f39-Paper-Conference.pdf.

A. Ruszczyński and W. Syski. *A method of aggregate stochastic subgradients with on-line stepsize rules for convex stochastic programming problems*, pages 113–131. Springer Berlin Heidelberg, Berlin, Heidelberg, 1986. ISBN 978-3-642-00927-3. doi: 10.1007/BFb0121128. URL https://doi.org/10.1007/BFb0121128.

T. Sarkar and A. Rakhlin. Near optimal finite time identification of arbitrary linear dynamical systems. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 5610–5618. PMLR, 09–15 Jun 2019. URL https://proceedings.mlr.press/v97/sarkar19a.html.

C. Scott and R. Nowak. A neyman-pearson approach to statistical learning. *IEEE Transactions on Information Theory*, 51(11):3806–3819, 2005. doi: 10.1109/TIT.2005.856955.

T. Sun, Y. Sun, and W. Yin. On markov chain gradient descent. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL https://proceedings.neurips.cc/paper_files/paper/2018/file/1371bccec2447b5aa6d96d2a540fb401-Paper.pdf.

T. Sun, Y. Sun, Y. Xu, and W. Yin. Markov chain block coordinate descent. *Computational Optimization and Applications*, 75(1):35–61, 2020.

T. Sun, D. Li, and B. Wang. Adaptive random walk gradient descent for decentralized optimization. In K. Chaudhuri, S. Jegelka, L. Song, C. Szepesvari, G. Niu, and S. Sabato, editors, *Proceedings of the 39th International Conference on Machine Learning*, volume 162 of *Proceedings of Machine Learning Research*, pages 20790–20809. PMLR, 17–23 Jul 2022. URL `https://proceedings.mlr.press/v162/sun22b.html`.

T. Sun, D. Li, and B. Wang. On the decentralized stochastic gradient descent with markov chain sampling. *IEEE Transactions on Signal Processing*, 71:2895–2909, 2023. doi: 10.1109/TSP.2023.3297053.

P. Wang, Y. Lei, Y. Ying, and D.-X. Zhou. Stability and generalization for markov chain stochastic gradient methods. In S. Koyejo, S. Mohamed, A. Agarwal, D. Belgrave, K. Cho, and A. Oh, editors, *Advances in Neural Information Processing Systems*, volume 35, pages 37735–37748. Curran Associates, Inc., 2022. URL `https://proceedings.neurips.cc/paper_files/paper/2022/file/f61538f83b0f19f9306d9d801c15f41c-Paper-Conference.pdf`.

R. Ward, X. Wu, and L. Bottou. AdaGrad stepsizes: Sharp convergence over nonconvex landscapes. In K. Chaudhuri and R. Salakhutdinov, editors, *Proceedings of the 36th International Conference on Machine Learning*, volume 97 of *Proceedings of Machine Learning Research*, pages 6677–6686. PMLR, 09–15 Jun 2019. URL `https://proceedings.mlr.press/v97/ward19a.html`.

X. Wei, H. Yu, and M. J. Neely. Online primal-dual mirror descent under stochastic constraints. *Proc. ACM Meas. Anal. Comput. Syst.*, 4(2), jun 2020. doi: 10.1145/3392157. URL `https://doi.org/10.1145/3392157`.

X. Xiao. Penalized stochastic gradient methods for stochastic convex optimization with expectation constraints. Technical report, September 2019. URL `https://optimization-online.org/?p=15973`.

Y. Yan and Y. Xu. Adaptive primal-dual stochastic gradient method for expectation-constrained convex stochastic programs. *Mathematical Programming Computation*, 14(2):319–363, 2022. doi: 10.1007/s12532-021-00214-w. URL `https://doi.org/10.1007/s12532-021-00214-w`.

Z. Yang, Y. Lei, P. Wang, T. Yang, and Y. Ying. Simple stochastic and online gradient descent algorithms for pairwise learning. In M. Ranzato, A. Beygelzimer, Y. Dauphin, P. Liang, and J. W. Vaughan, editors, *Advances in Neural Information Processing Systems*, volume 34, pages 20160–20171. Curran Associates, Inc., 2021. URL `https://proceedings.neurips.cc/paper_files/paper/2021/file/a87d27f712df362cd22c7a8ef823e987-Paper.pdf`.

S. Yao and B. Huang. Beyond parity: Fairness objectives for collaborative filtering. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and

R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper_files/paper/2017/file/e6384711491713d29bc63fc5eeb5ba4f-Paper.pdf`.

E. Yazdandoost Hamedani, A. Jalilzadeh, and N. S. Aybat. Randomized primal-dual methods with adaptive step sizes. In F. Ruiz, J. Dy, and J.-W. van de Meent, editors, *Proceedings of The 26th International Conference on Artificial Intelligence and Statistics*, volume 206 of *Proceedings of Machine Learning Research*, pages 11185–11212. PMLR, 25–27 Apr 2023. URL `https://proceedings.mlr.press/v206/yazdandoost-hamedani23a.html`.

X. Yi, X. Li, T. Yang, L. Xie, T. Chai, and K. Johansson. Regret and cumulative constraint violation analysis for online convex optimization with long term constraints. In M. Meila and T. Zhang, editors, *Proceedings of the 38th International Conference on Machine Learning*, volume 139 of *Proceedings of Machine Learning Research*, pages 11998–12008. PMLR, 18–24 Jul 2021. URL `https://proceedings.mlr.press/v139/yi21b.html`.

H. Yu, M. Neely, and X. Wei. Online convex optimization with stochastic constraints. In I. Guyon, U. V. Luxburg, S. Bengio, H. Wallach, R. Fergus, S. Vishwanathan, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 30. Curran Associates, Inc., 2017. URL `https://proceedings.neurips.cc/paper/2017/file/da0d1111d2dc5d489242e60ebcbaf988-Paper.pdf`.

J. Yuan and A. Lamperski. Online convex optimization for cumulative constraints. In S. Bengio, H. Wallach, H. Larochelle, K. Grauman, N. Cesa-Bianchi, and R. Garnett, editors, *Advances in Neural Information Processing Systems*, volume 31. Curran Associates, Inc., 2018. URL `https://proceedings.neurips.cc/paper/2018/file/9cb9ed4f35cf7c2f295cc2bc6f732a84-Paper.pdf`.

M. B. Zafar, I. Valera, M. Gomez-Rodriguez, and K. P. Gummadi. Fairness constraints: A flexible approach for fair classification. *Journal of Machine Learning Research*, 20(75): 1–42, 2019. URL `http://jmlr.org/papers/v20/18-262.html`.

L. Zhang, Y. Zhang, J. Wu, and X. Xiao. Solving stochastic optimization with expectation constraints efficiently by a stochastic augmented lagrangian-type algorithm. *INFORMS Journal on Computing*, 34(6):2989–3006, 2022. doi: 10.1287/ijoc.2022.1228. URL `https://doi.org/10.1287/ijoc.2022.1228`.

L. Zhang, Y. Zhang, X. Xiao, and J. Wu. Stochastic approximation proximal method of multipliers for convex stochastic programming. *Mathematics of Operations Research*, 48 (1):177–193, 2023. doi: 10.1287/moor.2022.1257. URL `https://doi.org/10.1287/moor.2022.1257`.

R. Zhao. Accelerated stochastic algorithms for convex-concave saddle-point problems. *Mathematics of Operations Research*, 47(2):1443–1473, 2022. doi: 10.1287/moor.2021.1175. URL `https://doi.org/10.1287/moor.2021.1175`.

Y. Zhao, M. Li, L. Lai, N. Suda, D. Civin, and V. Chandra. Federated learning with non-iid data. 2018. doi: 10.48550/ARXIV.1806.00582. URL `https://arxiv.org/abs/1806.00582`.