Primal-Dual Contextual Bayesian Optimization for Control System Online Optimization with Time-Average Constraints

Wenjie Xu, Yuning Jiang, Bratislav Svetozarevic, Colin N. Jones

***Abstract*— This paper studies the problem of online performance optimization of constrained closed-loop control systems, where both the objective and the constraints are unknown black-box functions affected by exogenous time-varying contextual disturbances. A primal-dual contextual Bayesian optimization algorithm is proposed that achieves sublinear cumulative regret with respect to the dynamic optimal solution under certain regularity conditions. Furthermore, the algorithm achieves zero time-average constraint violation, ensuring that the average value of the constraint function satisfies the desired constraint. The method is applied to both sampled instances from Gaussian processes and a continuous stirred tank reactor parameter tuning problem; simulation results show that the method simultaneously provides close-to-optimal performance and maintains constraint feasibility on average. This contrasts current state-of-the-art methods, which either suffer from large cumulative regret or severe constraint violations for the case studies presented.**

I. Introduction

Constrained closed-loop control systems are critical to a wide range of applications, and optimizing their online performance by tuning control parameters is a common challenge. This paper studies the time-varying black-box optimization problems that arise in optimizing system performance with constraints, where both the objective and the constraints are unknown. For instance, consider building control, which aims to minimize energy consumption while meeting occupant comfort requirements [1]. Solving the controller tuning problem in such scenarios is challenging for several reasons. Firstly, it is hard to model the mapping from control parameters to the performance metric of the closed-loop system. Secondly, unknown constraint violations need to be taken care of during the optimization process. Finally, evaluating the system's performance is often expensive in practice.

To tackle these challenges, Bayesian optimization has shown promise as a sample-efficient, derivative-free black-box optimization method [2]. Bayesian optimization constructs a surrogate model using Gaussian process regression [3] and uses this model to guide the sampling of the black-box functions. Variants of Bayesian optimization methods have been proposed to handle constrained optimization problems. One common method is to maximize constrained expected improvement (CEI) [4], [5] to select the next sample in each step. Another line of research developed safe Bayesian optimization (Safe BO) methods by restricting the sampling to only feasible points [6]–[8]. These methods have found wide applications in control system optimization. For example, CEI method is applied to trajectory optimization for path following control [9]. Safe BO is applied to tune the PI controller of a room heating system subject to a group of safe constraints [10]. However, the CEI method may not guarantee constraint feasibility and can suffer from severe constraint violations, while safe Bayesian optimization methods can be too cautious and lose performance due to the strict requirement of sampling feasible points.

In the context of controller tuning, one additional challenge that needs to be considered is that the unknown objective and constraints are time-varying [11]. To address this challenge, several time-varying variants of Bayesian optimization have been proposed in the literature [11], [12]. In practice, these variations can be due to the change of 'contextual variables' observed by the decision-maker before selecting a new set of parameters [13]. Affected by the time-varying contextual disturbances, guaranteeing constraint satisfaction at every time instant can be challenging and may not be necessary [14]. But rather, it can be more of interest to satisfy the *time-average* constraint [15], especially when it represents some economic cost that accumulates over time. Examples include performance optimization of a machine subject to fatigue constraints (see, e.g., [16]) and data center cooling subject to the constraints on the number of delayed queries per unit of time [17].

Motivated by the aforementioned observations, we formulate the time-varying black-box optimization problem as a constrained contextual Bayesian optimization problem. To solve it, we extend the primal-dual Bayesian optimization framework [18] to the contextual setting [19]. In the proposed algorithm, the contextual variable that impacts the response surface of the black-box functions can be observed at each step before selecting a new set of parameters, and then the system performance is optimized over the current context. The detailed contribution is summarized as follows:

- We propose a **P**rimal-**D**ual **C**ontextual **B**ayesian **O**ptimization (PDCBO) algorithm for time-varying constrained black-box optimization problems. In contrast to the regret with respect to a static optimal value in the non-contextual setting [18], we provide bounds on the contextual regret with respect to the time-varying optimal solutions. Furthermore, our algorithm can achieve *zero* time-averaged constraint violations, even under the *adversarial* time-varying contextual setting;

This work was supported by the Swiss National Science Foundation under the NCCR Automation (grant agreement 51NF40_180545).

W. Xu, Y. Jiang, and C. N. Jones are with the Automatic Control Laboratory, EPFL, Switzerland. W. Xu and B. Svetozarevic are with Swiss Federal Laboratories for Materials Science and Technology (Empa), Switzerland. {wenjie.xu, yuning.jiang, colin.jones}@epfl.ch, bratislav.svetozarevic@empa.ch

- The proposed method is deployed to both sampled instances from Gaussian processes and to a continuous stirred-tank reactor parameter tuning problem. Compared to other state-of-the-art methods, our method can simultaneously achieve the lowest cumulative objective while satisfying all the constraints on average.

II. Problem Formulation

This work aims to sequentially optimize the control parameters $\theta_t \in \Theta \subset \mathbb{R}^{n_\theta}$ after observing some contextual variable $z_t \in \mathcal{Z} \subset \mathbb{R}^{n_z}$ in each step t, where Θ is the candidate set of the control parameters and $\mathcal{Z}$ is the set of possible contextual variables. Our problem can be thus, formulated as follows:

$$\min_{\theta \in \Theta} \quad f(\theta, z), \tag{1a}$$

$$\text{subject to:} \quad g_i(\theta, z) \leq 0, \quad \forall i \in [N], \tag{1b}$$

where $[N]$ is defined as the set $\{1, 2, \cdots, N\} \subset \mathbb{Z}$, $f : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_z} \to \mathbb{R}$ is the black-box objective function, and $g_i : \mathbb{R}^{n_\theta} \times \mathbb{R}^{n_z} \to \mathbb{R}, i \in [N]$ is a set of black-box constraints to be satisfied. We use $\theta^*(z)$ to denote the optimal solution of the problem (1), which depends on the current context variable z. We use g to denote the concatenation $(g_i)_{i=1}^N$ and $g(\theta, z)$ to denote the concatenation $(g_i(\theta, z))_{i=1}^N$.

We make some assumptions regarding the regularities of the problem.

Assumption 1 *Both $\mathcal{Z}$ and Θ are compact.*

Assumption 1 is commonly seen in practice. For example, we can restrict the control parameter set Θ to a hyper box and often know an apriori bound on the range of contextual variables.

Assumption 2 *$f \in \mathcal{H}_0, g_i \in \mathcal{H}_i$, where $\mathcal{H}_i, i \in \{0\} \cup [N]$ is a reproducing kernel Hilbert space (RKHS) equipped with kernel function $k_i(\cdot, \cdot)$ (See [20].). Furthermore, $\|f\| \leq C_0, \|g_i\| \leq C_i, \forall i \in [N]$, where $\|\cdot\|$ is the norm induced by the inner product of the corresponding RKHS without further notice.*

Assumption 2 requires that the underlying functions are regular in the sense of having a bounded norm in an RKHS. In the existing literature, having a bounded norm in an RKHS is a commonly adopted assumption (e.g., [18], [21]). Intuitively, it means the black-box function has a 'smoothness' property at least to a certain level (See [20]).

Assumption 3 *We only have access to a noisy zero-order oracle, which means each round of query θ_t with contextual variable z_t returns the noisy function evaluations,*

$$y_{0,t} = f(x_t) + \nu_{0,t} \ , \tag{2a}$$

$$y_{i,t} = g_i(x_t) + \nu_{i,t} \ , \quad i \in [N], \tag{2b}$$

where x_t denotes the concatenation (θ_t, z_t) and $\nu_{i,t}, i \in \{0\} \cup [N]$ are independent and identically distributed σ-sub-Gaussian noise.

Assumption 3 is reasonable as, in many applications, the performance value of a control system can be obtained through experiment or simulation with noise. However, the gradient or other higher-order information can be difficult to acquire.

Throughout the rest of this paper, we use the notation $X_t := (x_1, x_2, \cdots, x_t)$ to define the sequence of sampled points up to step t. Therefore, the historical evaluations are

$$\mathcal{D}_{1:t} := \{(x_\tau, y_{0,\tau}, \cdots, y_{N,\tau})\}_{\tau=1}^t .$$

Assumption 4 *The kernel function is normalized, such that, $k_i(x, x) \leq 1, \forall x \in \mathcal{X}, i \in \{0\} \cup [N]$, where $\mathcal{X} = \mathcal{Z} \times \Theta$.*

Assumption 5 *There exists $\xi > 0$ independent of the context $z \in \mathcal{Z}$ and $\bar{\theta}(z) \in \Theta$, which can be dependent on z, such that $g(\bar{\theta}(z), z) \leq -\xi e$, where $e \in \mathbb{R}^N$ is the vector with all 1s and the inequality is interpreted elementwise.*

Asssumption 5 is the so-called Uniform Slater Condition. As an immediate implication of Assumption 5, we have the following result.

Proposition 1 *There exists a probability distribution $\bar{\pi}(z)$ over Θ, such that $\mathbb{E}_{\bar{\pi}(z)}(g(\theta, z)) \leq -\xi e$.*

III. Performance metric

We compare the controller parameters selected by our algorithm with the optimal tracking solution of problem (1), given the context z taking the value z_t at step t.

We are interested in two metrics,

$$\mathcal{R}_T = \sum_{t=1}^T \left(f(\theta_t, z_t) - f(\theta^*(z_t), z_t)\right), \tag{3a}$$

$$\mathcal{V}_T = \left\| \left[\sum_{t=1}^T g(\theta_t, z_t) \right]^+ \right\|, \tag{3b}$$

which are the cumulative regret compared to the optimal tracking solutions and the cumulative constraint violations. The regret here is essentially the *contextual regret* [19], which is stronger than the regret considered in [18].

The form of constraint violation in (3a) is the violation of cumulative constraint value. $\mathcal{V}_T/T$ gives the violation of the average constraint value, which is common in practice. For example, when g represents some economic cost (e.g., monetary expenses and energy consumption), it is usually of more interest to bound the average constraint value during a period rather than the whole constraint function sequence.

IV. Gaussian process regression

As in the existing popular Bayesian optimization methods, we use Gaussian process surrogates to learn the unknown functions. As in [22], we artificially introduce a Gaussian process $\mathcal{GP}(0, k_0(\cdot, \cdot))$ for the surrogate modelling of the unknown black-box function f. We also adopt an i.i.d Gaussian zero-mean noise model with noise variance $\lambda > 0$, which can be chosen by the algorithm.

We use $X_t := \{x_1, x_2, \cdots, x_t\}$ to denote the sample sequence up to step t (Recall that x_t is defined as (θ_t, z_t)). And we introduce the following functions of (x, x'),

$$\mu_{0,t}(x) = k_0(x_{1:t}, x)^\top (K_{0,t} + \lambda I)^{-1} y_{0,1:t}, \tag{4a}$$

$$k_{0,t}(x, x') = k_0(x, x') - k_0(x_{1:t}, x)^\top (K_{0,t} + \lambda I)^{-1} k_0(x_{1:t}, x'), \tag{4b}$$

$$\sigma^2_{0,t}(x) = k_{0,t}(x, x) \tag{4c}$$

with $k_0(x_{1:t}, x) = [k_0(x_1, x), k_0(x_2, x), \cdots, k_0(x_t, x)]^\top$, $K_{0,t} = (k_0(x, x'))_{x,x' \in X_t}$, $y_{0,1:t} = [y_{0,1}, y_{0,2}, \cdots, y_{0,t}]^\top$. Similarly, we can get $\mu_{i,t}(\cdot), k_{i,t}(\cdot,\cdot), \sigma_{i,t}(\cdot)$, $\forall i \in [N]$ for the constraints.

To characterize the complexity of the Gaussian processes and the RKHSs corresponding to the kernel functions, we further introduce the maximum information gain for the objective f as in [21],

$$\gamma_{0,t} := \max_{A \subset \Theta \times \mathcal{Z}; |A| = t} \frac{1}{2} \log \left| I + \lambda^{-1} K_{0,A} \right|, \tag{5}$$

where $K_{0,A} = (k_0(x, x'))_{x,x' \in A}$. Similarly, we introduce $\gamma_{i,t}, \forall i \in [N]$ for the constraints.

Remark 1 *Note that the Gaussian process model above is* only *used to derive posterior mean functions, covariance functions, and maximum information gain for algorithm design and theoretical analysis. It does not change our set-up that f is a deterministic function and that the observation noise only needs to be sub-Gaussian.*

Based on the aforementioned preliminaries of Gaussian process regression, we then derive the lower confidence, and upper confidence bound functions.

Lemma 1 *Let Assumptions 1 and 2 hold. With probability at least $1 - \delta, \forall \delta \in (0, 1)$, the following holds for all $x \in \mathcal{X}$ and $t \geq 1$,*

$$g_i(x) \in [\underline{g}_{i,t}(x), \bar{g}_{i,t}(x)], \ \forall i \in [N] \tag{6}$$

$$\text{and} \ \ f(x) \in [\underline{f}_t(x), \bar{f}_t(x)], \tag{7}$$

where for all $i \in [N]$,

$$\underline{f}_t(x) := \max\{\mu_{0,t-1}(x) - \beta^{1/2}_{0,t} \sigma_{0,t-1}(x), -C_0\}, \tag{8a}$$

$$\bar{f}_t(x) := \min\{\mu_{0,t-1}(x) + \beta^{1/2}_{0,t} \sigma_{0,t-1}(x), C_0\}, \tag{8b}$$

$$\underline{g}_{i,t}(x) := \max\{\mu_{i,t-1}(x) - \beta^{1/2}_{i,t} \sigma_{i,t-1}(x), -C_i\}, \tag{8c}$$

$$\bar{g}_{i,t}(x) := \min\{\mu_{i,t-1}(x) + \beta^{1/2}_{i,t} \sigma_{i,t-1}(x), C_i\}, \tag{8d}$$

$$\beta^{1/2}_{i,t} := C_i + \sigma \sqrt{2\left(\gamma_{i,t-1} + 1 + \ln((N+1)/\delta)\right)}. \tag{8e}$$

Proof: By Corollary 2.1, [23] and that $g_i(x) \in [-C_i, C_i], f(x) \in [-C_0, C_0]$ holds for sure due to $|g_i(x)| = |\langle g_i, k_i(x, \cdot)\rangle| \leq \|g_i\| \|k_i(x, \cdot)\| \leq C_i$ (similar for $f(x)$). ■

Without further notice, all the following results are conditioned on the high probability event in Lem. 1 happening.

V. Algorithm and Theoretical Guarantees

Our primal-dual algorithm is shown in Alg. 1, where $0 < \epsilon \leq \frac{\xi}{2}$ is a slackness parameter and $[\cdot]^+$ is interpreted element-wise.

Algorithm 1 **P**rimal-**D**ual **C**ontextual **B**ayesian **O**ptimization (PDCBO).

1: **for** $t \in [T]$ **do**
2: Observe contextual variables z_t.
3: **Primal update:**

$$\theta_t = \arg\min_{\theta \in \Theta} \{\underline{f}_t(\theta, z_t) + \eta \lambda_t^T \underline{g}_t(\theta, z_t)\}. \tag{9}$$

4: **Dual update:**

$$\lambda_{t+1} = [\lambda_t + \underline{g}_t(\theta_t, z_t) + \epsilon e]^+. \tag{10}$$

5: Evaluate f and $g_i, i \in [N]$ at $x_t = (\theta_t, z_t)$ with noise.
6: Update $(\mu_{i,t}, \sigma_{i,t}), i \in \{0\} \cup [N]$ with the new data.
7: **end for**

The Alg. 1 is based on the well-known primal-dual optimization method. The Lagrangian of the original problem is $\mathcal{L}(\theta, z, \phi) = f(\theta, z) + \phi^T g(\theta, z)$ and the dual function is $\mathcal{D}(z, \phi) = \min_\theta \mathcal{L}(\theta, z, \phi)$. In the primal update step, we optimize the Lagrangian. In the dual update step, we apply the dual ascent method. In primal and dual updates, we replace the unknown black-box functions with their lower confidence bounds, following the principle of *optimism in the face of uncertainty*.

The primal update (9) requires solving a potentially non-convex auxiliary optimization problem. When the dimension of θ is small (e.g., < 6) we can use a pure grid search method to solve it. For larger dimensions, one can, for example, apply the gradient descent method from multiple different initial points.

Remark 1 *In Alg. 1, λ_t is not exactly the dual variable, but the dual variable scaled by $\frac{1}{\eta}$. Indeed, λ_t can be interpreted as virtual queue length [18]. The intuition of ϵ is to introduce some level of constant pessimistic drift to control the cumulative violation.*

We now give the theoretical guarantees in Thm. 1.

Theorem 1 *Let the Assumptions 1, 2, 3, 4 and 5 hold. We further assume the maximum information gain term satisfies $\lim_{T\to\infty} \frac{\|\gamma^g_T\|}{\sqrt{T}} = 0$, where $\gamma^g_T := (\gamma_{1,T}, \cdots, \gamma_{N,T})$. We set $\eta = \frac{1}{\sqrt{T}}$, $\lambda_1 = (\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi})e$ and*

$$\epsilon = \frac{\sqrt{N\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2 + \frac{4C_0}{\eta} + 4\|C\|^2} + 8\|\beta^g_T\|\sqrt{T\|\gamma^g_T\|}}{T},$$

where $\beta^g_T := (\beta_{1,T}, \cdots, \beta_{N,T})$ and $C := (C_1, \cdots, C_N)$. Let T be large enough such that $\epsilon = \mathcal{O}(\|\gamma^g_T\|/\sqrt{T}) \leq \xi/2$. We have

$$\mathcal{R}_T = \mathcal{O}((\gamma_{0,T} + \|\gamma^g_T\|)\sqrt{T}) \ \text{ and } \ \mathcal{V}_T \leq 0.$$

Proof Sketch Our proof involves three key ingredients. First, the cumulative cost paid to learn the black-box functions can be bounded by some term involving the maximum information gain (5). Second, the regret and the violations in the primal values can be bounded by the drift of dual variables. Third, by properly selecting the constant dual drift ϵ, we can balance the regret and violation such that *time-average* constraints are strictly satisfied while the cumulative *contextual regret* is bounded. The detailed proofs can be found in the Appendix.

Remark 2 *Thm. 1 highlights that for kernels with the maximum information gain terms that grow slower than $\sqrt{T}$, our algorithm achieves average contextual regret converging to zero while incurring no constraint violation on average.*

By applying the bounds on maximum information gains from [21], [24] to Thm. 1, we derive the kernel-specific bounds as in Tab. I, where $d = n_\theta + n_z$, ν is the smoothness parameter of the Mátern kernel, and $\tilde{\mathcal{O}}(\cdot)$ hides some polylogarithmic term.

TABLE I

KERNEL-SPECIFIC REGRET BOUNDS.

Kernel	Linear	Squared Exponential	Mátern ($\frac{\nu}{d} > \frac{1}{2}$)
$\mathcal{R}_T$	$\tilde{\mathcal{O}}(d\sqrt{T})$	$\tilde{\mathcal{O}}(\sqrt{T})$	$\tilde{\mathcal{O}}(T^{\frac{2\nu+3d}{4\nu+2d}})$

Remark 3 *Thm. 1 focuses on the case that the running steps T is known beforehand. To extend to the case where T is unknown, we can apply the doubling trick [25]. The basic idea is to start the algorithm with a small T and restart it with T doubled every time the running steps are exhausted.*

VI. EXPERIMENTS

We consider two sets of experiments. In the first set, we use the objective and constraint functions sampled from Gaussian processes. In the second set, we consider the classical Williams-Otto benchmark problem [26]. We compare our method to the state-of-the-art Bayesian optimization methods with constraints consideration, including SafeOPT [6] and CEI (**C**onstrained **E**xpected **I**mprovement) method [4], [5]. The experiments are implemented in python, based on the package GPy [27].

Choice of Parameters. The performance of PDCBO is mainly impacted by the choice of $\beta_{i,t}^{1/2}$. In practice, $\beta_{i,t}^{1/2}$ can usually be set as a constant. In our experiments, manually setting $\beta_{i,t}^{1/2} = 1.0$ works well. We also set $\lambda = 0.05^2$ as the noise variance for the Gaussian process modelling. We use the common squared exponential kernel functions. For the second example, we sample a few points randomly and maximize the likelihood function to get the hyperparameters of the kernel.

Computational Time. In our experiments, all the problems have low-dimensional inputs ($n_\theta \leq 3$). So we use pure grid search to solve the auxiliary problem for primal update, which is relatively cheap due to the explicit expressions of the lower confidence bound functions as compared to the simulation of the Williams-Otto benchmark problem.

A. Sampled Instances from Gaussian Process

We consider the problem (1) with only one constraint, where $\theta \in \Theta = [-10, 10]$, $z \in \mathcal{Z} = [-10, 10]$, and both f and g are unknown black-box functions sampled from a Gaussian process. We use the squared exponential kernel,

$$k((\theta_1, z_1), (\theta_2, z_2)) = \sigma_{\text{SE}}^2 \exp\left\{-\left(\frac{\theta_1 - \theta_2}{l_\theta}\right)^2 - \left(\frac{z_1 - z_2}{l_z}\right)^2\right\}, \tag{11}$$

where $\sigma_{\text{SE}}^2 = 2.0$, and $l_\theta = l_z = 1.0$. We run different

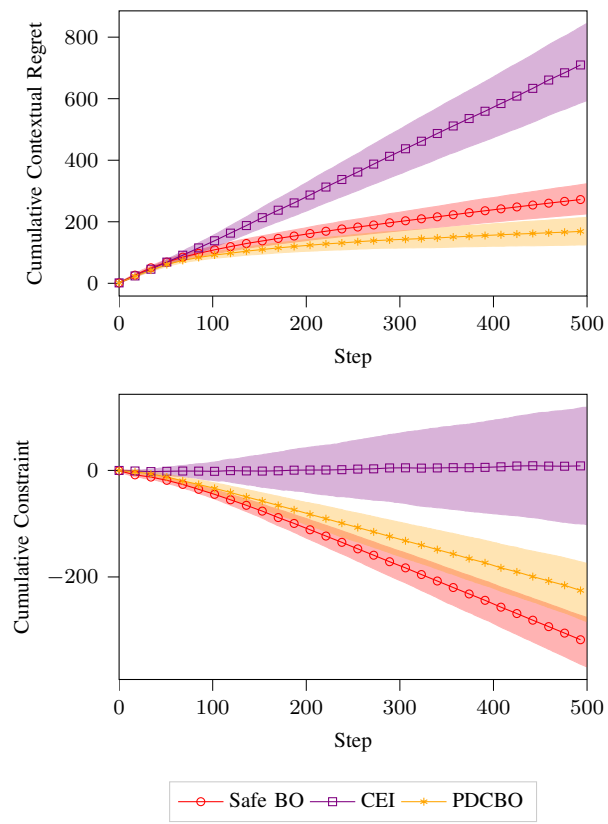

Fig. 1. The cumulative context regret and cumulative constraint value of different algorithms. The statistics are obtained over 50 different trajectories. The shaded area represents ± 0.5 standard deviation.

algorithms for 500 steps over 50 different sampled instances. The context sequence is randomly generated from a uniform distribution over the context set $\mathcal{Z}$. Fig. 1 shows the cumulative contextual regret and cumulative constraint value of different algorithms over 50 different trajectories. It can be observed that our PDCBO achieves sublinear cumulative contextual regret while keeping the cumulative constraint value below zero with high confidence. In contrast, the CEI method suffers from almost linear cumulative contextual regret and fails to keep average constraint feasibility with high probability. Although Safe BO can keep average constraint feasibility, its cumulative contextual regret is about 62%

higher than our PDCBO due to its being overly cautious to keep feasibility in every single step.

B. Williams-Otto Classical Problem

In this section, we consider the problem of the classical Williams-Otto benchmark problem [26]. In this problem, we feed a continuous stirred-tank reactor (CSTR) with two components A and B. The reactor then operates at the steady state under the reaction temperature T_{r}. During the chemical reaction, a byproduct G is produced. We use X_{A} and X_{G} to denote the steady-state residual mass fractions of A and G at the reactor outlet. Both X_{A} and X_{G} need to be managed. That is, we need to control the steady-state residual mass fractions of A and G to be lower than some thresholds. See [26, Sec. 4.1] for more details. Meanwhile, a key factor that impacts the cumulative profit of the chemical reaction process is the time-varying nature of the product price and the costs of the raw materials [28]. To incorporate these time-varying factors, we model them as contextual variables.

We want to adaptively tune the feed rate F_{B} of the component B and the reaction temperature T_{r} to maximize the cumulative economic profit from the reaction while managing X_{A} and X_{G}. Our problem can be formulated as,

$$\begin{aligned} \min_{F_{\mathrm{B}}, T_{\mathrm{r}}} \quad & J(F_{\mathrm{B}}, T_{\mathrm{r}}, P) \\ \text{subject to} \quad & \text{CSTR model [29]} \\ & g_1(F_{\mathrm{B}}, T_{\mathrm{r}}) := X_{\mathrm{A}}(F_{\mathrm{B}}, T_{\mathrm{r}}) - 0.12 \leq 0 \\ & g_2(F_{\mathrm{B}}, T_{\mathrm{r}}) := X_{\mathrm{G}}(F_{\mathrm{B}}, T_{\mathrm{r}}) - 0.08 \leq 0, \\ & F_{\mathrm{B}} \in [4, 7], T_{\mathrm{r}} \in [70, 100] \end{aligned} \tag{12}$$

where $J(F_{\mathrm{B}}, T_{\mathrm{r}}, P)$ is the minimization objective that is opposite to the net economic profit, $P \in \mathbb{R}^4$ is the price vector of the product and the raw materials, $g_1(F_{\mathrm{B}}, T_{\mathrm{r}})$ and $g_2(F_{\mathrm{B}}, T_{\mathrm{r}})$ are threshold constraints on the residual mass fractions. To generate the context P^t at step t, we randomly sample P_i^t uniformly from the interval $[(1-\eta)\bar{P}_i, (1+\eta)\bar{P}_i]$, where $\bar{P}_i$ is a predefined normal price and $\eta = 0.2$.

Fig. 2 shows the cumulative cost and the running average of the constraints with 50 different context trajectories. We observe that our PDCBO achieves the lowest cumulative cost (the highest cumulative profit, equivalently) while satisfying the constraints on average. In contrast, Safe BO is overly cautious and can not achieve all the potential economic profit. Finally, the generic CEI method suffers from both low profit and severe constraint violation for the second constraint.

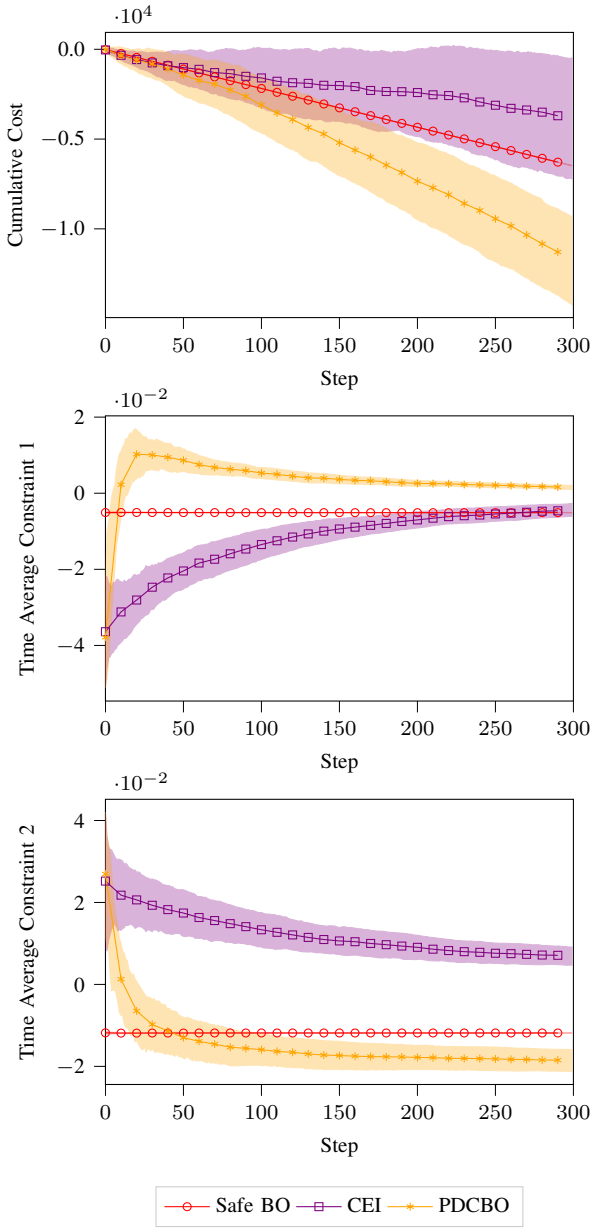

Fig. 2. Cumulative cost and average constraints of the chemical reaction over 50 different context trajectories. The shaded area represents ± 1.0 standard deviation.

VII. Conclusion

This paper studied the problem of optimizing the closed-loop performance of constrained control systems, where both the objective and the constraints are unknown black-box functions affected by exogenous time-varying contextual disturbances. A primal-dual contextual Bayesian optimization algorithm was proposed, and it was shown to achieve sub-linear cumulative regret with respect to the dynamic optimal solution under certain regularity conditions. Furthermore, the algorithm achieves zero time-average constraint violation, ensuring that the average value of the constraint function satisfies the desired constraint. Our method is particularly useful when the constraint functions correspond to some economic costs that accumulate with time. The method is applied to both sampled instances from Gaussian processes and to a continuous stirred tank reactor feedrate and reaction temperature tuning problem; simulation results on both problems show that the method simultaneously provides close-to-optimal performance and maintains constraint feasibility on average. This contrasts current state-of-the-art methods, which either suffer from large cumulative regret or severe constraint violations for the case studies presented.

REFERENCES

[1] W. Xu, C. N. Jones, B. Svetozarevic, C. R. Laughman, and A. Chakrabarty, "VABO: Violation-Aware Bayesian Optimization for closed-loop control performance optimization with unmodeled constraints," in *2022 American Control Conference (ACC)*. IEEE, 2022, pp. 5288–5293.

[2] W. Xu, Y. Jiang, E. T. Maddalena, and C. N. Jones, "Lower bounds on the worst-case complexity of efficient global optimization," *arXiv preprint arXiv:2209.09655*, 2022.

[3] C. K. Williams and C. E. Rasmussen, *Gaussian processes for machine learning*. MIT press Cambridge, MA, 2006, vol. 2, no. 3.

[4] M. A. Gelbart, J. Snoek, and R. P. Adams, "Bayesian optimization with unknown constraints," in *Proc. of the 30th Conf. on Uncertainty in Artif. Intell.*, ser. UAI'14. Arlington, Virginia, USA: AUAI Press, 2014, p. 250–259.

[5] J. R. Gardner, M. J. Kusner, Z. E. Xu, K. Q. Weinberger, and J. P. Cunningham, "Bayesian optimization with inequality constraints." in *ICML*, vol. 2014, 2014, pp. 937–945.

[6] Y. Sui, A. Gotovos, J. Burdick, and A. Krause, "Safe exploration for optimization with Gaussian processes," in *Proc. of the Int. Conf. on Mach. Learn.*, 2015, pp. 997–1005.

[7] Y. Sui, J. Burdick, Y. Yue *et al.*, "Stage-wise safe Bayesian optimization with Gaussian processes," in *Proc. of the Int. Conf. on Mach. Learn.*, 2018, pp. 4781–4789.

[8] D. Bergmann and K. Graichen, "Safe Bayesian optimization under unknown constraints," in *2020 59th IEEE Conference on Decision and Control (CDC)*. IEEE, 2020, pp. 3592–3597.

[9] A. Rupenyan, M. Khosravi, and J. Lygeros, "Performance-based trajectory optimization for path following control using Bayesian optimization," in *2021 60th IEEE Conference on Decision and Control (CDC)*. IEEE, 2021, pp. 2116–2121.

[10] M. Fiducioso, S. Curi, B. Schumacher, M. Gwerder, and A. Krause, "Safe contextual Bayesian optimization for sustainable room temperature PID control tuning," *arXiv preprint arXiv:1906.12086*, 2019.

[11] P. Brunzema, A. Von Rohr, and S. Trimpe, "On controller tuning with time-varying Bayesian optimization," in *2022 IEEE 61st Conference on Decision and Control (CDC)*. IEEE, 2022, pp. 4046–4052.

[12] I. Bogunovic, J. Scarlett, and V. Cevher, "Time-varying Gaussian process bandit optimization," in *Artificial Intelligence and Statistics*. PMLR, 2016, pp. 314–323.

[13] W. Xu, C. N. Jones, B. Svetozarevic, C. R. Laughman, and A. Chakrabarty, "Violation-aware contextual Bayesian optimization for controller performance optimization with unmodeled constraints," *arXiv preprint arXiv:2301.12099*, 2023.

[14] J. Fleming and M. Cannon, "Time-average constraints in stochastic Model Predictive Control," in *2017 American Control Conference (ACC)*. IEEE, 2017, pp. 5648–5653.

[15] M. Korda, R. Gondhalekar, F. Oldewurtel, and C. N. Jones, "Stochastic MPC framework for controlling the average constraint violation," *IEEE Transactions on Automatic Control*, vol. 59, no. 7, pp. 1706–1721, 2014.

[16] M. Cannon, B. Kouvaritakis, and X. Wu, "Probabilistic constrained MPC for multiplicative and additive stochastic uncertainty," *IEEE Transactions on Automatic Control*, vol. 54, no. 7, pp. 1626–1632, 2009.

[17] J. Kim, M. Ruggiero, and D. Atienza, "Free cooling-aware dynamic power management for green datacenters," in *2012 International Conference on High Performance Computing & Simulation (HPCS)*. IEEE, 2012, pp. 140–146.

[18] X. Zhou and B. Ji, "On kernelized multi-armed bandits with constraints," in *Advances in Neural Information Processing Systems*, 2022.

[19] A. Krause and C. Ong, "Contextual Gaussian process bandit optimization," *Advances in Neural Information Processing Systems*, vol. 24, 2011.

[20] B. Schölkopf, R. Herbrich, and A. J. Smola, "A generalized representer theorem," in *International conference on computational learning theory*. Springer, 2001, pp. 416–426.

[21] N. Srinivas, A. Krause, S. M. Kakade, and M. W. Seeger, "Information-theoretic regret bounds for Gaussian process optimization in the bandit setting," *IEEE Transactions on Information Theory*, vol. 58, no. 5, pp. 3250–3265, 2012.

[22] S. R. Chowdhury and A. Gopalan, "On kernelized multi-armed bandits," in *International Conference on Machine Learning*. PMLR, 2017, pp. 844–853.

[23] W. Xu, Y. Jiang, and C. N. Jones, "Constrained efficient global optimization of expensive black-box functions," *arXiv preprint arXiv:2211.00162*, 2022.

[24] S. Vakili, K. Khezeli, and V. Picheny, "On information gain and regret bounds in Gaussian process bandits," in *International Conference on Artificial Intelligence and Statistics*. PMLR, 2021, pp. 82–90.

[25] L. Besson and E. Kaufmann, "What doubling tricks can and can't do for multi-armed bandits," *arXiv preprint arXiv:1803.06971*, 2018.

[26] E. A. del Rio Chanona, P. Petsagkourakis, E. Bradford, J. A. Graciano, and B. Chachuat, "Real-time optimization meets Bayesian optimization and derivative-free optimization: A tale of modifier adaptation," *Comput. & Chem. Eng.*, vol. 147, p. 107249, 2021.

[27] GPy, "GPy: A Gaussian process framework in python," http://github.com/SheffieldML/GPy, since 2012.

[28] R. Yadbantung and P. Bumroongsri, "Periodically time-varying economic model predictive control with applications to nonlinear continuous stirred tank reactors," *Computers & Chemical Engineering*, vol. 157, p. 107602, 2022.

[29] D. F. Mendoza, J. E. A. Graciano, F. dos Santos Liporace, and G. A. C. Le Roux, "Assessing the reliability of different real-time optimization methodologies," *The Can. J. of Chem. Eng.*, vol. 94, no. 3, pp. 485–497, 2016.

[30] S. R. Chowdhury and A. Gopalan, "On kernelized multi-armed bandits," *arXiv preprint arXiv:1704.00445*, 2017.

APPENDIX

We give the detailed proof of the Thm. 1. Note that all the results are under the same assumptions as the Thm. 1 without further notice. The learning cost comes from the uncertainties of the unknown functions, measured by the posterior standard deviations, and can be bounded as in Lem. 2.

Lemma 2 (Lemma 4, [30]) *Given a sequence of points $x_1, x_2, \cdots, x_T$ from $\Theta \times \mathcal{Z}$, we have,*

$$\sum_{t=1}^{T} \sigma_{i,t-1}(x_t) \leq \sqrt{4(T+2)\gamma_{i,T}}. \tag{13}$$

To connect the drift of the dual variables and the primal values, we introduce a function in the dual space,

$$V(\lambda_t) = \frac{1}{2}\|\lambda_t\|^2. \tag{14}$$

We consider,

$$\begin{aligned}
\Delta_t :=& V(\lambda_{t+1}) - V(\lambda_t) && \text{(15a)}\\
=& \frac{1}{2}\left(\|[\lambda_t + \underline{g}_t(\theta_t, z_t) + \epsilon e]^+\|^2 - \|\lambda_t\|^2\right) && \text{(15b)}\\
\leq& \frac{1}{2}\left(\|\lambda_t + \underline{g}_t(\theta_t, z_t) + \epsilon e\|^2 - \|\lambda_t\|^2\right) && \text{(15c)}\\
=& \lambda_t^T(\underline{g}_t(\theta_t, z_t) + \epsilon e) + \frac{1}{2}\|\underline{g}_t(\theta_t, z_t) + \epsilon e\|^2, && \text{(15d)}
\end{aligned}$$

where the inequality follows by discussion on the sign of $\lambda_t + \underline{g}_t(\theta_t, z_t) + \epsilon e$. To characterize the tradeoff between feasibility and optimality, we introduce the perturbed problem,

$$\begin{aligned}
\min_{\pi \in \Pi(\Theta)} \quad & \mathbb{E}_\pi(f(\theta, z)), && \text{(16a)}\\
\text{subject to:} \quad & \mathbb{E}_\pi(g(\theta, z)) + \epsilon e \leq 0, \quad \forall i \in [N], && \text{(16b)}
\end{aligned}$$

where the feasible set is relaxed to the set of all distributions over the set Θ. Such a relaxation results in a linear programming problem in distribution, which is easier for sensitivity analysis. We use $\pi_\epsilon^*(z)$ to denote the optimal solution to the above problem. We then have the following lemma.

Lemma 3

$$\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(f(\theta, z_t)) - \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi^*(z_t)}(f(\theta, z_t)) \leq \frac{2C_0 T\epsilon}{\xi}, \quad (17)$$

where $\pi^*(z_t)$ is the optimal distribution for Problem (16) *with $\epsilon = 0$.*

Proof: Let $\pi_\epsilon(z_t) = (1 - \frac{\epsilon}{\xi})\pi^*(z_t) + \frac{\epsilon}{\xi}\bar{\pi}(z_t)$, where $\mathbb{E}_{\bar{\pi}(z_t)}(g(\theta, z_t)) \leq -\xi e$ (Recall the Proposition 1). Then

$$\mathop{\mathbb{E}}_{\pi_\epsilon(z_t)}(g(\theta, z_t)) = (1 - \frac{\epsilon}{\xi}) \mathop{\mathbb{E}}_{\pi^*(z_t)}(g(\theta, z_t)) + \frac{\epsilon}{\xi} \mathop{\mathbb{E}}_{\bar{\pi}(z_t)}(g(\theta, z_t)) \leq -\epsilon e.$$

Hence, $\pi_\epsilon(z_t)$ is a feasible solution to the slightly perturbed problem. So, we have,

$$\begin{aligned}
&\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi_\epsilon^*(z)}(f(\theta, z_t)) - \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi^*(z)}(f(\theta, z_t)) \\
\leq &\sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi_\epsilon(z)}(f(\theta, z_t)) - \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi^*(z)}(f(\theta, z_t)) \\
= &\frac{\epsilon}{\xi} \sum_{t=1}^{T} \left(\mathop{\mathbb{E}}_{\bar{\pi}(z_t)}(f(\theta, z_t)) - \mathop{\mathbb{E}}_{\pi^*(z_t)}(f(\theta, z_t)) \right) \\
\leq &\frac{2C_0 T\epsilon}{\xi},
\end{aligned}$$

where the first inequality follows by the optimality of π_ϵ^*, the equality follows by the definition of π_ϵ and the last inequality follows by Assumption 2, which implies that $|f(x)| = |\langle f, k(x, \cdot)\rangle| \leq \|f\|\|k(x, \cdot)\| \leq C_0$. ■

It will be seen that ϵ plays a key role in trading some regret for strict time-average feasibility.

1) Bound Cumulative Regret: We have the following lemma to bound $\underline{f}_t(\theta_t, z_t) - \mathbb{E}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t))$, which approximates the single-step regret.

Lemma 4

$$\underline{f}_t(\theta_t, z_t) - \mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t)) \leq 2\eta\|C\|^2 - \eta\Delta_t,$$

where $\|C\|^2 = \sum_{i=1}^{N} C_i^2$ and ϵ is set to be small enough such that $\epsilon \leq C_i, \forall i \in [N]$.

Proof:

$$\begin{aligned}
\Delta_t &= V(\lambda_{t+1}) - V(\lambda_t) \\
&\leq \lambda_t^T(\underline{g}_t(\theta_t, z_t) + \epsilon e) + \frac{1}{2}\|\underline{g}_t(\theta_t, z_t) + \epsilon e\|^2 \\
&\leq \lambda_t^T \underline{g}_t(\theta_t, z_t) + \frac{1}{\eta}\underline{f}_t(\theta_t, z_t) + \epsilon\lambda_t^T e - \frac{1}{\eta}\underline{f}_t(\theta_t, z_t) \\
&\quad + \frac{1}{2}\sum_{i=1}^{N}(C_i + \epsilon)^2 \\
&\leq \lambda_t^T \mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(\underline{g}_t(\theta, z_t)) + \frac{1}{\eta} \mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t)) + \epsilon\lambda_t^T e \\
&\quad - \frac{1}{\eta}\underline{f}_t(\theta_t, z_t) + 2\sum_{i=1}^{N} C_i^2 \\
&\leq \frac{1}{\eta}(\mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t)) - \underline{f}_t(\theta_t, z_t)) + 2\|C\|^2,
\end{aligned}$$

where the first inequality follows by the inequality (15), the second inequality follows by adding and subtracting $\frac{1}{\eta}\underline{f}_t(\theta_t, z_t)$ and the projection operation to $[-C_i, C_i]$ as shown in (8), the third inequality follows by the optimality of θ_t for the primal update problem (9) and the assumption that $\epsilon \leq C_i$, and the last inequality follows by the feasibility of π_ϵ^* for the problem (16). Rearrangement of the above inequality gives the desired result. ■

We are then ready to upper bound the cumulative regret.

Lemma 5 (Cumulative Regret Bound)

$$\mathcal{R}_T \leq 2\beta_{0,T}^{1/2}\sqrt{4(T+2)\gamma_T} + 2\eta T\|C\|^2 + \eta V(\lambda_1) + \frac{2C_0 T\epsilon}{\xi}.$$

Proof:

$$\begin{aligned}
\mathcal{R}_T \leq &\sum_{t=1}^{T} \left(f(\theta_t, z_t) - \mathbb{E}_{\pi^*(z_t)}(f(\theta, z_t))\right) \\
= &\sum_{t=1}^{T} \left(f(\theta_t, z_t) - \underline{f}_t(\theta_t, z_t)\right) \\
&+ \sum_{t=1}^{T} \left(\underline{f}_t(\theta_t, z_t) - \mathbb{E}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t))\right) \\
&+ \sum_{t=1}^{T} \mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t) - f(\theta, z_t)) \\
&+ \sum_{t=1}^{T} \left(\mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(f(\theta, z_t)) - \mathop{\mathbb{E}}_{\pi^*(z_t)}(f(\theta, z_t)) \right),
\end{aligned}$$

where the last inequality follows by that relaxed optimal value $\mathbb{E}_{\pi^*(z_t)}(f(\theta, z_t))$ is smaller or equal to the original optimal value, and the equality splits the original term into four terms. For the first term,

$$\begin{aligned}
&\sum_{t=1}^{T} \left(f(\theta_t, z_t) - \underline{f}_t(\theta_t, z_t)\right) \leq \sum_{t=1}^{T} 2\beta_{0,t}^{1/2}\sigma_t(\theta_t, z_t) \\
\leq &2\beta_{0,T}^{1/2}\sum_{t=1}^{T}\sigma_t(\theta_t, z_t) \leq 2\beta_{0,T}^{1/2}\sqrt{4(T+2)\gamma_T},
\end{aligned}$$

where the first inequality follows by Lem. 1, the second inequality follows by the monotonicity of $\beta_{0,t}^{1/2}$, and the last inequality follows by Lem. 2. For the second term, by Lem. 4, we have,

$$\begin{aligned}
&\sum_{t=1}^{T} \left(\underline{f}_t(\theta_t, z_t) - \mathop{\mathbb{E}}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t)) \right) \leq \sum_{t=1}^{T}(2\eta\|C\|^2 - \eta\Delta_t) \\
&= 2\eta T\|C\|^2 + \eta V(\lambda_1) - \eta V(\lambda_{T+1}) \leq 2\eta T\|C\|^2 + \eta V(\lambda_1).
\end{aligned}$$

The third term is non-positive due to Lem. 1. Combining the three bounds and the Lem. 3 gives the desired result. ■

2) Bound Cumulative Violation: The dual update indicates that violations are reflected in the dual variable. So we first upper bound the dual variable. The idea is to show that whenever the dual variable is very large, it will be decreased.

Lemma 6 *If* $V(\lambda_t) \geq \frac{N}{2}\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2$ *and* $0 < \epsilon \leq \frac{\xi}{2}$, *we have* $V(\lambda_{t+1}) \leq V(\lambda_t)$.

Proof: By the primal updating rule,

$$\begin{aligned}
&\underline{f}_t(\theta_t, z_t) + \eta\lambda_t^T \underline{g}_t(\theta_t, z_t) \\
\leq &\underline{f}_t(\bar{\theta}(z_t), z_t) + \eta\lambda_t^T \underline{g}_t(\bar{\theta}(z_t), z_t) \\
\leq &\underline{f}_t(\bar{\theta}(z_t), z_t) + \eta\lambda_t^T g(\bar{\theta}(z_t), z_t) \\
\leq &C_0 + \eta(-\xi)\lambda_t^T e,
\end{aligned}$$

where the first inequality follows by the optimality of θ_t for the primal update problem, the second inequality follows by that both η and λ_t are non-negative, and the third inequality follows by Lem. 1 and Assumption 5. On the other hand,

$$\underline{f}_t(\theta_t, z_t) + \eta\lambda_t^T \underline{g}_t(\theta_t, z_t) \geq -C_0 + \eta\lambda_t^T \underline{g}_t(\theta_t, z_t),$$

by the Lem. 1. Therefore,

$$C_0 + \eta(-\xi)\lambda_t^T e \geq -C_0 + \eta\lambda_t^T \underline{g}_t(\theta_t, z_t),$$

which implies

$$\lambda_t^T \underline{g}_t(\theta_t, z_t) \leq \frac{2C_0}{\eta} - \xi\lambda_t^T e.$$

So we can get

$$\begin{aligned}
&V(\lambda_{t+1}) - V(\lambda_t) \\
\leq &\lambda_t^T(\underline{g}_t(\theta_t, z_t) + \epsilon e) + \frac{1}{2}\|\underline{g}_t(\theta_t, z_t) + \epsilon e\|^2 \\
\leq &\frac{2C_0}{\eta} - \frac{\xi}{2}\lambda_t^T e + \frac{1}{2}\|\underline{g}_t(\theta_t, z_t) + \epsilon e\|^2 \\
\leq &\frac{2C_0}{\eta} - \frac{\xi}{2}\sqrt{\frac{\|\lambda_t\|^2}{N}} + 2\|C\|^2 \leq 0,
\end{aligned}$$

where the first inequality follows by the inequality (15), the second inequality follows by that $\epsilon \leq \frac{\xi}{2}$, the third inequality follows by that $\lambda_t \geq 0$ and the Lem. 1, and the last inequality follows by $V(\lambda_t) \geq \frac{N}{2}\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2$. ■

Consequently, we have,

Lemma 7 *Let* $\lambda_1 \leq (\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi})e$, *we have for any* t,

$$V(\lambda_t) \leq C_V(\eta),$$

where $C_V(\eta) = \frac{N}{2}\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2 + \frac{2C_0}{\eta} + 2\|C\|^2$.

Proof: We use induction. λ_1 satisfies the conclusion. We now discuss conditioned on the value λ_t.

Case 1: $V(\lambda_t) \leq \frac{N}{2}\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2$, then $V(\lambda_{t+1}) \leq V(\lambda_t) + \frac{1}{\eta}(\mathbb{E}_{\pi_\epsilon^*(z_t)}(\underline{f}_t(\theta, z_t)) - \underline{f}_t(\theta_t, z_t)) + 2\|C\|^2 \leq \frac{N}{2}\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2 + \frac{2C_0}{\eta} + 2\|C\|^2$.

Case 2: $V(\lambda_t) \geq \frac{N}{2}\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2$, then $V(\lambda_{t+1}) \leq V(\lambda_t) \leq \frac{N}{2}\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2 + \frac{2C_0}{\eta} + 2\|C\|^2$.

By induction, the conclusion holds for any t. ■

We now can upper bound the cumulative violation.

Lemma 8 (Cumulative Violation Bound)

$$\mathcal{V}_T \leq \left\|\left[\lambda_{T+1} + 2\beta_T^g\sqrt{4(T+2)\gamma_T^g} - T\epsilon e\right]^+\right\|,$$

where $\beta_T^g = (\beta_{1,T}, \cdots, \beta_{N,T})$, $\gamma_T^g = (\gamma_{1,T}, \cdots, \gamma_{N,T})$ *and multiplication is interpreted elementwise.*

Proof: By the dual updating rule, we have $\lambda_{t+1} \geq \lambda_t + \underline{g}_t(\theta_t, z_t) + \epsilon$. By summing up from $t = 1$ to T, we get,

$$\lambda_{T+1} \geq \lambda_1 + \sum_{t=1}^T \underline{g}_t(\theta_t, z_t) + T\epsilon e.$$

Rearranging the above inequality gives,

$$\sum_{t=1}^T \underline{g}_t(\theta_t, z_t) \leq \lambda_{T+1} - \lambda_1 - T\epsilon e. \tag{18}$$

We thus have,

$$\begin{aligned}
\sum_{t=1}^T g(\theta_t, z_t) = &\sum_{t=1}^T \underline{g}_t(\theta_t, z_t) + \sum_{t=1}^T (g(\theta_t, z_t) - \underline{g}_t(\theta_t, z_t)) \\
\leq &\lambda_{T+1} - \lambda_1 - T\epsilon e + 2\beta_T^g\sqrt{4(T+2)\gamma_T^g},
\end{aligned}$$

where the inequality follows by combining the inequality (18), the monotonicity of β_T^g and Lem. 2.. Therefore,

$$\begin{aligned}
\mathcal{V}_T = &\left\|\left[\sum_{t=1}^T g(\theta_t, z_t)\right]^+\right\| \\
\leq &\left\|\left[\lambda_{T+1} + 2\beta_T^g\sqrt{4(T+2)\gamma_T^g} - T\epsilon e\right]^+\right\|.
\end{aligned}$$

■

A. Proof of Thm. 1

$$\lambda_{T+1} + 2\beta_T^g\sqrt{4(T+2)\gamma_T^g} \tag{19a}$$

$$\leq \left(\|\lambda_{T+1}\| + 8\|\beta_T^g\|\sqrt{T\|\gamma_T^g\|}\right)e \tag{19b}$$

$$\leq \left(\sqrt{N\left(\frac{4C_0}{\eta\xi} + \frac{4\|C\|^2}{\xi}\right)^2 + \frac{4C_0}{\eta} + 4\|C\|^2} + 8\|\beta_T^g\|\sqrt{T\|\gamma_T^g\|}\right)e = T\epsilon e \tag{19c}$$

where the first inequality follows by simple algebraic manipulation and the second inequality follows by Lem. 7. Combining the above inequality and Lem. 8 gives

$$\mathcal{V}_T \leq 0.$$

Plugging the values of η, λ_1 and ϵ into the Lem. 5 gives,

$$\mathcal{R}_T = \mathcal{O}((\gamma_T + \|\gamma_T^g\|)\sqrt{T}).$$