Journal of Machine Learning Research 25 (2024) 1-54 Submitted 8/22; Revised 12/23; Published 3/24

Log Barriers for Safe Black-box Optimization with Application to Safe Reinforcement Learning

Ilnura Usmanova ILNURA.USMANOVA@PSI.CH
Swiss Data Science Center,
Paul Scherrer Institute, 5232 Villigen, Switzerland

Yarden As YARDEN.AS@INF.ETHZ.CH
Institute for Machine Learning, D-INFK,
ETH Zürich, 8092 Zurich, Switzerland

Maryam Kamgarpour* MARYAM.KAMGARPOUR@EPFL.CH
STI-IGM-Sycamore, EPFL, 1015 Lausanne, Switzerland

Andreas Krause* KRAUSEA@ETHZ.CH
Institute for Machine Learning, D-INFK,
ETH Zürich, 8092 Zurich, Switzerland

Editor: Csaba Szepesvari

Abstract

Optimizing noisy functions online, when evaluating the objective requires experiments on a deployed system, is a crucial task arising in manufacturing, robotics and various other domains. Often, constraints on safe inputs are unknown ahead of time, and we only obtain noisy information, indicating how close we are to violating the constraints. Yet, safety must be guaranteed at all times, not only for the final output of the algorithm.

We introduce a general approach for seeking a stationary point in high dimensional non-linear stochastic optimization problems in which maintaining *safety* during learning is crucial. Our approach called LB-SGD, is based on applying stochastic gradient descent (SGD) with a carefully chosen adaptive step size to a logarithmic barrier approximation of the original problem. We provide a complete convergence analysis of non-convex, convex, and strongly-convex smooth constrained problems, with first-order and zeroth-order feedback. Our approach yields efficient updates and scales better with dimensionality compared to existing approaches.

We empirically compare the sample complexity and the computational cost of our method with existing safe learning approaches. Beyond synthetic benchmarks, we demonstrate the effectiveness of our approach on minimizing constraint violation in policy search tasks in safe reinforcement learning (RL). [1]

Keywords: Stochastic optimization, safe learning, black-box optimization, smooth constrained optimization, reinforcement learning

1. Introduction

Many optimization tasks in robotics, manufacturing, health sciences, and finance require minimizing a loss function under constraints and uncertainties. In several applications, these

1. * Equal supervision

constraints are unknown at the outset of optimization, and one can infer the feasibility of inputs only from noisy measurements. For example, in manufacturing, the learner may want to tune the parameters of a machine. However, they can only observe noisy measurements of the constraints. Alternatively, in learning-based control tasks, e.g., in robotics, one may want to iteratively collect measurements and improve a pre-trained control policy in new environments. In such cases, during the optimization process, it is crucial to only query points (decision vectors) that satisfy the safety constraints, i.e., lie inside the feasible set, since querying infeasible points could lead to harmful consequences (Kirschner et al., 2019; Berkenkamp et al., 2016b). In such settings, even if the state constraints are known in advance, the learner may only have an approximate model of the true dynamics, e.g., through a simulator or a learned model. This implies that in the control policy space, exact constraints are also unknown, which makes non-violation of safety constraints while learning a challenging and important task. In the manufacturing example, one wants to sequentially update the parameters and take measurements of the machine performance while not violating a temperature limit during learning, which can impair the machine. In the robotics example, one wants to perform only safe policy updates, so as to avoid dangerous situations while collecting data in the new environment. This problem is known as *safe learning*.

In this work, we consider two general settings of safe learning. In the first case, we can only access the objective and constraints from noisy value measurements. This is referred to the zeroth-order (black-box) noisy information setting. In the second case, noisy gradient measurements are also available. This is referred to as the first-order noisy information setting. The literature on safe learning has focused mainly on the zeroth-order, black-box setting. To compare various methods in this setting, we can estimate their sample and computational complexity. The sample complexity represents the number of oracle queries in total that the learner has to make during the optimization to achieve the specific final accuracy $\varepsilon > 0$. Computational complexity is the total number of arithmetical operations the algorithm requires to achieve accuracy ε.

In the safe learning case, a large body of work is based on safe Bayesian optimization (BO) (Sui et al., 2015a; Berkenkamp et al., 2020). These approaches are typically based upon fitting a Gaussian process (GP) as a surrogate of the unknown objective and constraints functions based on the collected measurements. GPs are built using the predefined kernel function.

Although safe BO algorithms are provably safe and globally optimal, without simplifying assumptions, BO methods suffer from *the curse of dimensionality* (Frazier, 2018; Moriconi et al., 2019; Eriksson and Jankowiak, 2021). That is, their sample complexity might depend exponentially on the dimensionality for most commonly used kernels, including the squared exponential kernel (Srinivas et al., 2012). Moreover, their computational cost can also scale exponentially on the dimensionality since BO methods have to solve a non-linear programming (NLP) sub-problem at each iteration, which is, in general, an NP-hard problem. Together with this challenge, their computational cost scales cubically with the number of measurements, which provides a significant additional restriction on the number of measurements. These challenges make BO methods harder to employ on medium-to-big scale problems. [2]

2. Few works are extending BO approaches to high dimensions, see Snoek et al. (2015); Kirschner et al. (2019), Eriksson and Poloczek (2021).

Thus, our primary motivation is to find an algorithm such that *1)* its computational complexity scales efficiently to a large number of data points; *2)* its sample complexity scales efficiently to high dimensions; and *3)* it keeps optimization iterates within the feasible set of parameters with high probability.

We propose *Log Barriers SGD (LB-SGD)*, an algorithm that addresses the safe learning task by minimizing the log barrier approximation of the problem. This minimization is done by using Stochastic Gradient Descent (SGD) with a carefully chosen adaptive step size. We prove safety and derive the convergence rate of the algorithm for the convex and non-convex case (to the stationary point) and demonstrate LB-SGD's performance compared to other safe BO optimization algorithms on a series of experiments with various scales.

Our contributions We summarize our contributions below:

- We propose a unified approach for safe learning given a zeroth-order or first-order stochastic oracle. We prove that our approach generates feasible iterations with high probability and converges to a stationary point. Each iteration of the proposed method is computationally cheap and does not require solving any subproblems.

- In contrast to our past work on the log barrier approach (Usmanova et al., 2020), we develop a less conservative adaptive step size based on the smoothness constant instead of the Lipschitz constant of the constraints, enabling a tighter analysis.

- We provide a unified analysis, deriving the convergence rate of our algorithm for the stochastic non-convex, convex, and strongly-convex problems. We establish convergence despite the non-smoothness of the log barrier and the increasingly high variance of the log barrier gradient estimator.

- We empirically demonstrate that our method can scale to problems with high dimensions, in which previous methods fail. Moreover, we show the effectiveness of our approach in minimizing constraint violation in policy search in a high-dimensional constrained reinforcement learning (RL) problem.

Related work Although first-order stochastic optimization is widely explored (Nemirovsky and Yudin, 1985; Juditsky et al., 2013; Lan, 2020), we are not aware of any work addressing *safe* learning for first-order stochastic optimization. Therefore, even though our work also covers the first-order information case, we focus our review on the zeroth-order (black-box) optimization. Most relevant to our work are two areas: 1) *feasible* optimization approaches addressing smooth problems with *known* constraints; 2) existing *safe* approaches addressing smooth *unknown* objectives and constraints (including linear constraints). Here, by *unknown* constraints, we mean that we only have access to a noisy zeroth-order oracle of the constraints while solving the constrained optimization problem. By feasible optimization approaches, we refer to constrained optimization methods that generate a feasible optimization trajectory. For example, the Projected Gradient Descent and Frank-Wolfe are feasible, whereas dual approaches such as Augmented Lagrangian are infeasible.

There also exists another large body of work addressing *probabilistic* or *chance constraints* (Shapiro et al., 2009). This line of work aims to solve an optimization problem with probabilistic constraints in the form $\mathbb{P}\{F^i(x,\xi) \leq 0\} \geq 1-\delta$. However, the main difference is that in our safe learning task, we aim to satisfy the uncertain constraints with high probability *during the learning process*, not only in the end. Another issue is that the chance constraints problem, in general, is also a complex task with no universal solution. Typically, to address it, one requires either a significant number of (unsafe) measurements (e.g., in the scenario approach) or assumes some knowledge about the structure of the constraints a priori (e.g., in robust optimization). In contrast, in the current work, we propose a way to address the safe learning task without prior knowledge of the structure of the constraints.

We summarize the discussion of the algorithms from the past work as well as the best known lower bounds in Table 1. All works in Table 1 consider one-point feedback, except for Balasubramanian and Ghadimi (2018) who consider two-point feedback. *Two-point* feedback allows access to the function measurements with the same noise disturbance in at least two different points, whereas *one-point* feedback cannot guarantee this, and the noise can change at each single measurement. In our current work we also consider one-point feedback due to its generality. Next, we provide a detailed discussion of the past work.

Known constraints We start with smooth zeroth-order optimization with *known* constraints.

For *convex* problems with known constraints, several approaches address zeroth-order optimization with and without projections. Flaxman et al. (2005) propose an algorithm achieving a sample complexity of $O(\frac{d^2}{\varepsilon^4})$ using projections, where ε is the target accuracy, and d is the dimensionality of the problem. Bach and Perchet (2016) achieve $O\left(\frac{d^2}{\varepsilon^3}\right)$ sample complexity for smooth convex problems, and $O\left(\frac{d^2}{\mu\varepsilon^2}\right)$ for smooth μ-strongly-convex problems. Since the projections might be computationally expensive, in the projection-free setting, Chen et al. (2019) propose an algorithm achieving a sample complexity of $O(\frac{d^5}{\varepsilon^5})$ for stochastic optimization. Garber and Kretzu (2020) improve the bound for projection-free methods to $O(\frac{d^4}{\varepsilon^4})$ sample complexity. Instead of projections, both of the above works require solving linear programming (LP) sub-problems at each iteration. Bubeck et al. (2017) propose a kernel-based method for adversarial learning achieving $O(d^{9.5}T^{1/2})$ regret, and conjecture that a modified version of their algorithm can achieve $O(\frac{d^3}{\varepsilon^2})$ sample complexity for stochastic black-box convex optimization. This method uses a specific annealing schedule for exponential weights, and is quite complex; at each iteration $t > 0$ it requires sampling from a specific distribution p_t, which can be done in $\mathrm{poly}(d, \log(T))T$-time. For the smooth and strongly-convex case, Hazan and Luo (2016) propose a method that achieves $O(\frac{d^3}{\varepsilon^2})$. The general lower bound for the convex black-box stochastic optimization $O\left(\frac{d^2}{\varepsilon^2}\right)$ is proposed by Shamir (2013). To the best of our knowledge, there is no proposed lower bound for the safe convex black-box optimization with unknown constraints.

For *non-convex* optimization, Balasubramanian and Ghadimi (2018) provide a comprehensive analysis of the performance of several zeroth-order algorithms allowing two-point zeroth-order feedback.[3]

There exist also other classical derivative-free optimization methods addressing non-convex optimization based on various heuristics. One example is the Nelder-Mead approach,

3. The difference between one-point and two-point feedback is that two-point feedback allows access to the function with the same noise in multiple points, which is a significantly stronger assumption than the one-point feedback (Duchi et al., 2015).

Algorithm	Sample complexity	Computational complexity	Constraints	Convexity
Bach and Perchet (2016)	$O\left(\frac{d^2}{\varepsilon^3}\right)$	projections	known	yes
Bach and Perchet (2016)	$O\left(\frac{d^2}{\mu\varepsilon^2}\right)$	projections	known	μ-strongly-convex
Bubeck et al. (2017)	$O\left(\frac{d^3}{\varepsilon^2}\right)$	samplings from p_t distribution	known	yes
Balasubramanian and Ghadimi (2018)	$O\left(\frac{d}{\varepsilon^4}\right)$	$O\left(\frac{1}{\varepsilon^2}\right)$ LPs	known	no
Garber and Kretzu (2020)	$O\left(\frac{d^4}{\varepsilon^4}\right)$	$O\left(\frac{1}{\varepsilon^2}\right)$ LPs	known	yes
Usmanova et al. (2019)	$O\left(\frac{d^3}{\varepsilon^3}\right)$	$O\left(\frac{1}{\varepsilon^2}\right)$ LPs	unknown, linear	yes
Fereydounian et al. (2020)	$\tilde{O}\left(\frac{d^2}{\varepsilon^4}\right)$	$O\left(\frac{1}{\varepsilon^2}\right)$ LPs	unknown, linear	yes/no
Berkenkamp et al. (2016a)	$\tilde{O}\left(\frac{\gamma(d)}{\varepsilon^2}\right)$	$O\left(\frac{\gamma(d)}{\varepsilon^2}\right)$ NLPs	unknown	no
This work	$O\left(\frac{d^2}{\varepsilon^7}\right)$	$O\left(\frac{1}{\varepsilon^3}\right)$ gradient steps	unknown	no
This work	$\tilde{O}\left(\frac{d^2}{\varepsilon^6}\right)$	$\tilde{O}\left(\frac{1}{\varepsilon^2}\right)$ gradient steps	unknown	yes
This work	$\tilde{O}(\frac{d^2}{\mu\varepsilon^4})$	$\tilde{O}\left(\frac{1}{\varepsilon^2}\right)$ gradient steps	unknown	μ-strongly-convex
Lower bound (Shamir, 2013)	$O\left(\frac{d^2}{\varepsilon^2}\right)$	-	known	yes

Table 1: Zeroth-order safe smooth optimization algorithms. Here ε is the target accuracy, and d is the dimension of the decision variable. Many of the cited works provide the bounds in terms of regret, which can be converted to stochastic optimization accuracy. In SafeOpt, $\gamma(d)$ depends on the kernel, and might be exponential in d. All the above works consider one-point feedback, except for Balasubramanian and Ghadimi (2018) who consider two-point feedback. In Bubeck et al. (2017), at each iteration the sampling from a a specifically updated distribution p_t can be done in poly$(d, \log(T))T$-time. Under $\tilde{O}(\cdot)$, we hide a multiplicative logarithmic factor.

also known as simplex downhill (Nelder and Mead, 1965). To handle constraints it uses penalty functions (Luersen et al., 2004), or barrier functions (Price, 2019). Another example are various evolutionary algorithms (Storn and Price, 1997; Kennedy and Eberhart, 1995; Rechenberg, 1989; Hansen and Ostermeier, 2001). Nevertheless, all of these approaches are based on heuristics and thus do not provide theoretical convergence rate guarantees, at best establishing asymptotic convergence.

Unknown constraints There are much fewer works on safe learning for problems with a non-convex objective and *unknown* constraints. A significant line of work covers objectives and constraints with bounded reproducing kernel Hilbert space (RKHS) norm (Sui et al., 2015b; Berkenkamp et al., 2016a), based on Bayesian Optimization (BO). Also, for the *linear* bandits problem, Amani et al. (2019) design a Bayesian algorithm handling safety constraints. These works build Bayesian models of the constraints and the objective using Gaussian processes (Rasmussen and Williams, 2005, GP) and crucially require a suitable GP prior. In contrast, in our work, we *do not* use GP models and do not require a prior model for the functions. Additionally, most of these approaches do not scale to high-dimensional problems. Kirschner et al. (2019) proposes an adaptation to higher dimensions using line search called LineBO, which demonstrates strong performance in safe and non-safe learning in practical applications. However, they derive the convergence rate only for the unconstrained case, whereas for the constrained case, they only prove safety without convergence. We empirically compare our approach with their method in high dimensions and demonstrate that our approach can solve problems where LineBO struggles. There are other more recent works on practically scalable constrained BO approaches, such as by Eriksson and Poloczek (2021), although they do not provide theoretical convergence rates and safety guarantees.

From the optimization side, in the case of *unknown* constraints, projection-based optimization techniques or Frank-Wolfe-based ones cannot be directly applied. Such approaches require solving subproblems with respect to the constraint set, and thus the learner requires at least an approximate model of it. One can build such a model in the special case of polytopic constraints. For example, Usmanova et al. (2019) propose a safe algorithm for convex learning with smooth objective and *linear* constraints based on the Frank-Wolfe algorithm. Building on the above, Fereydounian et al. (2020) propose an algorithm for both convex and non-convex objective and *linear* constraints. Both these methods consider first-order noisy objective oracle and zeroth-order noisy constraints oracle.

For the more general case of non-linear programming, there are recent safe optimization approaches based on the interior point method (IPM). Usmanova et al. (2020) propose using the log barrier gradient-based algorithm for the non-convex non-smooth problem with zeroth-order information. They show the sample complexity to be $\tilde{O}\left(\frac{d^3}{\varepsilon^9}\right)$. Here, we hide a multiplicative logarithmic factor under $\tilde{O}(\cdot)$. The work above is built on the idea of Hinder and Ye (2019) who propose the analysis of the gradient-based approach to solving the log barrier optimization (in the deterministic case). The first-order approach of Hinder and Ye (2019) has sample complexity of $O(\frac{1}{\varepsilon^3})$, compared to which Usmanova et al. (2020) is much slower due to harder non-smoothness and zeroth-order conditions. In the current paper, we extend the above works to smooth non-convex ($\tilde{O}(\frac{d^2}{\varepsilon^7})$), convex ($\tilde{O}(\frac{d^2}{\varepsilon^6})$) and strongly-convex ($\tilde{O}(\frac{d^2}{\varepsilon^4})$) problems for both first-order and zeroth-order stochastic information. Note that IPM is a feasible optimization approach by definition. By using self-concordance properties of specifically chosen barriers and second-order information, IPM is highly efficient in solving LPs, QPs, and conic optimization problems. However, constructing barriers with self-concordance properties is not possible for unknown constraints. Therefore, we focus on logarithmic barriers.

Price of safety To finalize Table 1, compared to the state-of-the-art works with tractable algorithms and *known* constraints (Bach and Perchet, 2016), we pay a price of an order $\tilde{O}(\varepsilon^{-3})$ in zeroth-order optimization just for the *safety* with respect to unknown constraints both in convex and strongly-convex cases. In the non-convex case, we pay $O(d\varepsilon^{-3})$ both for safety and having one-point feedback compared to Balasubramanian and Ghadimi (2018) considering 2-point feedback. As for the computational complexity, our method is projection-free and does not require solving any subproblems compared to the above methods.

Paper organization We organize our paper as follows. In Section 2, we formalize the problem, and define the assumptions for the first-order stochastic setting. In Section 3, we describe our main approach for solving this problem, describe our main theoretical results about it, and establish its safety. In Section 4, we specialize our approach for the non-convex, convex and strongly-convex cases. For each of these cases, we also provide the suitable optimality criterion and the convergence rate analysis. Then, in Section 4.5 we specifically analyse the setting of the zeroth-order information, and derive the sample complexity of all variants of our method in this setting. In Section 5 we compare our approach empirically with other existing safe learning approaches and additionally demonstrate its performance on a high-dimensional constrained reinforcement learning (RL) problem.

2. Problem Statement

We consider a general constrained optimization problem:

$$\begin{aligned} &\min f^0(x) \\ &\text{s.t. } f^i(x) \leq 0, i = 1, \ldots, m, \end{aligned} \tag{P}$$

where the objective function $f^0 : \mathbb{R}^d \to \mathbb{R}$ and the constraints $f^i : \mathbb{R}^d \to \mathbb{R}$ are *unknown*, possibly non-convex functions.

We denote by $\mathcal{X}$ the feasible set $\mathcal{X} := \{x \in \mathbb{R}^d : f^i(x) \leq 0, i \in [m]\}$, where $[m] := \{1, \ldots, m\}$. By $Int(\mathcal{X})$ we denote the interior of the set $\mathcal{X}$. By $\|\cdot\|$ we denote the Euclidean ℓ_2-norm. A function $f : \mathbb{R}^d \to \mathbb{R}$ is called *L-Lipschitz continuous* on $\mathcal{X}$ if $|f(x)-f(y)| \leq L\|x-y\|\ \forall x, y \in \mathcal{X}$. It is called *$M$-smooth* on $\mathcal{X}$ if $f(x) \leq f(y)+\langle\nabla f(y), x-y\rangle + \frac{M}{2}\|x-y\|^2\ \forall x, y \in \mathcal{X}$. It is called *$\mu$-strongly convex* on $\mathcal{X}$ if $f(x) \geq f(y) + \langle\nabla f(y), x-y\rangle + \frac{\mu}{2}\|x-y\|^2\ \forall x, y \in \mathcal{X}$. By $\mathcal{L}(x, \lambda) := f^0(x) + \sum_{i=1}^m \lambda^i f^i(x)$ we denote the Lagrangian function of a problem $\min_{x\in\mathbb{R}^d} f^0(x)$ s.t. $f^i(x) \leq 0\ \forall i \in [m]$, where $\lambda \in \mathbb{R}^m$ is the dual vector.

Our goal is to solve the *safe learning* problem. That is, we need to find the solution to the constrained problem (P) while keeping all the iterates x_t of the optimization procedure feasible $x_t \in \mathcal{X}$ with high probability during the learning process. Throughout this paper we make the following assumptions:

Assumption 1 *Let $\mathcal{X}$ have a bounded diameter, that is, $\exists R > 0$ such that for any $x, y \in \mathcal{X}$ we have $\|x - y\| \leq R$.*

Assumption 2 *The objective and the constraint functions $f^i(x)$ for $i \in \{0, \dots, m\}$ are M_i-smooth and L_i-Lipschitz continuous on $\mathcal{X}$ with constants $L_i, M_i > 0$. We denote by $L := \max_{i\in\{0,\dots,m\}}\{L_i\}$ and $M := \max_{i\in\{0,\dots,m\}}\{M_i\}$.*

The above two assumptions are standard in the optimization literature. Without any assumptions on the constraints, guaranteeing safety is impossible. We assume the upper bounds on the Lipschitz and smoothness constants to be known.

Assumption 3 *There exists a known starting point $x_0 \in \mathcal{X}$ at which $\max_{i\in[m]} f^i(x_0) \leq -\beta$, for $\beta > 0$.*

The third assumption ensures that we have a safe starting point, away from the boundary. In the absence of such an assumption, even the first iterate may be unsafe.

Assumption 4 *For a given $\rho > 0$ let $\mathcal{I}_\rho(x) := \{i \in [m] | f^i(x) \geq -\rho\}$ be the set of ρ-approximately active constraints at x. There exists $\rho \in (0, \frac{\beta}{2}]$ such that for any point $x \in \mathcal{X}$ there exists a direction $s_x \in \mathbb{R}^d : \|s_x\| = 1$ and $l > 0$, such that $\langle s_x, \nabla f^i(x)\rangle > l, \forall i \in \mathcal{I}_\rho(x)$.*

The last assumption is the extended Mangasarian-Fromovitz constraint qualification (MFCQ). The classic MFCQ (Mangasarian and Fromovitz, 1967) is a regularity assumption on the constraints, guaranteeing that they have a uniform descent direction for all constraints at a local optimum. Our extended MFCQ guarantees this regularity condition at all points ρ-close to the boundary. For the classic MFCQ and further details on our extension to it, please refer to Appendix A.1. This assumption holds for example for convex problems with the constraint set having a non-empty interior, as shown in Section 4.3.

2.1 Oracle

Typically in the applications we consider, the information available to the learner is noisy. For example, one can only observe perturbed gradients and values of $f^i, \forall i = 0, \dots, m$ at the requested points x_t. Therefore, formally we consider access to the first-order stochastic oracle for every $f^i(x)$, providing the pair of value and gradient stochastic measurements:

$$\mathcal{O}(f^i, x, \xi) = (F^i(x, \xi), G^i(x, \xi)). \tag{1}$$

Note that the formulation allows (but does not require) that $F^i(x, \xi)$ and $G^i(x, \xi)$ are correlated. In particular, this formulation allows to define the vector of $\xi = \{(\xi_0^i, \xi_1^i)\}_{i=0,\dots,m}$ such that each $F^i(x, \xi) = F^i(x, \xi_0^i)$ and $G^i(x, \xi) = G^i(x, \xi_1^i)$. In this formulation, $\{(\xi_0^i, \xi_1^i)\}_{i=0,\dots,m}$ can be either correlated or independent of each-other. The parts of the oracle are given as follows:

1) **Stochastic value $F^i(x, \xi)$.** We assume $F^i(x, \xi)$ is unbiased

$$\mathbb{E}[F^i(x, \xi)] = f^i(x),$$

and sub-Gaussian with variance bounded by σ_i^2, that is,

$$\mathbb{P}\left\{|F^i(x, \xi) - f^i(x)| \leq \sigma_i\sqrt{\ln\frac{1}{\delta}}\right\} \geq 1 - \delta,\ i \in \{0, \dots, m\}.$$

2) **Stochastic gradient $G^i(x,\xi)$.** We assume that its bias is bounded by

$$\|\mathbb{E}G^i(x,\xi) - \nabla f^i(x)\| \leq \hat{b}_i,$$

where $\hat{b}_i \geq 0$, and it is sub-Gaussian with the variance such that $\mathbb{E}[\|G^i(x,\xi) - \mathbb{E}G^i(x,\xi)\|^2] \leq \hat{\sigma}_i^2$.

Note that the variances of $F^i(x,\xi)$ and $G^i(x,\xi)$ are fixed and given by the nature of the problem. However, we can decrease these variances by taking several measurements per iteration and replacing $(F^i(x,\xi), G^i(x,\xi))$ with

$$F_n^i(x,\xi) := \frac{\sum_{j=1}^n F^i(x,\xi_j)}{n} \text{ and } G_n^i(x,\xi) := \frac{\sum_{j=1}^n G^i(x,\xi_j)}{n}. \tag{2}$$

In the above, we abuse the notation and replace the dependence $F_n^i(x,\xi_1,\ldots,\xi_n)$ by $F_n^i(x,\xi)$ for simplicity. Then, their variances become respectively such that

$$\mathbb{E}[\|F_n^i(x,\xi) - \mathbb{E}F_n^i(x,\xi)\|^2] \leq \sigma_i^2(n) := \frac{\sigma_i^2}{n}, \tag{3}$$

$$\mathbb{E}[\|G_n^i(x,\xi) - \mathbb{E}G_n^i(x,\xi)\|^2] \leq \hat{\sigma}_i^2(n) := \frac{\hat{\sigma}_i^2}{n}. \tag{4}$$

Our goal is given the provided first-order stochastic information, to find an approximate solution of problem (P) while not making value and gradient queries outside the feasibility set $\mathcal{X}$ with high probability. To do so, we introduce the log barrier optimization approach.

3. General Approach

In this section we describe our main approach and provide the main theoretical results of our paper.

3.1 Safe learning with log barriers

The main idea of the approach is to replace the original constrained problem (P) by its unconstrained *log barrier* surrogate $\min_{x\in\mathbb{R}^d} B_\eta(x)$, where $B_\eta(x)$ and its gradient $\nabla B_\eta(x)$ are defined as follows

$$B_\eta(x) = f^0(x) + \eta\sum_{i=1}^m -\log(-f^i(x)), \tag{5}$$

$$\nabla B_\eta(x) = \nabla f^0(x) + \eta\sum_{i=1}^m \frac{\nabla f^i(x)}{-f^i(x)}. \tag{6}$$

This surrogate $B_\eta(x)$ grows to infinity as the argument converges to the boundary of the set $\mathcal{X}$, and is defined only in the interior of the set $Int(\mathcal{X})$. Therefore, under Assumptions 1 to 4, a major advantage of this method for the problems we consider is that by carefully choosing the optimization step-size, *the feasibility of all iterates is maintained automatically.* We run Stochastic Gradient Descent (SGD) with the specifically chosen step size applied to the log barrier surrogate $\min_{x\in\mathbb{R}^d} B_\eta(x)$.

The main intuition is that the descent direction of the log barrier pushes the iterates away from the boundary, at the same time converging to an approximate KKT point for the non-convex case, and to an approximate minimizer for the convex case. To measure the approximation in the non-convex case, we define the *ε-approximate KKT point* **(ε-KKT)**. Specifically, for $\varepsilon > 0$ and a pair (x, λ), such point satisfies the following conditions:

$$\lambda^i, -f^i(x) \geq 0, \ \forall i \in [m] \tag{ε-KKT.1}$$

$$\lambda^i(-f^i(x)) \leq \varepsilon, \ \forall i \in [m] \tag{ε-KKT.2}$$

$$\|\nabla_x \mathcal{L}(x, \lambda)\| \leq \varepsilon. \tag{ε-KKT.3}$$

Hereby, λ is the vector of dual variables and $\mathcal{L}(x, \lambda) := f^0(x) + \sum_{i=1}^m \lambda^i f^i(x)$ is the Lagrangian function of (P). Later, we show that SGD on the η-log barrier surrogate converges to an ε-approximate KKT point with $\varepsilon = \eta$. We show it by demonstrating that the small barrier gradient norm $\|\nabla B_\eta(\hat{x})\| \leq \eta$ corresponds to a small gradient of the Lagrangian $\|\nabla_x \mathcal{L}(x, \lambda)\|$ with specifically chosen vector of dual variables $\lambda \in \mathbb{R}^m$ (Hinder and Ye, 2019; Usmanova et al., 2020). In the convex case, the approximate optimality in the value $B_\eta(\hat{x}) - B_\eta(x^*_\eta) \leq \eta$ itself implies that $\hat{x}$ is an ε-approximate solution of the original problem: $f^0(\hat{x}) - \min_{x \in \mathcal{X}} f^0(x) \leq \varepsilon$ with $\varepsilon > 0$ linearly dependent on η up to a logarithmic factor.

3.2 Main results

We propose to apply SGD with an adaptive step-size to minimize the unconstrained log barrier objective B_η. We name our approach LB-SGD. We show that LB-SGD (with confidence parameter $\delta = \frac{\hat{\delta}}{T(2m+1)}$) achieves the following convergence results for the target probability $1 - \hat{\delta}$:

1. For the non-convex case, after at most $T = O(\frac{1}{\varepsilon^3})$ iterations, and with $\sigma_i(n) = O(\varepsilon^2)$ and $\hat{\sigma}_i(n) = O(\varepsilon)$, LB-SGD outputs $x_{\hat{t}}$ which is an ε-KKT point with probability $1 - \hat{\delta}$. In total, we require $N = Tn = O(\frac{1}{\varepsilon^7})$ oracle queries $\mathcal{O}(f^i, x, \xi)$ for all $i \in \{0, \dots, m\}$. (Theorem 8)

2. For the convex case, after at most $T = \tilde{O}(\frac{\|x_0 - x^*\|^2}{\varepsilon^2})$ iterations of LB-SGD, and with $\sigma_i(n) = \tilde{O}(\varepsilon^2)$ and $\hat{\sigma}_i(n) = \tilde{O}(\varepsilon)$, we obtain output $\bar{x}_T$ such that with probability $1 - \hat{\delta}$: $f^0(\bar{x}_T) - \min_{x \in \mathcal{X}} f^0(x) \leq \varepsilon$. In total, we require $N = Tn = \tilde{O}(\frac{1}{\varepsilon^6})$ calls of the oracle $\mathcal{O}(f^i, x, \xi)$ for all $i \in \{0, \dots, m\}$. (Theorem 10)

3. For the μ-strongly-convex case, after at most $T = \tilde{O}(\frac{1}{\mu\varepsilon} \log \frac{1}{\varepsilon})$ iterations of LB-SGD and with $\sigma_i(n) = \tilde{O}(\varepsilon^{1.5})$ and $\hat{\sigma}_i(n) = \tilde{O}(\sqrt{\varepsilon})$, for the output $\hat{x}_K$ we have with probability $1 - \hat{\delta}$: $f^0(\hat{x}_K) - \min_{x \in \mathcal{X}} f^0(x) \leq \varepsilon$. In total, we require $N = \tilde{O}\left(\frac{1}{\varepsilon^4}\right)$ calls of the oracle $\mathcal{O}(f^i, x, \xi)$ for all $i \in \{0, \dots, m\}$. (Theorem 11)

4. For the zeroth-order information case, estimating the function gradients using finite difference, we obtain the following bounds on the number of measurements (Corollary 15):

    - $N = O(\frac{d^2}{\varepsilon^7})$ to get an ε-approximate KKT point in the non-convex case;

- $N = \tilde{O}(\frac{d^2}{\varepsilon^6})$ to get an ε-approximate minimizer in the convex case;
- $N = \tilde{O}(\frac{d^2}{\varepsilon^4})$ to get an ε-approximate minimizer in the strongly-convex case;

5. In all of the above cases the safety is guaranteed with probability $1-\delta$ for all the measurements. (Theorem 4, Corollary 15)

In the above, $\tilde{O}(\cdot)$ denotes $O(\cdot)$ dependence up to a multiplicative logarithmic factor. Note that for zeroth-order information case we only pay the price of a multiplicative factor d^2.

3.3 Our approach

To minimize the log barrier function, we employ SGD using the stochastic first-order oracle providing $(F^i(x,\xi), G^i(x,\xi))$ with an adaptive step size, and derive convergence rate of our methods dependent on the noise level of this oracle. At iteration t we make the step in the form:

$$x_{t+1} \leftarrow x_t - \gamma_t g_t, \tag{7}$$

where γ_t is a safe adaptive step size, g_t being the log barrier gradient estimator. In Section 3.3.1, we show how to build the estimator g_t of the log barrier gradient. Following that, in Section 3.3.2, we explain how to choose γ_t.

As mentioned before, the log barrier function is not a smooth function due to the fact that close to the boundaries of $\mathcal{X}$ it converges to infinity. To address non-smooth stochastic problems, optimization schemes in the literature typically require bounded sub-gradients. For the log barrier function even this condition does not hold in general. Hence, we cannot expect the classical analysis with the standard predefined step size to hold when applying SGD to the log barrier problem. Contrary to that, by making the step size *adaptive*, we can guarantee *local-smoothness* of the log barrier. Intuitively, this is done by restricting the growth of the constraints. We leverage this property in our analysis. In particular, let γ_t be such that $f^i(x_{t+1}) \leq \frac{f^i(x_t)}{2}$ for every constraint. Then, the log barrier is locally-smooth at point x_t with constant $M_2(x_t)$

$$M_2(x_t) \leq M_0 + 10\eta \sum_{i=1}^{m} \frac{M_i}{\alpha_t^i} + 8\eta \sum_{i=1}^{m} \frac{(\theta_t^i)^2}{(\alpha_t^i)^2}, \tag{8}$$

where $\theta_t^i = \langle \nabla f^i(x_t), \frac{g_t}{\|g_t\|}\rangle$, and $\alpha_t^i = -f^i(x_t)$ for all $i \in [m]$. The growth on the constraints can be bounded by any constant in $(0,1)$ – we pick $\frac{1}{2}$ for simplicity, similarly to Hinder and Ye (2018). We further analyze this adaptivity property and the local smoothness of $M_2(x_t)$ in Section 3.3.2. Importantly, our local smoothness $M_2(x)$ bound is more accurate since it is constructed by exploiting the smoothness of the constraints and takes into account the gradient measurements. In contrast, the bound of Hinder and Ye (2019) relies on Lipschitz continuity without considering the gradient measurements.

3.3.1 The log barrier gradient estimator

The key ingredient of the log barrier method together with the safe step size is estimating the log barrier gradient.

Estimating the gradient Recall that the log barrier gradient by definition is:

$$\nabla B_\eta(x_t) = \nabla f^0(x_t) + \eta \sum_{i=1}^m \frac{\nabla f^i(x_t)}{\alpha_t^i}.$$

Since we only have the stochastic information, we estimate the log barrier gradient as described in Algorithm 3. In the above, we allow to take a batch of measurements per call

Algorithm 1 Log Barrier Gradient estimator $\mathcal{O}_\eta(x_t, n)$

1: *Input:* Oracles $F^i(\cdot,\xi), G^i(\cdot,\xi),\ \forall i \in \{0,\dots,m\}$, $x_t \in \mathcal{X}$, $\eta > 0$, number of measurements n
2: $g_t \leftarrow G_n^0(x_t,\xi_t) + \eta \sum_{i=1}^m \frac{G_n^i(x_t,\xi_t)}{-F_n^i(x_t,\xi_t)}$;
3: *Output:* g_t

and average them as defined in (2) in order to reduce the variances $\sigma_i^2(n) := \frac{\sigma_i^2}{n}, \hat{\sigma}_i^2(n) := \frac{\hat{\sigma}_i^2}{n}$.

Properties of the estimator The log barrier gradient estimator defined above is biased and can be heavy tailed, since a part of it is a ratio of two sub-Gaussian random variables. Therefore, in the following lemma we provide a general upper confidence bound on the deviation. We denote $\bar{\alpha}_t^i := -F_n^i(x_t,\xi_t)$.

Lemma 1 *Given that $x_t \in Int\{\mathcal{X}\}$, that is, $\alpha_t^i > 0\ \forall i \in [m]$, the deviation of the log barrier gradient estimator $\Delta_t := g_t - \nabla B_\eta(x_t)$ satisfies:*

$$\mathbb{P}\left\{\|\Delta_t\| \le \hat{b}_0 + \hat{\sigma}_0(n)\sqrt{\ln\frac{1}{\delta}} + \sum_{i=1}^m \frac{\eta}{\alpha_t^i}\left(\hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}}\right) + \sum_{i=1}^m \hat{L}_i \frac{\eta\sigma_i(n)}{\alpha_t^i[\bar{\alpha}_t^i]_+}\sqrt{\ln\frac{1}{\delta}}\right\} \ge 1-\delta, \tag{9}$$

where $\hat{L}_i = L_i + \hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}}$.

From the above bound, we can see that the closer we are to the boundary, the smaller α_t^i becomes, and the smaller variance σ_i we require to keep the same level of disturbance of the barrier gradient estimator. That is, the closer to the boundary, the more measurements we require to stay safe despite the disturbance, which is quite natural. For the proof see Appendix A.2.

The above deviation consists of the variance part and the bias part. Note that the bias is non-zero even if the biases of the gradient estimators are zero. It can be bounded as follows (see Appendix A.2.1):

$$\|\mathbb{E}\Delta_t\| \le \sum_{i=1}^m \frac{\eta L_i \sigma_i(n)}{(\alpha_t^i)^2} + \hat{b}_0 + \sum_{i=1}^m \frac{\eta}{\alpha_t^i}\hat{b}_i. \tag{10}$$

In the above, the expectation is taken given fixed x_t. This bias comes from the fact that we are estimating the ratio of two sub-Gaussian distributions, which is often heavy-tailed and even for Gaussian variables might behave very badly if the mean of the denominator is smaller than its variance (Díaz-Francés and Rubio, 2013). This fact influences the SGD analysis, and does not allow getting convergence guarantees with larger noise. Therefore, our algorithm is very sensitive to the noise $\sigma_i(n)$ and might require many samples per iteration to reduce this noise.

3.3.2 Adaptive step-size γ_t

First, recall that the log barrier is non-smooth on $\mathcal{X}$, since it grows to infinity on the boundary. However, we can use the notion of the $M_2(x_t)$-local smoothness, that guarantees smoothness in a bounded region around the current point $x_t \in \mathcal{X}$ along the step direction g_t defined by:

$$\mathcal{S}(x_t) := \left\{ y \in \mathbb{R}^d : y = x_t - u g_t, \forall u \in [0, \gamma_t] \text{ such that } f^i(y) \leq \frac{f^i(x_t)}{2}, \forall i \in [m] \right\}. \tag{11}$$

Then, the function B_η we call $M_2(x_t)$-*locally smooth* around x_t if for any $x, y \in \mathcal{S}(x_t)$ we have

$$\|\nabla B_\eta(x) - \nabla B_\eta(y)\| \leq M_2(x_t)\|x - y\|.$$

The local smoothness of the Log Barrier $B_\eta(x)$ is required for our convergence analysis of the SGD.

$M_2(x_t)$-local smoothness constant for the log barrier We derive our local smoothness constant based on the M_i-smoothness of the objective and constraints f^i for $i = 0, \dots, m$. Compared to the Lipschitz constant-based approach (used in Hinder and Ye (2019); Usmanova et al. (2020)), our way to bound the local smoothness constant $M_2(x_t)$ allows to estimate it more tightly since we use the quadratic upper bound instead of the linear bound.

Lemma 2 *On the bounded area $\mathcal{S}(x_t)$ around x_t along the step direction g_t within the step-size γ_t such that the next iterate is restricted by $f^i(y) \leq \frac{f^i(x_t)}{2}, y = x_t - u g_t, u \in [0, \gamma_t]$ the log barrier $B_\eta(x_t)$ is locally-smooth with*

$$M_2(x_t) := M_0 + 2\eta \sum_{i=1}^m \frac{M_i}{\alpha_t^i} + 4\eta \sum_{i=1}^m \frac{\langle \nabla f^i(x_{t+1}), \frac{g_t}{\|g_t\|}\rangle^2}{(\alpha_t^i)^2}. \tag{12}$$

Moreover, if $\gamma_t \leq \frac{\alpha_t^i}{2|\theta_t^i| + \sqrt{\alpha_t^i M_i}} \frac{1}{\|g_t\|}$, then

$$M_2(x_t) := M_0 + 10\eta \sum_{i=1}^m \frac{M_i}{\alpha_t^i} + 8\eta \sum_{i=1}^m \frac{(\theta_t^i)^2}{(\alpha_t^i)^2}, \tag{13}$$

where $\theta_t^i = \langle \nabla f^i(x_t), \frac{g_t}{\|g_t\|}\rangle$.

For the proof of Lemma 2 see Appendix A.4. As detailed in the next paragraph, the condition on γ_t defined in the second part of the lemma is sufficient to ensure a bounded constraints growth (11). In the case with inexact measurements, we have to use lower bounds on α_t^i and upper bounds on θ_t^i. We denote by $\underline{\alpha}_t^i := \bar{\alpha}_t^i - \sigma_i(n)\sqrt{\ln \frac{1}{\delta}}$ a lower bound on $\alpha_t^i : \mathbb{P}\{\alpha_t^i \geq \underline{\alpha}_t^i\} \geq 1 - \delta$. We denote an upper bound on θ_t^i by $\hat{\theta}_t^i := |\langle G_n^i(x, \xi), \frac{g_t}{\|g_t\|}\rangle| + \hat{b}_i + \hat{\sigma}_i(n)\sqrt{\log \frac{1}{\delta}},\ \forall i \in [m]$ such that $\mathbb{P}\{\theta_t^i \leq \hat{\theta}_t^i\} \geq 1 - \delta$. Then, an upper bound on $M_2(x_t)$ can be computed as follows

$$\hat{M}_2(x_t) = M_0 + 10\eta \sum_{i=1}^m \frac{M_i}{\underline{\alpha}_t^i} + 8\eta \sum_{i=1}^m \frac{(\hat{\theta}_t^i)^2}{(\underline{\alpha}_t^i)^2}. \tag{14}$$

Adaptivity of the step-size In the above, we bound the local smoothness of the log barrier at the next iterate $x_{t+1} = x_t - \gamma_t g_t$ by carefully choosing a step size γ_t in a way that ensures the next iterate to remain in the set (11), i.e., such that $f^i(x_{t+1}) \leq \frac{f^i(x_t)}{2}$. This guarantees the feasibility of x_{t+1} given the feasibility of x_t.

One way to get the adaptive step size γ_t is to use the Lipschitz constants L_i of f^i to bound γ_t (see Hinder and Ye (2019); Usmanova et al. (2020)):

$$\gamma_t \leq \min_{i \in [m]} \frac{-f^i(x_t)}{2L_i} \frac{1}{\|g_t\|}.$$

In practice, L_i are typically unknown or overestimated. For example, even in the quadratic case $f^i(x) = \|x\|^2$, L_i depends on the diameter of the set $\mathcal{X}$, and thus might be very conservative in the middle of the set. Again, we propose to use the smoothness constants M_i for safety instead.[4] Indeed, smoothness parameters are susceptible to overestimation as are Liptschitz constants. However, problem-adaptive techniques (Vaswani et al., 2021) and approaches to efficiently estimate such constants (Fazlyab et al., 2019) make promising venue for relaxing this challenge.

Lemma 3 *The adaptive safe step size γ_t bounded by*

$$\gamma_t \leq \min_{i \in [m]} \left\{ \frac{\alpha_t^i}{2|\theta_t^i| + \sqrt{\alpha_t^i M_i}} \right\} \frac{1}{\|g_t\|},$$

guarantees $f^i(y) \leq \frac{f^i(x_t)}{2}$ for all $y = x_t - u g_t$ with $u \in [0, \gamma_t]$, including $y = x_{t+1}$.

The proof is based on the smoothness bound on the constraint growth:

$$f^i(x_{t+1}) \leq f^i(x_t) - \gamma_t \langle \nabla f^i(x_t), g_t \rangle + \gamma_t^2 \frac{M_i}{2} \|g_t\|^2.$$

For the full proof see Appendix A.3. We illustrate the principle of choosing this adaptive bound on Figure 1. Then, finally, we set the step size to:

$$\gamma_t = \min \left\{ \min_{i \in [m]} \left\{ \frac{\underline{\alpha}_t^i}{2|\hat{\theta}_t^i| + \sqrt{\underline{\alpha}_t^i M_i}} \right\} \frac{1}{\|g_t\|}, \frac{1}{\hat{M}_2(x_t)} \right\}. \tag{15}$$

3.3.3 Basic algorithm

To sum up, below we propose our basic algorithm, but emphasize that it can be instantiated differently for different problem classes. We showcase possible instantiations of LB-SGD in following Section 4.

4. *Firstly,* the Lipschitz constant, even if tight, provides the first-order linear upper bound on the constraint growth, whereas using the smoothness constant we can exploit more reliable and tight second order upper bound on the constraint growth. *Secondly,* Lipschitz constant is often much harder to estimate since it might strongly depend on the size of the set. By the same reason, in practice, even for the hard functions modeled by a neural network with smooth activation functions, we can estimate the smoothness parameters, but it is much less clear how to estimate the Lipschitz constants properly.

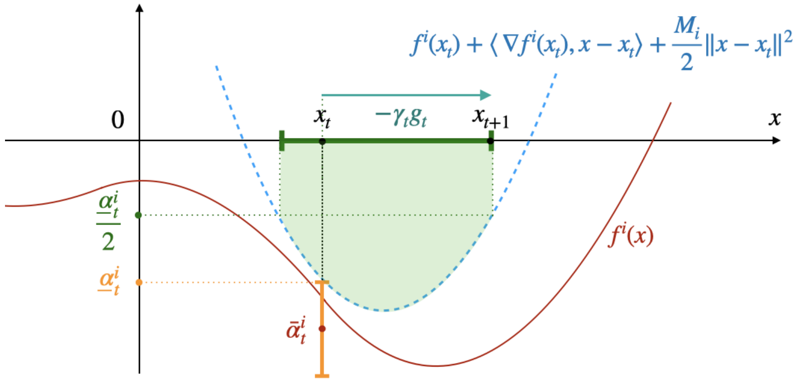

Figure 1: Illustration of the step size adaptivity. LB-SGD chooses γ_t such that the quadratic smoothness upper bound (*black*) on the constraint guarantees $f^i(x_{t+1}) \leq f^i(x_t)/2$. $\underline{\alpha}_t^i$ is the lower bound on $\alpha_t^i = -f^i(x_t)$, constructed based on the mean estimator $\bar{\alpha}_t^i$. The *orange* interval denotes the confidence interval for α_t^i. The *green* interval denotes the adaptive region for x_{t+1} based on the requirement $f^i(x_{t+1}) \leq f^i(x_t)/2$.

Algorithm 2 LB-SGD(x_0, η, T, n)

Input: $M_i, \sigma_i, \hat{\sigma}_i, \hat{b}_i \in \mathbb{R}_+ \ \forall i \in \{0, \dots, m\}$, $R \in \mathbb{R}_+$, $\eta \in \mathbb{R}_+$, $n \in \mathbb{N}$, $T \in \mathbb{N}$, $\delta \in [0, 1]$;
for $t = 1, \dots, T$ **do**
  Set $g_t \leftarrow \mathcal{O}_\eta(x_t, n)$ by taking a batch of measurements of size n at x_t;
  Compute lower bounds $\underline{\alpha}_t^i := \bar{\alpha}_t^i - \sigma_i(n)\sqrt{\ln \frac{1}{\delta}}$, $\forall i \in [m]$;
  Compute upper bounds $\hat{\theta}_t^i = |\langle G_n^i(x, \xi), \frac{g_t}{\|g_t\|}\rangle| + \hat{b}_i + \hat{\sigma}_i(n)\sqrt{\log \frac{1}{\delta}}$, $\forall i \in [m]$;
  Compute $\hat{M}_2(x_t)$ using (13);
  $\gamma_t \leftarrow \min\left\{ \min_{i \in [m]} \left\{ \frac{\underline{\alpha}_t^i}{2|\hat{\theta}_t^i| + \sqrt{\underline{\alpha}_t^i M_i}} \right\} \frac{1}{\|g_t\|}, \frac{1}{\hat{M}_2(x_t)} \right\}$;
  $x_{t+1} \leftarrow x_t - \gamma_t g_t$;
end for
Output: $\{x_1, \dots, x_T\}$.

In the above, the input parameters are: the smoothness constant M_i of each function f^i for $i \in \{0, \dots, m\}$, the bound on the variance of its value measurements σ_i^2, the bound on the variance of its gradient measurements $\hat{\sigma}_i^2$ and the upper bound on the bias of its gradient measurements $\hat{b}_i$, the bound on the diameter R of the set $\mathcal{X}$, the log barrier parameter η, the number of measurements per iteration n, the number of iterations T, and the confidence parameter δ.

3.4 Safety

From the safety side, the adaptive step-size γ_t automatically guarantees the safety of all the iterates due to construction, for any procedure generating the iterations in the form $x_{t+1} = x_t - \gamma_t g_t$ where γ_t is bounded by (15). We guarantee the feasibility of the optimization trajectory with probability at least $1 - \hat{\delta}$ with $\hat{\delta} := (2m + 1)T\delta$.

Theorem 4 *Let $T > 0$ denote the total number of iterations of the form (7), and $\hat{\delta} \in (0, 1)$ denote the target confidence level. Then, for LB-SGD with parameter $\delta \leq \hat{\delta}/(2m + 1)T$, all the query points x_t are feasible with probability greater than $1 - \delta$.*

Proof Due to the adaptive step size γ_t, we have $x_t \in \mathcal{X}$ implies $x_{t+1} \in \mathcal{X}$ (see Lemma 3) since $f^i(x_{t+1}) \leq \frac{f^i(x_t)}{2}$ with probability $1 - \delta$. Then, using $x_0 \in \mathcal{X}$ and Boole's inequality, we conclude that the whole optimization trajectory $\{x_t\}_{t\in[T]}$ is feasible with probability at least $1 - mT\delta \geq 1 - \hat{\delta}$. ■

4. Method Variants and Convergence Analysis

First of all, let us show the following general property of the log barrier method, important for the further convergence analysis of any problem type that we discuss.

4.1 Keeping a distance away from the boundary

Suppose that $\underline{\alpha}_t^i$ becomes 0 for some iteration t during the learning. That would lead to $\gamma_t = 0$, and the algorithm will stop without converging, since there is no safe non-zero step-size. Moreover, the log barrier gradient at that point simply blows up. However, we can lower-bound the step sizes γ_t if we can provide a lower bound on $\underline{\alpha}_t^i$ for all $t > 0$ during the learning with high probability:

Lemma 5 *If $\min_{i\in[m]} \underline{\alpha}_t^i \geq c\eta$ for $c > 0$, then we have $\gamma_t \geq C\eta$ with C defined by*

$$C := \frac{c}{2L^2(1 + \frac{m}{c})} \frac{1}{\max\left\{4 + \frac{5Mc\eta}{L^2}, 1 + \sqrt{\frac{Mc\eta}{4L^2}}\right\}}. \tag{16}$$

The proof is shown in Appendix A.6.

Therefore, for convergence, we need to show that our algorithm's iterates x_t do not only stay inside the feasible set, but moreover keep a distance away from the boundary. Keeping distance is the key property, guaranteeing the regularity of the log barrier function in the sense of a bounded gradient norm, bounded local smoothness and bounded variance. For the exact information case without noise, the adaptive gradient descent on the log barrier is shown to converge without stating this property explicitly (Hinder and Ye, 2018). However, in the stochastic case, this property becomes crucial for establishing stable convergence. It guarantees that the method pushes the iterates x_t away from the boundary of the set $\mathcal{X}$ as soon as they come too close to the boundary. We formulate it below.

Lemma 6 *Let Assumptions 2 and 3 hold, Assumption 4 hold with $\rho \geq \eta$, and let $\hat{\sigma}_i(n) \leq \frac{\alpha_t^i L}{\eta\sqrt{\ln\frac{1}{\delta}}}$, $\hat{b}_i \leq \frac{\alpha_t^i L}{2\eta}$, and $\sigma_i(n) \leq \frac{(\alpha_t^i)^2}{2\eta\sqrt{\ln\frac{1}{\delta}}}$. Then, we can show that for all x_t for all iterations $t \in [T]$ generated by the optimization process $x_{t+1} = x_t - \gamma_t g_t$ the following holds:*

$$\mathbb{P}\{\forall t \in [T] \ \min_{i\in[m]} \underline{\alpha}_t^i \geq c\eta\} \geq 1 - \hat{\delta}, \tag{17}$$

with $c := \frac{1}{2}\left(\frac{l}{4L(2m+1)}\right)^m$, where $l > 0$ is defined as in Assumption 4.

Proof First, let us note the following fact demonstrating that the product of the smallest absolute constraint values is not decreasing if x_t is close enough to the boundary.

Fact 1 *Let Assumptions 2, 3 hold, Assumption 4 hold with $\rho \geq \eta$, and let $\hat{\sigma}_i(n) \leq \frac{\alpha_t^i L}{2\eta\sqrt{\ln \frac{1}{\delta}}}$, $\hat{b}_i \leq \frac{\alpha_t^i L}{2\eta}$, and $\sigma_i(n) \leq \frac{(\alpha_t^i)^2}{2\eta\sqrt{\ln \frac{1}{\delta}}}$. If at iteration t we have $\min_{i\in[m]} \alpha_t^i \leq \bar{c}\eta$ with $\bar{c} := \frac{l}{L}\frac{1}{2m+1}$, then, for the next iteration $t+1$ we get $\prod_{i\in\mathcal{I}} \alpha_{t+1}^i \geq \prod_{i\in\mathcal{I}} \alpha_t^i$ for any $\mathcal{I} : \mathcal{I}_t \subseteq \mathcal{I}$ with $\mathcal{I}_t := \{i \in [m] : \alpha_t^i \leq \eta\}$ with probability $1-\delta$.*

For the proof see Appendix A.5.

Note that if $\min_{i\in[m]} \alpha_t^i \geq \bar{c}\eta$ for all $t \geq 0$, then the statement of the Lemma holds automatically. Now, consider a consecutive set of steps $t = \{t_0, \dots, t_k\}$ on whose $\min_{i\in[m]} \alpha_t^i \leq \bar{c}\eta$. By definition, and using Fact 1, for any $t \in \{t_0, \dots, t_k\}$ we have with probability $1-\delta$

$$\prod_{i\in\mathcal{I}_{t+1}} \alpha_{t+1}^i = \frac{\prod_{i\in\mathcal{I}_t\cup\mathcal{I}_{t+1}} \alpha_{t+1}^i}{\prod_{i\in\mathcal{I}_t\setminus\mathcal{I}_{t+1}} \alpha_{t+1}^i} \geq \frac{\prod_{i\in\mathcal{I}_t\cup\mathcal{I}_{t+1}} \alpha_t^i}{\prod_{i\in\mathcal{I}_t\setminus\mathcal{I}_{t+1}} \alpha_{t+1}^i}.$$

By induction, applying the above sequentially for all $t \in \{t_0, \dots, t_k\}$ we can get

$$\prod_{i\in\mathcal{I}_{t_k}} \alpha_{t_k}^i \geq \frac{\prod_{i\in\mathcal{I}_{t_k}\cup\mathcal{I}_{t_{k-1}}\cup\ldots\cup\mathcal{I}_{t_0}} \alpha_{t_0}^i}{\prod_{i\in\mathcal{I}_{t_{k-1}}\setminus\mathcal{I}_{t_k}} \alpha_{t_k}^i \prod_{i\in\mathcal{I}_{t_{k-2}}\setminus(\mathcal{I}_{t_k}\cup\mathcal{I}_{t_{k-1}})} \alpha_{t_{k-1}}^i \cdots \prod_{i\in\mathcal{I}_{t_0}\setminus(\mathcal{I}_{t_k}\cup\ldots\cup\mathcal{I}_{t_1})} \alpha_{t_1}^i}$$

with probability $1-\hat{\delta}$ (using Boole's inequality).

Note that by definition of $\mathcal{I}_t$: $\alpha_t^i \leq \eta$ for $i \in \mathcal{I}_t$. At the same time, due to the step size choice, we have $\alpha_{t+1}^i \leq 2\alpha_t^i \leq 2\eta$. Also, note that the sum of the set cardinalities in the denominator equals to the cardinality of the set $|\mathcal{I}_{t_0} \cup \ldots \cup \mathcal{I}_{t_k} \setminus \mathcal{I}_{t_k}|$. Hence, with probability $1-\hat{\delta}$

$$\prod_{i\in\mathcal{I}_{t_k}} \alpha_{t_k}^i \geq \frac{\prod_{i\in\mathcal{I}_{t_k}\cup\mathcal{I}_{t_{k-1}}\cup\ldots\cup\mathcal{I}_{t_0}} \alpha_{t_0}^i}{(2\eta)^{|\mathcal{I}_{t_0}\cup\ldots\cup\mathcal{I}_{t_k}\setminus\mathcal{I}_{t_k}|}}.$$

Thus, for any $j \in \mathcal{I}_{t_k}$ we get the bound:

$$\alpha_{t_k}^j \geq \frac{\prod_{i\in\mathcal{I}_{t_k}\cup\mathcal{I}_{t_{k-1}}\cup\ldots\cup\mathcal{I}_{t_0}} \alpha_{t_0}^i}{(2\hat{c}\eta)^{|\mathcal{I}_{t_0}\cup\ldots\cup\mathcal{I}_{t_k}\setminus j|}} \geq \frac{(\bar{c}\eta/2)^{|\mathcal{I}_{t_0}\cup\ldots\cup\mathcal{I}_{t_k}|}}{2^{|\mathcal{I}_0\cup\ldots\cup\mathcal{I}_{t_k}|-1}} \geq \left(\frac{\bar{c}}{4}\right)^m \eta.$$

Since $\sigma_i(n) \leq \frac{\alpha_t^i}{2\sqrt{\ln \frac{1}{\delta}}}$, we get that

$$\underline{\alpha}_t^i \geq \frac{\alpha_t^i}{2} \geq \frac{1}{2}\left(\frac{\bar{c}}{4}\right)^m \eta \tag{18}$$

with probability $1-\delta$. Using the definition $\bar{c} := \frac{l}{L}\frac{1}{2m+1}$, we obtain the statement of the lemma. ■

4.2 Stochastic non-convex problems

For the non-convex problem we analyse LB-SGD(x_0, η, T, n) with the fixed parameter η, that uses the stopping criterion

$$\|g_t\| \leq 3\eta/4 \tag{19}$$

and outputs $x_{\hat{t}}$ with $\hat{t}$ corresponding to $\arg\min_{t \in T} \|g_t\|$.

4.2.1 Stationarity criterion in the non-convex case

Similarly to Usmanova et al. (2020), we can generally state case small gradient of the log barrier with parameter η leads to an η-approximate KKT point of the constrained problem. Let us set the pair of primal and dual variables to $(x, \lambda) := \left(x, \left[\frac{\eta}{-f^1(x)}, \dots, \frac{\eta}{-f^m(x)}\right]^T\right)$. Then, it satisfies:

$$\begin{aligned}
&1)\ \|\nabla_x \mathcal{L}(x, \lambda)\| = \|\nabla B_\eta(x)\|;\\
&2)\ \lambda^i(-f^i(x)) = \frac{\eta}{-f^i(x)}(-f^i(x)) = \eta;\\
&3)\ \lambda^i \geq 0, -f^i(x) \geq 0,\ i \in [m].
\end{aligned}$$

This insight immediately implies the following Lemma.

Lemma 7 *Consider problem (P) under Assumptions 2, and 3. Let $\hat{x}$ be an η-approximate solution to $\min_{x \in \mathbb{R}^d} B_\eta(x)$, the η-log barrier approximation of (P), such that $\|\nabla B_\eta(\hat{x})\| \leq \eta$. Then, $\hat{x}$ is an η-approximate KKT point to the original problem (P).*

Thus, we can use a small log barrier norm to guarantee stationarity for problem (P).

4.2.2 Convergence for the non-convex problem

Then, we get the following convergence result:

Theorem 8 *After at most T iterations of LB-SGD with $T \leq 11\frac{B_\eta(x_0) - \min_x B_\eta(x)}{C\eta^3}$, and with $\sigma_i(n) = O(\eta^2)$, $\hat{\sigma}_i(n) = O(\eta)$, and $\hat{b}_i = O(\eta)$, for the output $x_{\hat{t}}$ with $\hat{t} = \arg\min_{t \in T} \|g_t\|$ we have*

$$\mathbb{P}\left\{\|\nabla B_\eta(x_{\hat{t}})\| \leq \eta\right\} \geq 1 - \hat{\delta} \tag{20}$$

Therefore, given $\sigma_i(n) = \frac{\sigma_i}{\sqrt{n}}$ (3) and $\hat{\sigma}_i(n) = \frac{\hat{\sigma}_i}{\sqrt{n}}$ (4), for constant $\hat{\sigma}_i, \sigma_i$, we require $n = O(\frac{1}{\eta^4})$ oracle calls per iteration, and $N = O(\frac{1}{\eta^7})$ calls of the first-order stochastic oracle in total. Using Lemma 7, we get that $x_{\hat{t}}$ is an ε-approximate KKT point to the original problem (P) with $\varepsilon = \eta$.

Remark *Lower bound in the unconstrained non-safe case.* In a well-known model where algorithms access smooth, non-convex functions through queries to an unbiased stochastic gradient oracle with bounded variance, Arjevani et al. (2019) prove that in the worst case, any algorithm requires at least ε^{-4} queries to find an ε stationary point. Although, they allow d to depend on ε. Therefore, we "pay" extra ε^{-3} measurements for safety. From the methodology point of view, this happens due to the non-smoothness of the log-barrier on the boundary and the fact that the noise of the barrier gradient estimator is very sensitive to how close the iterates x_t are to the boundary.

Proof

Step 1. Probabilistic events. In the current theorem, we derive the convergence rate with probability $1-(2m+1)\delta = 1-\hat{\delta}$. For that, we prove the convergence under the condition of a particular event $\mathcal{A}_T$, which holds with probability $1-\hat{\delta}$. Let us first define the following events $\mathcal{A}_t^i$ of $1-\delta$-confidence intervals covering the true value and the true gradient of constraint functions f^i at step t respectively, which holds with probability $\mathbb{P}\{\mathcal{A}_t^i\} \geq 1-2\delta$:

$$\mathcal{A}_t^i := \mathbb{I}\left\{\left|\alpha_t^i - \bar{\alpha}_t^i\right| \leq \sigma_i(n)\sqrt{\ln\frac{1}{\delta}} \text{ and } \left\|G_n^i(x_t) - \nabla f^i(x_t)\right\| \leq \hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}}\right\}. \tag{21}$$

For the objective, we denote $\mathcal{A}_t^0 := \mathbb{I}\left\{\left\|G_n^0(x_t) - \nabla f^0(x_t)\right\| \leq \hat{b}_0 + \hat{\sigma}_0(n)\sqrt{\ln\frac{1}{\delta}}\right\}$, which holds with probability $\mathbb{P}\{\mathcal{A}_t^0\} \geq 1-\delta$. We also denote $\mathcal{A}_t := \cap_{i\in\{0,\dots,m\}}\mathcal{A}_t^i$, which holds with probability at least $1-(2m+1)\delta$, using the Boole's inequality. And finally, we denote an event $\mathcal{A}_T := \cap_{t\leq T}\mathcal{A}_t$, which holds with probability at least $1-(2m+1)T\delta = 1-\hat{\delta}$, again using the Boole's inequality.

Note that event $\mathcal{A}_T$ almost surely implies that for all t: 1) $M_2(x_t) \leq \hat{M}_2(x_t)$, and therefore $\gamma_t \leq \frac{1}{M_2(x_t)}$. 2) It also almost surely implies the upper bound on $\|\Delta_t\|$ from Lemma 1 (9). 3) It guarantees the "keeping distance" property $\min_{i\in[m]} \underline{\alpha}_t^i \geq c\eta$ (17) with probability 1, that can be seen from the proof of Lemma 6 (due to $\underline{\alpha}_t^i \geq \frac{\alpha_t^i}{2}$ with probability 1 in (18) given that $\mathcal{A}_T$ holds.) 4) And therefore, it implies the lower bound $\gamma_t \geq C\eta$ (Lemma 3). For the rest of the proof steps below we assume that event $\mathcal{A}_T$ holds, hence we will use all four statements made above.

Step 2. Bounding number of iterations T. First, let us denote $\hat{\gamma}_t := \gamma_t\|g_t\|$. At each iteration of Algorithm 2 with the fixed η the value of the logarithmic barrier decreases at least by the following value:

$$\begin{aligned} B_\eta(x_t) - B_\eta(x_{t+1}) &\overset{1}{\geq} \gamma_t\left\langle\nabla B_\eta(x_t), g_t\right\rangle - \frac{1}{2}M_2(x_t)\gamma_t^2\|g_t\|^2 \\ &= \hat{\gamma}_t\langle\Delta_t, \frac{g_t}{\|g_t\|}\rangle + \hat{\gamma}_t\left(1 - \frac{M_2(x_t)\gamma_t}{2}\right)\|g_t\| \\ &\overset{2}{\geq} \frac{1}{2}\hat{\gamma}_t\|g_t\| - \hat{\gamma}_t\|\Delta_t\| \end{aligned} \tag{22}$$

In the above, $M_2(x)$ is the local smoothness constant that we bound by (13). The first inequality 1 is due to the local smoothness of the barrier. 2 is due to the fact that $\mathbb{P}\{\gamma_t \leq \frac{1}{M_2(x_t)}\} \geq 1-\delta$, given $x_t \in Int(\mathcal{X})$. Note that under the condition of event $\mathcal{A}_T$, the above holds with probability 1. Summing up the above inequalities (22) for $t \in [T]$, we obtain the second inequality below:

$$T\min_{t\in[T]}\hat{\gamma}_t\left(\frac{1}{2}\|g_t\| - \|\Delta_t\|\right) \leq \sum_{t\in[T]}\hat{\gamma}_t\left(\frac{1}{2}\|g_t\| - \|\Delta_t\|\right) \leq B_\eta(x_0) - \min_{x\in X} B_\eta(x).$$

In the above, the first inequality is due to the fact that the minimum of summands is smaller than any of the summands. Recall that we stop the algorithm as soon as $\|g_t\| \leq 3\eta/4$, as stated in the beginning of the section in (19). Hence, for all T iterations $t \in [T]$ with $\|g_t\| \geq 3\eta/4$ we have $\hat{\gamma}_t \geq 0.75\eta\gamma_t$. Therefore, we get:

$$T \leq \frac{B^\eta(x_0) - \min_x B^\eta(x)}{\min_{t\in[T]}\{\hat{\gamma}_t\,(0.5\|g_t\| - \|\Delta_t\|)\}} \leq \frac{B^\eta(x_0) - \min_x B^\eta(x)}{0.75\eta\min_{t\in[T]}\{\gamma_t\,(0.5\|g_t\| - \|\Delta_t\|)\}}. \tag{23}$$

To obtain a lower bound on the denominator, we use the result of Lemmas 5 and 6, under the condition of event $\mathcal{A}_T$, giving $\gamma_t \geq C\eta$ for all $t \in [T]$.

Step 3. Bounding $\|\Delta_t\|$**.** Next, we have to upper bound $\|\Delta_t\|$ with high probability. Recall from Lemma 1 (9):

$$\mathbb{P}\left\{\|\Delta_t\| \leq b_0 + \hat{\sigma}_0(n)\sqrt{\ln\frac{1}{\delta}} + \sum_{i=1}^m \frac{\eta}{\bar{\alpha}_t^i}\left(\hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}}\right) + \sum_{i=1}^m L_i\frac{\eta\sigma_i(n)}{\alpha_t^i\bar{\alpha}_t^i}\sqrt{\ln\frac{1}{\delta}}\right\} \geq 1-\delta.$$

The above holds with probability 1 given $\mathcal{A}_T$. Hence, we get $\|\Delta_t\| \leq \frac{\eta}{4}\forall t$ given $\mathcal{A}_T$ if for all $i \in [m]$:

$$\hat{\sigma}_0(n) \leq \frac{\eta}{4(3m+2)\sqrt{\ln\frac{1}{\delta}}}, \hat{\sigma}_i(n) \leq \frac{\underline{\alpha}_t^i}{4(3m+2)\sqrt{\ln\frac{1}{\delta}}}, \tag{24}$$

$$\hat{b}_0 \leq \frac{\eta}{4(3m+2)}, \hat{b}_i \leq \frac{\underline{\alpha}_t^i}{4(3m+2)}, \sigma_i(n) \leq \frac{(\underline{\alpha}_t^i)^2}{4(3m+2)L\sqrt{\ln\frac{1}{\delta}}}. \tag{25}$$

Using the lower bound on $\underline{\alpha}_t^i$ by Lemma 6, we get that for all $i \in [m]$ we require $\hat{\sigma}_i(n) = O(\eta)$, $\sigma_i(n) = O(\eta^2)$, $\hat{b}_i = O(\eta)$.

Step 4. Finalising the bounds. Then, using $\|g_t\| \leq 3\eta/4$ for all $t \in [T]$, we can claim that given $\mathcal{A}_T$ for any $t \in [T]$ the following holds

$$0.5\|g_t\| - \|\Delta_t\| \geq \frac{3}{8}\eta - \frac{1}{4}\eta = \frac{\eta}{8}.$$

Combining it with inequality (23), the algorithm stops after at most T iterations with

$$T \leq 8\frac{B_\eta(x_0) - \min_x B_\eta(x)}{0.75C\eta^3} \leq 11\frac{B_\eta(x_0) - \min_x B_\eta(x)}{C\eta^3}.$$

Finally, using the fact that $\|\Delta_t\| \leq \frac{\eta}{4}\forall t$ given $\mathcal{A}_T$, also using $\mathbb{P}\{\mathcal{A}_T\} \geq 1-\hat{\delta}$, and the stopping criterion $\|g_t\| \leq 3\eta/4$ (19), we obtain

$$\mathbb{P}\left\{\|\nabla B_\eta(x_{\hat{t}})\| \leq \|g_{\hat{t}}\| + \|\Delta_{\hat{t}}\| \leq \eta\right\} \geq 1-\hat{\delta}.$$

■

4.3 Stochastic convex problems

For the convex case, we propose to use LB-SGD(x_0, η, T, n) with the *output:* $\bar{x}_T := \frac{\sum_{t=1}^T \gamma_t x_t}{\sum_{t=1}^T \gamma_t}$. Next, we discuss the optimality criterion for convex problems.

4.3.1 Optimality criterion in the convex case

In the convex case, we can relate an approximate solution of the log barrier problem to an ε-approximate solution of the original problem in terms of the objective value.

Assumption 5 *The objective and the constraint functions $f^i(x)$ for all $i \in \{0, \dots, m\}$ are convex.*

Recall that convexity of functions $f^i(x)$ also implies the convexity of $-\log(-f^i(x))$ over the domain $\mathcal{X}$ simply noting that $\log(\cdot)$ is monotonously increasing function, and $-f^i(x)$ is concave. Therefore, we have the convexity of the logarithmic barrier $B_\eta(x)$.

Also, note that Assumption 3 implies non-emptiness on $Int(\mathcal{X})$ which is called Slater Constraint Qualification. In the convex setting, it in turn implies the extended MFCQ:

Fact 2 *Let Assumptions 1, 3, and 5 hold. Then, Assumption 4 holds with $s_x := \frac{x-x_0}{\|x-x_0\|}$, such that $\langle \nabla f^i(x), s_x \rangle \geq \frac{\beta-\rho}{R}$ for all $i \in \mathcal{I}_\rho(x)$ for any $0 < \rho < \beta$.*

Proof Indeed, for any point $x \in \mathcal{X}$ and for any convex constraint f^i such that $f^i(x) \geq -\rho$, due to convexity we have $f^i(x) - f^i(x_0) \leq \langle \nabla f^i(x), x - x_0 \rangle$. Given the bounded diameter of the set $\|x_0 - x\| \leq R$, we get $\langle \nabla f^i(x), s_x \rangle \geq \frac{\beta-\rho}{R}$. ■

Then, we can relate an η-approximate solution by the log barrier value with an ε-approximate solution for the original problem, where ε depends on η linearly up to a logarithmic factor. We formulate that in the following lemma:

Lemma 9 *Consider problem (P) under Assumptions 1, 2, 3, and the convexity Assumption 5. Assume that $\hat{x}$ is an η-approximate solution to the η-log barrier approximation, that is,*

$$B_\eta(\hat{x}) - B_\eta(x^*_\eta) \leq \eta,$$

where x^*_η is a solution of the $\min B_\eta$ minimization problem, with $\eta \leq \beta/2$. Then, $\hat{x}$ is an ε-approximate solution to the original problem (P) with $\varepsilon = \eta(m+1) + \eta m \log\left(\frac{2mLR\hat{\beta}}{\eta\beta}\right)$, that is, $f^0(\hat{x}) - \min_{x\in\mathcal{X}} f^0(x) \leq \varepsilon$, where $\hat{\beta} > 0$ is such that $\forall i \in [m] \forall x \in \mathcal{X}\ |f^i(x)| \leq \hat{\beta}$. Since the constraints are smooth and the set $\mathcal{X}$ is bounded, such $\hat{\beta}$ exists.

Proof sketch Let $\hat{x}$ be an approximately optimal point for the log barrier: $B_\eta(\hat{x}) - B_\eta(x^*_\eta) \leq \eta$, and x^*_η be an optimal point for the log barrier. Then, using the definition, we can bound: $f^0(\hat{x}) - f^0(x^*_\eta) \leq \eta + \eta \sum_{i=1}^m -\log\frac{-f^i(x^*_\eta)}{-f^i(\hat{x})}$. Combining Fact 2 with the first order stationarity criterion, we can derive: $\min_{i\in[m]}\{-f^i(x^*_\eta)\} \geq \frac{\eta\beta}{2mLR}$. Hence, combining the above two inequalities, we get the following relation of point $\hat{x}$ and point x^*_η: $f^0(\hat{x}) - f^0(x^*_\eta) \leq \eta\left(1 + m\log\left(\frac{2mLR\hat{\beta}}{\eta\beta}\right)\right)$ using $-f^i(\hat{x}) \leq \hat{\beta}$. Using the Lagrangian definition for stationarity of the optimal point of the initial problem x^*, we get the following relation between x^* and x^*_η: $f^0(x^*_\eta) - f^0(x^*) \leq m\eta$. Combining it with the above, we get the statement of the Lemma

$$f^0(\hat{x}) - \min_{x\in\mathcal{X}} f^0(x) \leq \eta + \eta m \log\left(\frac{2mLR\hat{\beta}}{\eta\beta}\right) + m\eta.$$

For the full proof see Appendix A.7.

4.3.2 Convergence in the convex case

As already discussed in the optimality criterion Section 4.3.1, for the convex problem we only require the convergence in terms of the value of the log barrier. Thus, we get the following convergence result for this method.

Theorem 10 *Let Assumptions 1, 2, 3, 5 hold, $B_\eta(x) := f^0(x) - \eta\sum_{i=1}^m \log(-f^i(x))$ be a log barrier function with parameter $\eta > 0$, and $x_0 \in \mathbb{R}^d$ be the starting point. Let x^*_η be a minimizer of $B_\eta(x)$. Then, after $T \geq \frac{\|x_0 - x^*_\eta\|^2}{C\eta^2}$ iterations of LB-SGD, and with $\sigma_i(n) = O(\frac{\eta^2}{R})$, $\hat{\sigma}_i(n) = O(\frac{\eta}{R})$, and $\hat{b}_i = O(\frac{\eta}{R})$, for the point $\bar{x}_T := \frac{\sum_{t=1}^T \gamma_t x_t}{\sum_{t=1}^T \gamma_t}$ we obtain:*

$$\mathbb{P}\left\{B_\eta(\bar{x}_T) - B_\eta(x^*_\eta) \leq \eta\right\} \geq 1 - \hat{\delta},$$

where $\hat{\delta} = 2mT\delta$. For the noise with constant variances $\hat{\sigma}_i, \sigma_i$, given $\sigma_i(n) = \frac{\sigma_i}{\sqrt{n}}$ (3) and $\hat{\sigma}_i(n) = \frac{\hat{\sigma}_i}{\sqrt{n}}$ (4), we require $n = O(\frac{1}{\eta^4})$ oracle calls per iteration, and $O(\frac{1}{\eta^6})$ measurements of the first-order oracle in total. Using Lemma 9 we get $\bar{x}_T$ is an ε-approximate solution to the original problem (P) with $\varepsilon = \eta(m+1) + \eta m \log\left(\frac{2mLR\hat{\beta}}{\eta\beta}\right)$, that is, $f^0(\hat{x}) - \min_{x\in\mathcal{X}} f^0(x) \leq \varepsilon$.

Proof Similarly to the proof of the previous theorem, we use the notion of event $\mathcal{A}^i_t$ as defined in Eq. (21), we denote $\mathcal{A}_T := \cap_{t\leq T}\cap_{i\in[m]}\mathcal{A}_t$, which holds with probability at least $1 - 2mT\delta = 1 - \hat{\delta}$. For the rest of the proof, we assume that event $\mathcal{A}_T$ holds. Note the following

$$\begin{aligned}
B_\eta(x_{t+1}) - B_\eta(x^*_\eta) &\overset{①}{\leq} B_\eta(x_t) + \langle \nabla B_\eta(x_t), x_{t+1} - x_t\rangle + \frac{M_2(x_t)}{2}\|x_t - x_{t+1}\|^2 - B_\eta(x^*_\eta) \\
&\overset{②}{\leq} \langle \nabla B_\eta(x_t), x_{t+1} - x_t\rangle + \frac{M_2(x_t)}{2}\|x_t - x_{t+1}\|^2 + \langle \nabla B_\eta(x_t), x_t - x^*_\eta\rangle \\
&= \frac{M_2(x_t)}{2}\|x_t - x_{t+1}\|^2 + \langle g_t, x_{t+1} - x^*_\eta\rangle - \langle \Delta_t, x_{t+1} - x^*_\eta\rangle \\
&\overset{③}{\leq} \frac{\|x_t - x^*_\eta\|^2}{2\gamma_t} - \frac{\|x_{t+1} - x^*_\eta\|^2}{2\gamma_t} - \left(\frac{1}{2\gamma_t} - \frac{M_2(x_t)}{2}\right)\|x_{t+1} - x_t\|^2 - \langle \Delta_t, x_{t+1} - x^*_\eta\rangle \\
&\overset{④}{\leq} \frac{\|x_t - x^*_\eta\|^2}{2\gamma_t} - \frac{\|x_{t+1} - x^*_\eta\|^2}{2\gamma_t} - \langle \Delta_t, x_{t+1} - x^*_\eta\rangle.
\end{aligned}$$

The first inequality ① is due to the $M_2(x_t)$-local smoothness of the log barrier, the second one ② is due to convexity. The third inequality ③ uses the fact that: $\forall u \in \mathbb{R}^d : \langle g_t, x_{t+1} - u\rangle = \frac{\|x_t - u\|^2}{2\gamma_t} - \frac{\|x_{t+1} - u\|^2}{2\gamma_t} - \frac{\|x_{t+1} - x_t\|^2}{2\gamma_t}$. And the last one ④ is due to $\gamma_t \leq \frac{1}{M_2(x_t)}$. By multiplying both sides by γ_t, we get:

$$2\gamma_t(B_\eta(x_{t+1}) - B_\eta(x^*_\eta)) \leq \|x_t - x^*_\eta\|^2 - \|x_{t+1} - x^*_\eta\|^2 - 2\gamma_t\langle \Delta_t, x_{t+1} - x^*_\eta\rangle. \tag{26}$$

Then, by summing up the above for all $t \in [T]$ we get, and using the Jensen's inequality:

$$\begin{aligned}
B_\eta\left(\frac{\sum_{t=1}^T \gamma_t x_t}{\sum \gamma_t}\right) - B_\eta(x_\eta^*) &\leq \frac{1}{\sum_{t=1}^T \gamma_t} \sum_{t=1}^T \gamma_t (B_\eta(x_t) - B_\eta(x_\eta^*)) \\
&\leq \frac{1}{2\sum_{t=1}^T \gamma_t} \sum_{t=1}^T \left(\|x_t - x_\eta^*\|^2 - \|x_{t+1} - x_\eta^*\|^2 - 2\gamma_t \langle \Delta_t, x_{t+1} - x_\eta^* \rangle\right) \\
&\leq \frac{\|x_0 - x_\eta^*\|^2}{2\sum_{t=1}^T \gamma_t} - \frac{\sum_{t=1}^T \gamma_t \langle \Delta_t, x_{t+1} - x_\eta^* \rangle}{2\sum_{t=1}^T \gamma_t}. \qquad (27)
\end{aligned}$$

That is, we can bound the accuracy by

$$B_\eta\left(\bar{x}_T\right) - B_\eta(x_\eta^*) \leq \frac{\|x_0 - x_\eta^*\|^2}{2\sum_{t=1}^T \gamma_t} + \frac{\max_t \langle \Delta_t, x_{t+1} - x_\eta^* \rangle}{2}. \qquad (28)$$

Using Lemma 5 we can prove for $\hat{\sigma}_i(n) \leq \frac{L\alpha_t^i}{3\eta\sqrt{\ln\frac{1}{\delta}}}$ that $\gamma_t \geq C\eta$. Recall from Lemma 1 (9):

$$\mathbb{P}\left\{\|\Delta_t\| \leq b_0 + \hat{\sigma}_0(n)\sqrt{\ln\frac{1}{\delta}} + \sum_{i=1}^m \frac{\eta}{\bar{\alpha}_t^i}\left(\hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}}\right) + \sum_{i=1}^m L_i \frac{\eta\sigma_i(n)}{\alpha_t^i \bar{\alpha}_t^i}\sqrt{\ln\frac{1}{\delta}}\right\} \geq 1 - \delta.$$

Hence, we can guarantee $\|\Delta_t\| \leq \frac{\eta}{R}$ for all $t \in [T]$ with probability $1 - \hat{\delta}$ if for all $i \in [m]$

$$\begin{aligned}
\hat{\sigma}_0(n) &\leq \frac{\eta}{(3m+2)R\sqrt{\ln\frac{1}{\delta}}}, \hat{\sigma}_i(n) \leq \frac{\underline{\alpha}_t^i}{(3m+2)R\sqrt{\ln\frac{1}{\delta}}}, \qquad (29) \\
\hat{b}_0 &\leq \frac{\eta}{(3m+2)R}, \hat{b}_i \leq \frac{\underline{\alpha}_t^i}{(3m+2)R}, \sigma_i(n) \leq \frac{(\underline{\alpha}_t^i)^2}{(3m+2)LR\sqrt{\ln\frac{1}{\delta}}}.
\end{aligned}$$

Using the lower bound on $\underline{\alpha}_t^i$ by Lemma 6, we get that for all $i \in [m]$ we require $\hat{\sigma}_i(n) = O(\eta)$, $\sigma_i(n) = O(\eta^2)$, $\hat{b}_i = O(\eta)$.
Therefore, we get for the $\bar{x}_T$ the following bound on the accuracy:

$$B_\eta\left(\bar{x}_T\right) - B_\eta(x_\eta^*) \leq \frac{\|x_0 - x_\eta^*\|^2}{TC\eta} + \frac{\max_t \|\Delta_t\| R}{2} \leq \frac{\|x_0 - x_\eta^*\|^2}{TC\eta} + \frac{\eta}{2}. \qquad (30)$$

Thus, for $T \geq \frac{\|x_0 - x_\eta^*\|^2}{2C\eta^2}$ we obtain $\mathbb{P}\{B_\eta\left(\bar{x}_T\right) - B_\eta(x_\eta^*) \leq \eta\} \geq 1 - \hat{\delta}$. In order to satisfy conditions on the variance (29), we require at each iteration $n = O\left(\frac{1}{\eta^4}\right)$ measurements, and therefore $N = Tn = O(\frac{1}{\eta^6})$ measurements in total. ■

4.4 Strongly-convex problems

For strongly-convex problems, we can get the following convergence result.

Theorem 11 *Let Assumptions 1, 2, 3, 5 hold, and the log barrier function with parameter η: $B_\eta(x) := f^0(x) - \eta \sum_{i=1}^m \log(-f^i(x))$ be μ-strongly-convex. We use $\|g_t\| \leq \frac{2\sqrt{\mu\eta}}{4.5}$ as a stopping criterion. Then, after at most $T = \frac{4}{\mu C\eta} \log \frac{2(B_\eta(x_0) - B_\eta(x^*))}{\eta}$ iterations of LB-SGD with constant η (Algorithm 4), and with $\sigma_i(n) = O(\eta^{1.5})$, $\hat{b}_i = O(\sqrt{\eta})$, $\hat{\sigma}_i(n) = O(\sqrt{\eta})$, we obtain:*

$$\mathbb{P}\left\{B_\eta(\hat{x}_K) - \min_{x\in\mathcal{X}} B_\eta(x) \leq \eta\right\} \geq 1 - \hat{\delta},$$

where $\hat{\delta} = 2mT\delta$. Hence, we require $n = O(\frac{1}{\eta^3})$ measurements per iteration, and $N = \tilde{O}(\frac{1}{\eta^4})$ measurements of the first-order oracle in total. Using Lemma 9, we obtain that $\hat{x}$ is an ε-approximate solution to the original problem (P) with $\varepsilon = \eta(m+1) + \eta m \log\left(\frac{2mLR\hat{\beta}}{\eta\beta}\right)$.

Proof The proof uses techniques similar to Ni and Kamgarpour (2023).
First, similarly to the proof of the previous theorems, we use notion of event $\mathcal{A}_t^i$ as defined in (21):

$$\mathcal{A}_t^i := \mathbb{I}\left\{\left|\alpha_t^i - \bar{\alpha}_t^i\right| \leq \sigma_i(n)\sqrt{\ln\frac{1}{\delta}} \text{ and } \left\|G_n^i(x_t) - \nabla f^i(x_t)\right\| \leq \hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}}\right\}.$$

We denote $\mathcal{A}_T := \cap_{t\leq T} \cap_{i\in[m]} \mathcal{A}_t$, which holds with probability at least $1 - 2mT\delta = 1 - \hat{\delta}$. For the rest of the proof, we assume that event $\mathcal{A}_T$ holds.
Step 1. Improvement for a single iteration. Using $M_2(x_t)$ local smoothness of the log barrier we can lower bound the $B_\eta(x_t) - B_\eta(x_{t+1})$ by:

$$\begin{aligned}
&B_\eta(x_t) - B_\eta(x_{t+1}) \geq \gamma_t\langle\nabla B_\eta(x_t), g_t\rangle - \frac{M_2(x_t)}{2}\gamma_t^2\|g_t\|^2 \\
&\geq \gamma_t\langle\nabla B_\eta(x_t), \nabla B_\eta(x_t) + \Delta_t\rangle - \frac{\gamma_t}{2}\|\nabla B_\eta(x_t) + \Delta_t\|^2 - \frac{\gamma_t}{2}\|\nabla B_\eta(x_t)\|^2 + \frac{\gamma_t}{2}\|\nabla B_\eta(x_t)\|^2 \\
&= \frac{\gamma_t}{2}\|\nabla B_\eta(x_t)\|^2 - \frac{\gamma_t}{2}\|\Delta_t\|^2, \qquad (31)
\end{aligned}$$

where the second inequality we get by adding and subtracting $\frac{\gamma_t}{2}\|\nabla B_\eta(x_t)\|^2$ and using $\gamma_t \leq \frac{1}{M_2(x_t)}$.
Step 2. Case when the stopping criterion is not reached at any iteration $t \leq T$. Since we use $\|g_t\| \leq \frac{2\sqrt{\mu\eta}}{4.5}$ as the stopping criterion, for all time steps $t \leq T$ we can assume $\|g_t\| > \frac{2\sqrt{\mu\eta}}{4.5}$. In this case, using that with high probability the deviation is bounded by $\|\Delta_t\| \leq \frac{\sqrt{\mu\eta}}{4.5}$, we conclude $\|\nabla B_\eta(x_t)\| \geq \|g_t\| - \|\Delta_t\| \geq \frac{\sqrt{\mu\eta}}{4.5}$, and consequently $\|\nabla B_\eta(x_t)\|^2 - \|\Delta_t\|^2 \geq 0$. Therefore, we can further write (31) as follows:

$$B_\eta(x_t) - B_\eta(x_{t+1}) \geq \frac{\gamma_t}{2}\left(\|\nabla B_\eta(x_t)\|^2 - \|\Delta_t\|^2\right) \geq \frac{C\eta}{2}\left(\|\nabla B_\eta(x_t)\|^2 - \frac{\mu\eta}{20}\right), \qquad (32)$$

where we also used $\gamma_t \geq C\eta$ given that event $\mathcal{A}_T$ holds (see Lemma 5). From strong convexity we have $\frac{1}{2}\|\nabla B_\eta(x_t)\|^2 \geq \mu(B_\eta(x_t) - B_\eta(x^*))$. By adding and subtracting $B_\eta(x^*)$ we get $(B_\eta(x_t) - B_\eta(x^*)) - (B_\eta(x_{t+1}) - B_\eta(x^*)) \geq \frac{\mu C\eta}{2}(B_\eta(x_t) - B_\eta(x^*)) - \frac{\mu C\eta^2}{40}$. The above directly implies

$$B_\eta(x_{t+1}) - B_\eta(x^*) \leq \left(1 - \frac{\mu C\eta}{2}\right)(B_\eta(x_t) - B_\eta(x^*)) + \frac{\mu C\eta^2}{40}. \qquad (33)$$

By recursion of the above inequality over t, setting $\eta < \frac{2}{\mu C}$ and using that $\sum_i \frac{1}{z^i} \to \frac{1}{1-z}$, we get:

$$
\begin{aligned}
B_\eta(x_T) - B_\eta(x^*) &\leq \left(1 - \frac{\mu C \eta}{2}\right)^T (B_\eta(x_0) - B_\eta(x^*)) + \frac{\mu C \eta^2}{40} \sum_{i=0}^{T-1} \left(1 - \frac{\mu C \eta}{2}\right)^i \\
&\leq \left(1 - \frac{\mu C \eta}{2}\right)^T (B_\eta(x_0) - B_\eta(x^*)) + \frac{\eta}{20}.
\end{aligned}
$$

We get accuracy η if

$$
\begin{aligned}
\left(1 - \frac{\mu C \eta}{2}\right)^T (B_\eta(x_0) - B_\eta(x^*)) &\leq \frac{\eta}{2} \\
\left(\frac{1}{1 - \frac{\mu C \eta}{2}}\right)^T &\leq \frac{2(B_\eta(x_0) - B_\eta(x^*))}{\eta} \\
T\mu C\eta/4 \leq T\left(-\ln\left(1 - \frac{\mu C \eta}{2}\right)\right) &\leq \ln \frac{2(B_\eta(x_0) - B_\eta(x^*))}{\eta},
\end{aligned}
$$

where the second last inequality is due to the fact that $\frac{z}{2} \leq \frac{z}{1+z} \leq \ln(1+z)$ for $z > -1$, taking $z = -\mu C\eta/2$. Thus, after at most $T = \frac{4}{\mu C \eta} \log \frac{2(B_\eta(x_0) - B_\eta(x^*))}{\eta}$ we get $B_\eta(x_T) - B_\eta(x^*) \leq \eta$.
Step 3. Case when the stopping criterion reached at iteration $t \leq T$. Alternatively, in case we reached $\|g_t\| \leq \frac{2\sqrt{\mu\eta}}{4.5}$ earlier, and due to $\|\Delta_t\| \leq \frac{\sqrt{\mu\eta}}{4.5}$ (given $\mathcal{A}_T$), using triangle inequality we get $\|\nabla B_\eta(x_t)\| \leq \|g_t\| + \|\Delta_t\| \leq \frac{2\sqrt{\mu\eta}}{4.5} + \frac{\sqrt{\mu\eta}}{4.5} = \frac{\sqrt{\mu\eta}}{1.5} \leq \sqrt{\frac{\mu\eta}{2}}$. Thus, using strong convexity:

$$
B_\eta(x_t) - B_\eta(x^*) \leq \frac{2\|\nabla B_\eta(x_t)\|^2}{\mu} \leq \frac{2\mu\eta}{2\mu} = \eta.
$$

Step 4. Sample complexity Recall from Lemma 1 (9):

$$
\mathbb{P}\left\{\|\Delta_t\| \leq b_0 + \hat{\sigma}_0(n)\sqrt{\ln\frac{1}{\delta}} + \sum_{i=1}^m \frac{\eta}{\alpha_t^i}\left(\hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}}\right) + \sum_{i=1}^m L_i \frac{\eta\sigma_i(n)}{\alpha_t^i[\bar{\alpha}_t^i]_+}\sqrt{\ln\frac{1}{\delta}}\right\} \geq 1 - \delta.
$$

Hence, we can guarantee $\|\Delta_t\| \leq \frac{\sqrt{\mu\eta}}{4.5} \forall t$, given that $\mathcal{A}_T$ holds and that for all $i \in [m]$:

$$
\hat{\sigma}_0(n) \leq \frac{\sqrt{\mu\eta}}{4(3m+2)\sqrt{\ln\frac{1}{\delta}}}, \hat{\sigma}_i(n) \leq \frac{\underline{\alpha}_t^i\sqrt{\mu}}{4(3m+2)\sqrt{\eta}\sqrt{\ln\frac{1}{\delta}}}, \tag{34}
$$

$$
\hat{b}_0 \leq \frac{\sqrt{\mu\eta}}{4(3m+2)}, \hat{b}_i \leq \frac{\underline{\alpha}_t^i\sqrt{\mu}}{4(3m+2)\sqrt{\eta}}, \sigma_i(n) \leq \frac{(\underline{\alpha}_t^i)^2\sqrt{\mu}}{4(3m+2)L\sqrt{\eta}\sqrt{\ln\frac{1}{\delta}}}. \tag{35}
$$

Noting that $\underline{\alpha}_t^i \geq c\eta$ if $\mathcal{A}_T$ holds, also using $\mathbb{P}\{\mathcal{A}_T\} \geq 1 - \hat{\delta}$, and $\hat{\sigma}_0(n) = \frac{\hat{\sigma}_0}{\sqrt{n}}, \hat{\sigma}_i(n) = \frac{\hat{\sigma}_i}{\sqrt{n}}, \hat{\sigma}_i(n) = \frac{\hat{\sigma}_i}{\sqrt{n}}$, we see that the required number of samples per iteration is $n = O(\frac{1}{\eta^3})$ with probability $1 - \hat{\delta}$. Hence, in total, we require $N = nT = \tilde{O}(\frac{1}{\eta^4})$. ■

4.5 Black-box optimization

A special case of stochastic optimization is zeroth-order optimization, in which one can access only the value measurements of f^i. In many applications, for example in physical systems with measurements collected by noisy sensors, we only have access to noisy evaluations of the functions.

4.5.1 Stochastic zeroth-order oracle

Formally we assume access to a *one-point stochastic zeroth-order oracle*, defined as follows. For any $i \in \{0, \dots, m\}$ this oracle provides noisy function evaluations at the requested point x_j: $F^i(x_j, \xi_j^i) = f^i(x_j) + \xi_j^i$, where ξ_j^i is a zero-mean σ_i-sub-Gaussian noise. We assume that noise values ξ_j^i may differ over iterations j and indices i even for the close points, i.e., we cannot access the evaluations of f^i with the same noise by two different queries: $\xi_j^i \neq \xi_{j+1}^i$ for any $F^i(x_j, \xi_j^i)$ and $F^i(x_{j+1}, \xi_{j+1}^i)$ even if $x_j = x_{j+1}$. Also, we assume that the noise vectors ξ_j of the measurements taken around the same point are i.i.d. random variables.

4.5.2 Zeroth-order gradient estimator

One way to tackle zeroth-order optimization is to sample a random point $x_t + \nu s_t$ around x_t at iteration t, and approximate the stochastic gradient $G^i(x, \xi)$ using finite differences. A classical choice for sampling is to use the Gaussian distribution, referred to as Gaussian sampling. However, since the Gaussian distribution has infinite support, one has an additional risk of sampling a point in the unsafe region arbitrarily far from the point, which is inappropriate for *safe learning.* Therefore, we propose to use the uniform distribution $\mathcal{U}(\mathbb{S}^d)$ on the unit sphere for sampling. In particular, in the case where we only have access to a noisy zeroth-order oracle, we estimate the gradient in the following way.

We need to estimate the descent directions of f^i using the zeroth-order information. For any point x, we can estimate the gradient of the function ∇f^i by sampling directions s_j uniformly at random on the unit sphere $s_j \sim \mathcal{U}(\mathbb{S}^d)$, and using the finite difference as follows:

$$G^i_{\nu,n}(x, \xi) := \frac{d}{n} \sum_{j=1}^{n} \frac{F^i(x + \nu s_j, \xi_j^{i+}) - F^i(x, \xi_j^{i-})}{\nu} s_j, \tag{36}$$

where $\xi_j^{i\pm}$ are sampled from σ_i-sub-Gaussian distribution. Note that s_j also satisfy the sub-Gaussian condition. [5]

There are also several other ways to sample directions to estimate the gradient from finite-differences. Berahas et al. (2021) compared various zeroth-order gradient approximation methods and showed that their sample complexity has a similar dependence on the dimensionality d required for a precise gradient approximation. Deterministic coordinate

5. There is also an option of using the one-point estimator $G^i_{\nu,n}(x, \xi) := \frac{d}{n} \sum_{j=1}^{n} \frac{F^i(x+\nu s_j, \xi_j^{i+})}{\nu} s_j$, but the variance of this estimator might be much higher. Note that even with zero-noise its variance grows to infinity while $\nu \to 0$. Its variance would depend on $\frac{\max_{x \in D} |f^i(x)|}{\nu}$, while the two-point estimator's variance depends on the Lipschitz constant L_i, which might be significantly smaller. Also, in the case of differentiable f^i with small noise ξ the two-point estimator becomes a finite difference directional derivative estimator with the accuracy dependent on ν only, in contrast to the one-point estimator.

sampling requires fewer samples due to smaller constants. However, we stick with sampling on the sphere because deterministic coordinate sampling requires the number of samples to be divisible by d. We want to keep flexibility on how many samples we can take per iteration; this number might be provided by the application. However, we note that any other sampling procedure can also be used.

Then, the estimator $G^i_{\nu,n}(x,\xi)$ defined above is a biased estimator of the gradient $\nabla f^i(x)$ and an unbiased estimator of the smoothed function gradient $\nabla f^i_\nu(x)$. The smoothed approximation f^i_ν of each function f^i is defined as follows:

Definition 12 *The ν-smoothed approximation of the function $f(x)$ is defined by $f_\nu(x) := \mathbb{E}_b f(x+\nu b)$, where b is uniformly distributed in the unit ball $\mathbb{B}^d$, and $\nu \geq 0$ is the sampling radius.*

Algorithm 3 Zeroth-order gradient-value estimator $(F^i_n(x,\xi), G^i_{\nu,n}(x,\xi))$

1: *Input:* $F^i(\cdot,\xi), i \in \{0,\dots,m\}$, $x \in \mathcal{X}$, $\nu > 0$, $n \in \mathbb{N}$;
2: Sample n directions $s_j \sim \mathcal{U}(\mathbb{S}^d)$, sample $F^i(x+\nu s_j, \xi^{i+}_j)$ and $F^i(x,\xi^{i-}_j)$, $j \in [n]$;
3: *Output:*

$$F^i_n(x,\xi) := \frac{\sum_{j=1}^n F^i(x,\xi^{i-}_j)}{n}$$
$$G^i_{\nu,n}(x,\xi) := \frac{d}{n}\sum_{j=1}^n \frac{F^i(x+\nu s_j,\xi^{i+}_j) - F^i(x,\xi^{i-}_j)}{\nu} s_j$$

Lemma 13 *Let $f^i_\nu(x)$ be the ν-smoothed approximation of $f^i(x)$. Then $\mathbb{E}G^i_{\nu,n}(x,\xi) = \nabla f^i_\nu(x)$, where the expectation is taken over both s_j and $\xi^{i\pm}_j$ for all $j \in [n]$.*

Proof First note that $\mathbb{E}G^i_{\nu,n}(x,\xi) = \underbrace{\mathbb{E}\frac{d}{n}\sum_{j=1}^n \frac{f^i(x+\nu s_j) - f^i(x)}{\nu} s_j}_{(1)} + \underbrace{\mathbb{E}\frac{d}{n}\sum_{j=1}^n \frac{\xi^{i+}_j - \xi^{i-}_j}{\nu} s_j}_{(2)}.$

Recall that $\xi^{i\pm}_j$ are independent on s_j and zero-mean, hence $(2) = 0$. The proof that $(1) = \nabla f^i_\nu(x)$ is classical (Flaxman et al., 2005) and is based on Stokes' theorem. ■

The following lemma shows important properties of the above zeroth-order gradient-value estimators.

Lemma 14 *Let $F^i(x,\xi)$ have variance $\sigma_i > 0$ and let the estimator $G^i_{\nu,n}(x,\xi)$ be defined as in (36) by sampling s_j uniformly from the unit sphere $\mathcal{U}(\mathbb{S}^d)$, then $f^i_\nu(x)$, and $G^i_{\nu,n}(x,\xi)$ are biased approximations of $f^i(x)$ and $\nabla f^i(x)$ respectively, such that*

$$|f^i(x) - f^i_\nu(x)| \leq \nu^2 M_i,$$

the variance of $F^i_n(x,\xi)$ is $\sigma_i(n) = \frac{\sigma_i}{\sqrt{n}}$ and the bias of $G^i_{\nu,n}(x,\xi)$ is bounded by:

$$\hat{b}_i := \|\nabla f^i(x) - \nabla f^i_\nu(x)\| \leq \nu M_i, \;\; \forall i \in \{0,\dots,m\}. \tag{37}$$

The variance of $G^i_{\nu,n}(x,\xi)$ is bounded as follows:

$$\hat{\sigma}_i^2(n) := \mathbb{E}\|G^i_{\nu,n}(x,\xi) - \nabla f^i_\nu(x)\|^2 \leq \frac{3}{n}\left(d\|\nabla f^i(x)\|^2 + \frac{d^2 M_i^2 \nu^2}{4}\right) + 4\frac{d^2}{n}\frac{\sigma_i^2}{\nu^2} \;\forall i \in \{0,\dots,m\}. \tag{38}$$

Proof These properties are corollaries from Berahas et al. (2021). For the bias (37) we use the result of Equation (2.35) (Berahas et al., 2021), and for the variance (38) the result of Lemma 2.10 of the same paper, in both cases by setting the disturbance $\epsilon_f = 0$ in Berahas et al. (2021). The last term of the variance is coming from the additive noise. We set the disturbance ϵ_f to zero for their formulation and analyze the noise separately since they consider the disturbance without any assumptions on it. In contrast, we consider the zero-mean and sub-Gaussian noise which we can use explicitly. For further discussions and proof, see Appendix A.8. ∎

4.5.3 Setting the sample radius ν and bounding the sample complexity

The parameters of the estimator defined in Algorithm 3 that we can control are ν and n. We want to set them in such a way that the biases b_i and variances $\sigma_i, \hat{\sigma}_i$ satisfy requirements of Theorems 8, 10, 11. Based on them, we can bound the sample complexity of our approach in the zeroth-order setting.

According to Theorems 8, 10, for non-convex and convex problems, we require the bias to be bounded by $\hat{b}_i \leq \frac{\underline{\alpha}_t^i}{(3m+2)R}$, $\hat{b}_0 \leq \frac{\eta}{(3m+2)R}$. Therefore, since $\hat{b}_i \leq \nu M_i$, we need to set the sampling radius small enough $\nu \leq \min\left\{\frac{\underline{\alpha}_t^i}{(3m+2)M_i R}, \frac{\eta}{(3m+2)M_0 R}\right\}$. For the strongly-convex case, according to Theorem 11, we require the bias to be bounded by $\hat{b}_0 \leq \frac{\sqrt{\mu\eta}}{4(3m+2)}, \hat{b}_i \leq \frac{\underline{\alpha}_t^i\sqrt{\mu}}{4(3m+2)\sqrt{\eta}}$. Hence, in this case we need to set the sample radius to $\nu \leq \min\left\{\frac{\underline{\alpha}_t^i\sqrt{\mu}}{(3m+2)M_i\sqrt{\eta}}, \frac{\sqrt{\eta}}{(3m+2)M_0}\right\}$. Moreover, in order to guarantee *safety* of all the measurements within the sample radius ν around the current point $f^i(x_t + \nu s_t) \leq 0$ using the smoothness of each constraint

$$f^i(x_t + \nu s_t) \leq f^i(x_t) - \nu\langle\nabla f^i(x_t), s_t\rangle + \nu^2\frac{M_i}{2}\|s_t\|^2,$$

we require the sample radius to be $\nu \leq \frac{\underline{\alpha}_t^i}{2\|\nabla f^i(x_t)\| + \sqrt{\underline{\alpha}_t^i M_i}}$. This bound can be obtained using the same derivations as for the adaptive step size γ_t (Lemma 3). Hence, for non-convex and convex problems we set

$$\nu = \min\left\{\frac{\underline{\alpha}_t^i}{2\|\nabla f^i(x_t)\| + \sqrt{\underline{\alpha}_t^i M_i}}, \frac{\underline{\alpha}_t^i}{(3m+2)M_i R}, \frac{\eta}{(3m+2)M_0}\right\} = O(\eta) = \Omega(\eta).$$

For strongly-convex case, respectively

$$\nu = \min\left\{\frac{\underline{\alpha}_t^i}{2\|\nabla f^i(x_t)\| + \sqrt{\underline{\alpha}_t^i M_i}}, \frac{\underline{\alpha}_t^i\sqrt{\mu}}{(3m+2)M_i\sqrt{\eta}}, \frac{\sqrt{\eta}}{(3m+2)}M_0\right\} = O(\eta) = \Omega(\eta).$$

Thus, from Lemma 14 (38), the variance of the estimated gradient with $\nu = O(\varepsilon) = \Omega(\varepsilon)$ is

$$\hat{\sigma}_i^2(n) = \frac{1}{n} O\left(\max\left\{ \frac{d^2\sigma_i^2}{\varepsilon^2}, L_i^2, d^2 M_i^2 \varepsilon^2 \right\} \right). \tag{39}$$

Additionally, according to the previous Theorems 8, 10, we require the variances to be $\hat{\sigma}_i(n) = O(\varepsilon)$ and $\sigma_i(n) = O(\varepsilon^2)$. From the above Eq. (39), in order to have $\hat{\sigma}_i(n) = O(\varepsilon)$ we require $n = O\left(\max\{\frac{d^2\sigma_i^2}{\varepsilon^4}, \frac{L_i^2}{\varepsilon^2}, dM^2\}\right)$. From the properties of the zero-mean noise, to have $\sigma_i(n) = O(\varepsilon^2)$ we require $n = O(\frac{\sigma_i^2}{\varepsilon^4})$. For the strongly-convex case, according to Theorem 11, we require $\sigma_i(n) = O(\varepsilon^{1.5})$ and $\hat{\sigma}_i(n) = O(\varepsilon)$, that is, we need $n = O(\frac{\sigma_i^2}{\varepsilon^3})$. Thus, we can prove the following corollary of the previously proven Theorems 8, 10, 11 particularly for the zeroth-order information case:

Corollary 15 *We get the following sample complexities for the zeroth-order information case, using* $\nu = \min\left\{ \frac{\alpha_t^i}{2\|\nabla f^i(x_t)\| + \sqrt{\alpha_t^i M_i}}, \frac{\alpha_t^i}{(3m+2)M_i R}, \frac{\eta}{(3m+2)M_0} \right\}$ *:*

- *For the non-convex problem, LB-SGD returns x_t such that is ε-approximate KKT point after at most $N = O(\frac{d^2\sigma_i^2}{\varepsilon^7})$ measurements with probability $1 - \hat{\delta}$.*
- *For the convex problem, LB-SGD returns x_t such that $\mathbb{P}\{f^0(x_t) - \min_{x\in\mathcal{X}} f^0(x) \leq \varepsilon\} \geq 1 - \hat{\delta}$ after at most $N = \tilde{O}(\frac{d^2\sigma_i^2}{\varepsilon^6})$ measurements.*
- *For the strongly-convex problem, with $\nu = \min\left\{ \frac{\alpha_t^i}{2\|\nabla f^i(x_t)\| + \sqrt{\alpha_t^i M_i}}, \frac{\alpha_t^i\sqrt{\mu}}{(3m+2)M_i\sqrt{\eta}}, \frac{\sqrt{\eta}}{(3m+2)M_0} \right\}$, LB-SGD returns x_t such that $\mathbb{P}\{f^0(x_t) - \min_{x\in\mathcal{X}} f^0(x) \leq \varepsilon\} \geq 1 - \hat{\delta}$ after at most $N = \tilde{O}(\frac{d^2\sigma_i^2}{\varepsilon^4})$ measurements.*
- *Moreover, all the query points of LB-SGD are feasible for (P) with probability at least $1 - \hat{\delta}$.*

5. Experiments

In this section, we demonstrate the empirical performance of our method when optimizing synthetic functions, as well as on a complex case study in constrained reinforcement learning.

Numerical stability. First, we note that to improve numerical stability, we slightly modify the steps of our method for practical applications. Recall that the log barrier gradient estimator is $g_t \leftarrow G_n^0(x_t, \xi_t) + \eta \sum_{i=1}^m \frac{G_n^i(x_t,\xi_t)}{-F_n^i(x_t,\xi_t)}$. Due to noise, the value of $-F_n^i(x_t, \xi_t)$ might become infinitely close to zero or negative, which leads to g_t blowing up or being unreliable. Therefore, we denote by $\bar{\alpha}_t^i$ the truncated value measurements $-F_n^i(x_t, \xi_t)$ with small truncation parameter $a > 0$, that is $\bar{\alpha}_t^i = [-F_n^i(x_t, \xi_t)]_a := \begin{cases} -F_n^i(x_t, \xi_t) & , -F_n^i(x_t, \xi_t) > a \\ a & , -F_n^i(x_t, \xi_t) \leq a \end{cases}$. Based on the above, we use the following estimator for the first-order stochastic optimization at point x_t: $g_t = G_n^0(x_t, \xi_t) + \eta \sum_{i=1}^m \frac{G_n^i(x_t,\xi_t)}{\bar{\alpha}_t^i}$.

Iteratively decreasing parameter η. Additionally, for all the experiments we iteratively decrease parameter η in a classical way as follows, since in practice it shows better performance:

Algorithm 4 LB-SGD with decreasing η_k ($\eta \in \mathbb{R}_+, \eta_0 \in \mathbb{R}_+, x_0 \in \mathbb{R}^d, \omega \in (0,1), \{T_k\}, \{n_k\}$)

1: *Input:* $M_i > 0, i \in \{0, \dots, m\}, \eta_0 \in \mathbb{R}_+$, $\hat{x}_0 \leftarrow x_0$, $K \leftarrow [\log_{\omega^{-1}} \frac{\eta_0}{\eta}]$;
2: **for** $k = 1, \dots, K$ **do**
3: $\quad \hat{x}_k \leftarrow$ LB-SGD$(\hat{x}_{k-1}, \eta_{k-1}, T_{k-1}, n_{k-1})$;
4: $\quad \eta_k \leftarrow \omega\eta_{k-1}$, with $\omega \in (0,1)$;
5: **end for**
6: *Output:* $\hat{x}_K$.

In the above, η_0 is the initial log barrier parameter, $\omega \in (0,1)$ is the parameter reduction rate, K is the number of rounds, $\eta = \eta_K$ is a barrier parameter at the last round, at every round k we run LB-SGD with T_k iterations and n_k oracle calls per iteration.

5.1 Safe black-box learning

In this section, we demonstrate the performance of our method on simulations and compare it to other existing non-linear safe learning approaches. All the experiments in this subsection were carried out on a Mac Book Pro 13 with 2.3 GHz Quad-Core Intel Core i5 CPU and with 8 GB RAM. The code corresponding to the experiments in this subsection can be found under the following link: `https://github.com/Ilnura/LB_SGD`.

5.1.1 Convex objective and constraints

We first compare our safe method LB-SGD with SafeOpt (Sui et al., 2015b; Berkenkamp et al., 2016a) and LineBO (Kirschner et al., 2019), on a simple synthetic example.

We consider the quadratic problem with linear constraints $\min_{x \in \mathbb{R}^d} \|x - x_0\|^2/4d$, s.t. $Ax \leq b$, where $x_0 = [2, \dots, 2]$ and $A = \begin{bmatrix} I_d \\ -I_d \end{bmatrix}$, $b = \mathbf{1}/\sqrt{d}$. The optimum of this problem is on the boundary. We assume that the linearity of the constraints is unknown, hence for SafeOpt we use the Gaussian kernel. For dimensions $d = 2, 3, 4$ we carry out the simulations with standard deviation $\sigma = 0.001$ of an additive noise Figure 2 averaged over 10 different experiments. For $d = 2$ we run SafeOpt, and for $d = 3, 4$ we run SafeOptSwarm, which is a heuristic making SafeOpt updates more tractable for slightly higher dimensions (Berkenkamp et al., 2016a). For SafeOpt and LineBO methods, instead of plotting the accuracy and constraints corresponding to x_t, we plot the smallest accuracy and biggest constraint seen up to the step t (for sake of interpretability of the plots). Even for $d = 4$, LB-SGD is already notably more sample efficient compared to both SafeOpt and LineBO. Moreover, LB-SGD significantly outperforms SafeOpt over computational cost and memory usage. It is well known that SafeOpt's sample complexity and computational cost can exponentially depend on the dimensionality. In contrast, the complexity of LB-SGD depends on d polynomially. The runtimes of the above experiments, in seconds, are shown in Table 2 and Figure 4.

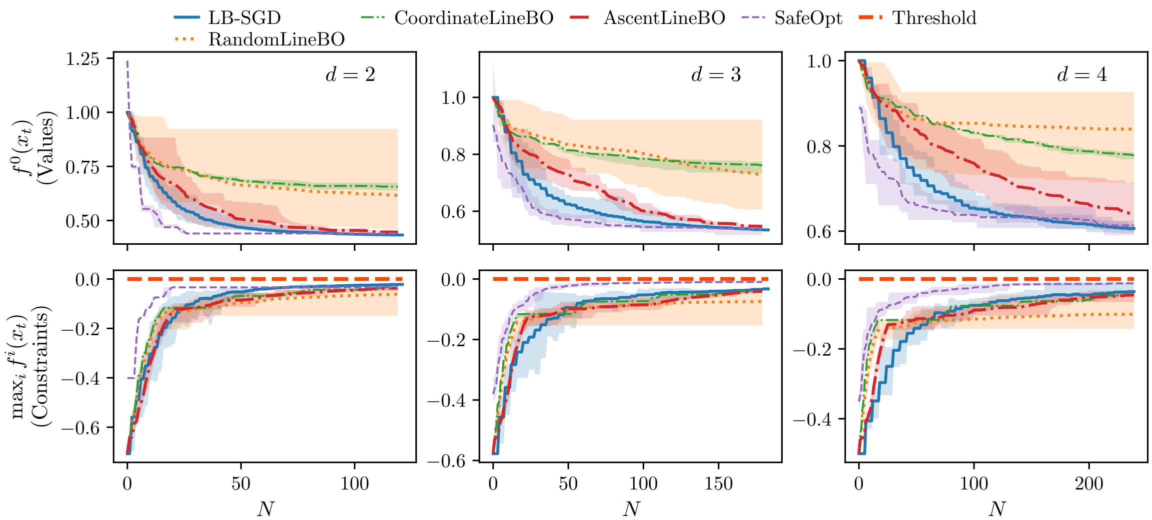

Figure 2: Accuracy (upper plots) and constraints (lower plots) of LB-SGD and SafeOpt for $d = 2, 3, 4$, averaged over 10 samples. t here is the amount of zeroth-order oracle calls. In these experiments, for LB-SGD we decrease $\eta_{k+1} = 0.7\eta_k$ gradually every $T_k = 7$ steps with $n_k = [\frac{d}{2}]$ value measurements at each step. Already for $d = 4$ LB-SGD starts outperforming all BO-based methods on this problem in terms of the sample complexity. Best viewed in color.

d	2	3	4
SafeOpt (SafeOptSwarm)	4.289	114.406	212.514
LineBO	8.180	17.837	40.8
LB-SGD	0.429	0.895	0.781

Table 2: Average runtime dependence on dimensionality d (in seconds). Importantly, the wall-clock time of LB-SGD remains roughly constant as we increase d, noting that we only increase the number of measurements per iteration, but not the number of iterations.

5.1.2 Non-convex objective and constraints

As a non-convex example, we consider the Rosenbrock function, a common benchmark for black-box optimization, with quadratic constraints. In particular, we consider the following problem

$$\min_{x \in \mathbb{R}^d} \sum_{i=1}^{d-1} 100\|x_i - x_{i+1}\|^2 - \|1 - x_i\|^2,$$
$$\text{s.t. } \|x\|^2 \leq r_1^2,\ \|x - \hat{x}\|^2 \leq r_2^2.$$

We set $r_1 = 0.1$, $r_2 = 0.2$, $\hat{x} = [-0.05, \dots, -0.05]$. The optimum of this problem is on the boundary of the constraint set. We show the comparison of LB-SGD and SafeOpt on Figure 3. Again, for $d = 2$ we run SafeOpt, and for $d = 3, 4$ we run SafeOptSwarm. Here, on the constraints plot of SafeOpt and LineBO we again plot the highest value of the constraints over all points explored so far.

The run-times of LineBO and SafeOpt are demonstrated in Table 3 and Figure 4. Note

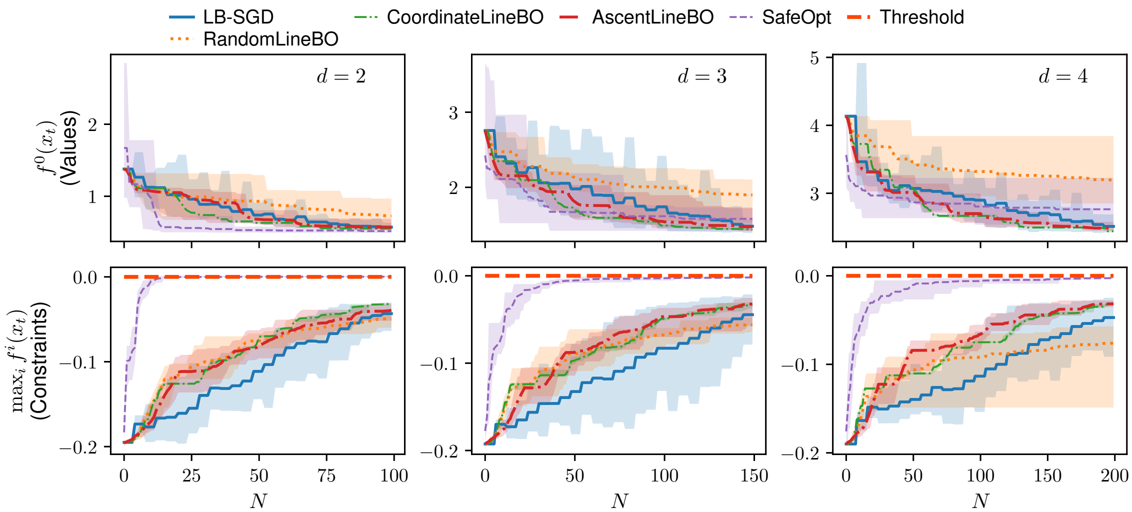

Figure 3: Accuracy and constraints of LB-SGD and SafeOpt for $d = 2, 3, 4$, averaged over 10 samples. t here is the amount of zeroth-order oracle calls. In these experiments, for LB-SGD we decrease $\eta_{k+1} = 0.7\eta_k$ gradually every $T_k = 5$ steps with $n_k = d - 1$ value measurements at each step. On this problem, for $d = 4$ we observe that LB-SGD performs better that SafeOpt, and comparable to LineBO. Best viewed in color.

d	2	3	4
SafeOpt (SafeOptSwarm)	26.960	44.909	63.019
LineBO	7.584	10.593	13.293
LB-SGD	0.294	0.332	0.324

Table 3: Run-time (in seconds) as dependent on dimensionality d (Rosenbrock benchmark).

that the second problem is easier for BO methods than the first one. It is related to the fact that in the first problem, the number of constraints (and therefore, the number of GPs) is higher and grows with dimensionality ($m = 4, 6, 8$). In contrast, there are always only two constraints for the second problem. As one can see, our approach is significantly cheaper in computational time than SafeOpt. This is, of course, at the price of finding only a local minimum, not the global one.

5.1.3 Comparison with LineBO in higher dimensions

In higher dimensions, it is well known that SafeOpt is not tractable. Therefore, we compare our method only with LineBO (Kirschner et al., 2019). This method scales significantly better with dimensionality than the classical BO approaches. The method was demonstrated to be efficient in the unconstrained case and in cases where the solution lies in the interior of the constraint set. The authors proved the theoretical convergence in the unconstrained case and the safety of the iterations in the constrained case. However, in contrast to our method, this approach has a drawback that we discuss below. At each iteration, LineBO samples a direction (at random, an ascent direction of the objective, or a coordinate direction). Then

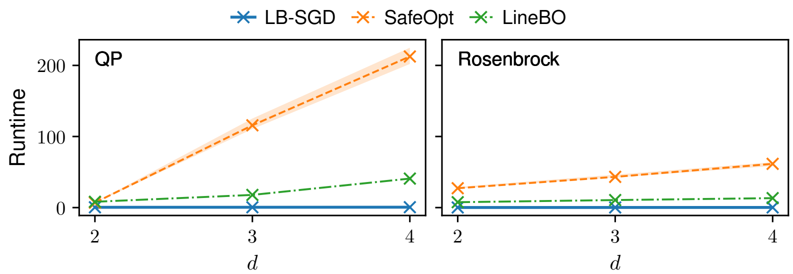

Figure 4: Run-times of LB-SGD and SafeOpt for $d = 2, 3, 4$, averaged over 10 samples, in seconds. t here is the amount of zeroth-order oracle calls. We can observe that LB-SGD is a significantly cheaper approach in terms of the computational cost compared to both BO-based methods with growing dimensions.

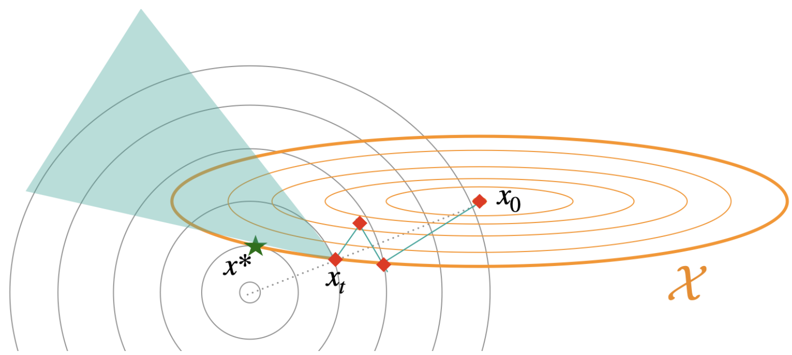

Figure 5: Illustration of the LineBO behavior. At point x_t not every direction allows the safe improvement (only the directions lying in the green sector). Therefore, the LineBO method might get stuck sampling the wrong directions. On this example, the closer to the solution, the narrower is the improvement sector.

it solves a *1-dimensional* constrained optimization along this direction, using SafeOpt. After optimizing along this direction, it samples another direction starting from the current point. The drawback of this approach is that when the solution is on the boundary, LineBO might get stuck on the wrong point on the boundary. In such a case, it might be difficult for it to find a safe direction of improvement too close to the boundary. See Figure 5 for the illustration of this potential problem. Furthermore, the higher dimension, the harder it is to sample a suitable direction. We demonstrate that empirically in application to the following problem:

$$\min_{x \in \mathbb{R}^d} -\exp^{-4\|x\|^2}, \tag{40}$$

$$\text{s.t. } \langle x - \hat{x}, A(x - \hat{x})\rangle \le r^2, \tag{41}$$

with $r = 0.5$ and $A = \text{diag}(3, 1.2, \ldots, 1.2)$. On Figure 6 we demonstrate the comparison of LineBO and LB-SGD methods on the above problem for dimensionalities $d = 2, 10, 20$. We report the run-times in Table 4.

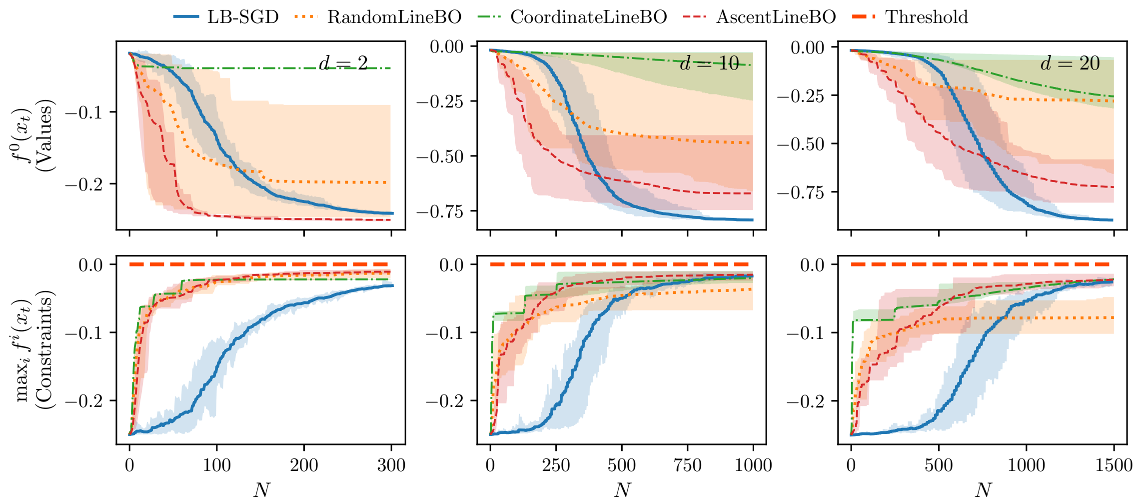

Figure 6: Accuracy and constraints of LB-SGD and SafeRandomLineBO for $d = 2, 10, 20$, averaged over 10 samples. The amount of zeroth-order oracle calls here is denoted by t. In these experiments, for LB-SGD we decrease $\eta_{k+1} = 0.85\eta_k$ gradually every $T_k = 3$ steps with $n_k = [\frac{d+1}{2}]$ value measurements at each step.

d	2	10	20
LB-SGD	0.828	2.186	2.676
LineBO	12.883	298.097	1038.459

Table 4: Runtime (in seconds) dependence on dimensionality d on the negative Gaussian minimization benchmark. We can observe that LineBO is significantly more expensive in computational cost (for the same number of queried points).

To compare, in the case when the solution is in the interior of the constraint set achieved by setting $r = 10$ (that is, if the constraints do not influence the solution), the LineBO approach does not have this issue and can still be very efficient (see Figure 7).

5.2 LB-SGD for safe reinforcement learning

We previously showed LB-SGD performance on smaller scale, classical black box benchmark problems. In this part, we showcase how LB-SGD scales to more complex, high-dimensional domains arising in RL. We consider the "Safety Gym" (Ray et al., 2019) benchmark suite designed to evaluate safe RL approaches on constrained Markov decision processes (CMDP). In Safety Gym, a robot needs to reach a goal area while avoiding obstacles on the way. Navigation is done by observing first-person-view of the robot.

5.2.1 Problem statement

Constrained Markov decision processes The problem of safe reinforcement learning can be viewed as finding a policy that solves a *constrained Markov decision process* (Altman, 1999). We define a discrete-time episodic CMDP as a tuple $(\mathcal{S}, \mathcal{A}, \rho, P, R, \gamma, \mathcal{C})$. At each time

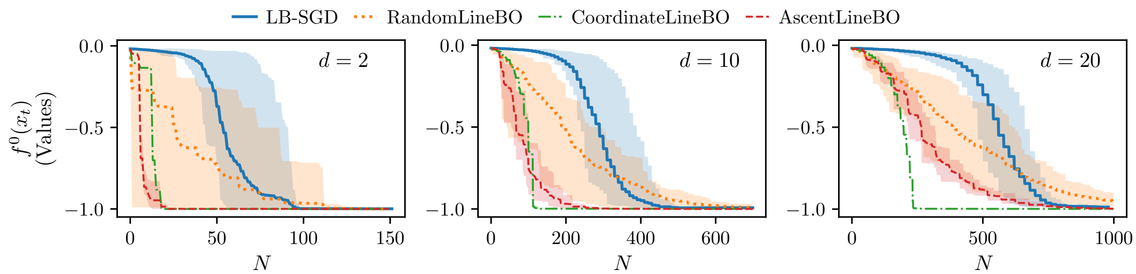

Figure 7: Accuracy of LB-SGD and SafeRandomLineBO for $d = 2, 10, 20$, averaged over 10 samples, when the solution is strictly in the interior. N is the amount of zeroth-order oracle calls. For LB-SGD we decrease $\eta_{k+1} = 0.85\eta_k$ gradually every $T_k = 3$ steps with $n_k = [\frac{d+1}{2}]$ value measurements at each step. Constraints were never violated by any of the methods.

step $\tau \in \{0, \dots, \mathrm{T}\}$, an agent observes a state $s_\tau \in \mathcal{S}$. The initial state s_0 is determined according to some unknown distribution $\rho(s_0)$ such that $s_0 \sim \rho(s_0)$. Given a state, the agent decides what action $a_\tau \in \mathcal{A}$ to take next. Then, an unknown transition density $P : \mathcal{S} \times \mathcal{A} \times \mathcal{S} \to [0, 1]$, $s_{\tau+1} \sim P(\cdot|s_\tau, a_\tau)$ generates a new state. $R : \mathcal{S} \times \mathcal{A} \to \mathbb{R}$ is a reward function that generates an immediate reward signal observed by the agent. The discount factor $\gamma \in (0, 1]$ weighs the importance of immediate rewards compared to future ones. Lastly, $\mathcal{C} = \left\{c^i : \mathcal{S} \times \mathcal{A} \to \mathbb{R} \;\middle|\; i \in [m]\right\}$ is a set of immediate cost signals that the agent observes alongside the reward. The goal is to find a policy $\pi : \mathcal{S} \times \mathcal{A} \to [0, 1]$ that solves the constrained problem:

$$\max_\pi \underbrace{\mathbb{E}\left[\sum_\tau \gamma^\tau R(s_\tau, a_\tau)\right]}_{-f^0(x)} \quad \text{s.t.} \quad \underbrace{\mathbb{E}\left[\sum_\tau \gamma^\tau c^i(s_\tau, a_\tau)\right] - d^i}_{f^i(x)} \leq 0 \; \forall i \in [m]. \tag{42}$$

In the above, $d^i, i \in [m]$ are predefined threshold values for the expected discounted return of costs. Note that we take the expectation with respect to all stochasticity induced by the CMDP and policy.

On-policy methods as black-box optimization problems A typical recipe for solving CMDPs at scale is to parameterize the policy with parameters x and use *on-policy* methods. On-policy methods use Monte-Carlo sampling to sample trajectories from the environment, evaluate the policy, and finally update it (Chow et al., 2015; Achiam et al., 2017; Ray et al., 2019). By using Monte-Carlo, these methods compute unbiased estimates of the constraints, objective and their gradients (e.g., via REINFORCE, cf. Sutton et al., 2000), equivalently to the assumptions in Section 2.1. In particular, the process of sampling trajectories from the CMDP and averaging them to estimate the objective and constraints in Equation (42) is equivalent to querying $f^0(x)$ and $f^i(x), i \in [m]$ and gives rise to a first-order, stochastic and unbiased oracle. However, without deliberately enforcing $x_t \in \mathcal{X}\ \forall t \in \{0, \dots, T\}$, these methods may use an unsafe policy during learning[6].

6. It is important to note the work of Dalal et al. (2018), which, under a more strict setting, treats this specific challenge, but uses an off-policy algorithm (Lillicrap et al., 2015, DDPG) to solve CMDPs.

Solving CMDPs with LB-SGD Another shortcoming of the previously mentioned algorithms is that using only Monte-Carlo sampling often leads to high variance estimates of $f^0(x)$ and $f^i(x)$ (Schulman et al., 2015). To reduce this variance, one must typically take an abundant number of queries of $f^0(x)$ and $f^i(x)$, making the aforementioned algorithms sample-inefficient. One way to improve sample efficiency, is to learn a model of the CMDP, and query it (instead of the real CMDP) to have *approximations* of $f^0(x)$ and $f^i(x)$. This allows us to trade off the high variance and sample inefficiency with some bias introduced my model errors. By making this compromise, model-based methods empirically exhibit improved sample efficiency compared to the previously mentioned on-policy methods (Deisenroth and Rasmussen, 2011; Chua et al., 2018; Hafner et al., 2021). Motivated by this insight, we use LAMBDA (As et al., 2022), a recent model-based approach for solving CMDPs. In short, LAMBDA learns the transition density P from image observations, and uses this learned model to try and find an optimal policy. To accommodate the complexity of learning a policy from a high dimensional input such as images, LAMBDA requires $\sim$588400 parameters to parameterize the policy. This allows us to demonstrate LB-SGD's ability to scale to problems with large dimensionality. To solve Equation (42), LAMBDA queries *approximations* of $f^0(x), f^i(x)$ by using its model of the CMDP together with its policy to sample *model-generated, on-policy trajectories*. As described before, these model-generated trajectories are subject to model errors which in turn makes the estimation of the objective and constraints biased. As a result, the assumptions in Section 2.1 do not necessarily, hold as in this case, the oracle is first-order, stochastic but *biased*. Nevertheless, this biasedness is subject only to LAMBDA's model inaccuracies, so LB-SGD can still produce safe policies with high utility, as we empirically show in the following section. As et al. (2022) use the "Augmented Lagrangian" (Nocedal and Wright, 2006) to solve the constrained problem with SGD. However, by using the Augmented Lagrangian, x_t are generally infeasible throughout training, even if $x_0 \in \mathcal{X}$. Therefore, we use LB-SGD instead and empirically show that by using it we get $x_t \in \mathcal{X}\ \forall t \in \{0, \dots, T\}$.

5.2.2 Experiments

Addressing the assumptions Let us briefly discuss the assumptions in Section 2 and explicitly state which of them do not hold. **Oracle.** As mentioned before, we cannot guarantee the assumptions in Section 2.1. LAMBDA uses neural networks to model the transition density and to learn an approximation of the objective and constraints[7]. For this reason, the assumption on unbiased zeroth-order queries, and assumptions of sub-Gaussian oracles do not hold. **Smoothness.** By choosing ELU activation function (Clevert et al., 2015) we ensure the smoothness of our approximation of the objective and constraints. **MFCQ.** In general, similarly to the assumptions on the oracle, this cannot be guaranteed. However, in our experiments, the CMDP is defined to have only one constraint ($m = 1$) so this assumption is satisfied by definition. **Safe initial policy.** This assumption exists in a large body of previous work (Berkenkamp et al., 2017; Koller et al., 2018; Wabersich and Zeilinger, 2021). Yet, it is not always clear how to design such a policy a-priori. In the following, we propose an experimental protocol in which this assumption empirically holds.

7. The approximation of the objective and constraint is done by learning their corresponding *value functions*. Please see As et al. (2022) for further details.

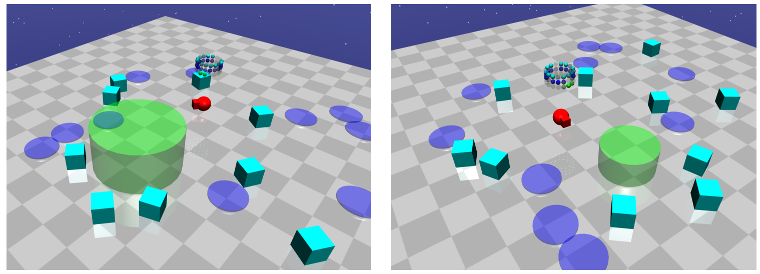

Figure 8: In Safety Gym, a robot should navigate to a goal area (green circle), while avoiding all other objects (turquoise and blue). The robot on the right picture should arrive to a smaller goal region, making navigation harder.

Experiment protocol To ensure LB-SGD starts from a safe policy, we warm-start it with a policy that was trained on a similar, but easier task. Specifically, we follow a similar experimental setup as As et al. (2022) but first train the agent with LAMBDA on a task in which the *goal area is larger*, as shown in Section 5.2.2. We use the policy parameters of the trained agent as a starting point for LB-SGD on a harder task, in which the goal area is smaller. As we later show, this allows the agent to start the second stage with a *safe but sub-optimal policy*. We verify this setup with all three available robots of the Safety-Gym benchmark suite (Ray et al., 2019), each run with 5 different random seeds. For our implementation, please see `https://github.com/lasgroup/lbsgd-rl`.

Results We first validate our experimental setup. In Figure 9 we show that by using either LB-SGD or the Augmented Lagrangian in the first stage, and *not updating* the policy in the second stage, LAMBDA's policy is *safe* but *sub-optimal*. With this, we empirically confirm the safe initial policy assumption on the second stage of training. Further, given such a policy, we compare LB-SGD with the Augmented Lagrangian on the second stage. In Figure 10 we demonstrate how the Augmented Lagrangian needs to "re-learn" a new value for the Lagrange multiplier and therefore fails to transfer safely to the harder task. However, by using LB-SGD, the agent is able to maintain safety after transitioning to the harder task. It is important to note that this safe transfer comes at the cost of limited exploration. As shown in Figures 9 and 10, LB-SGD finds slightly less performant policies compared to the Augmented Lagrangian. In Figure 10 we also compare with the classical constrained policy optimization (CPO) algorithm (Achiam et al., 2017) following the implementation in `https://github.com/lasgroup/jax-cpo` as a baseline. CPO is a model-free method and hence requires much more environment interactions. Therefore, we only show the final level CPO reaches after 10 million environment steps.

6. Conclusion

In this paper, we addressed the problem of sample and computationally efficient safe learning. We proposed an approach based on logarithmic barriers, which we optimize using SGD with adaptive step sizes. We analytically proved its safety during the learning and analyzed the convergence rates for non-convex, convex, and strongly-convex problems. We empirically demonstrated the performance of our method in comparison with other existing methods. We show that 1) its sample and computational complexity scale efficiently to high dimensions, and; 2) it keeps optimization iterates within the feasible set with high probability. Additionally,

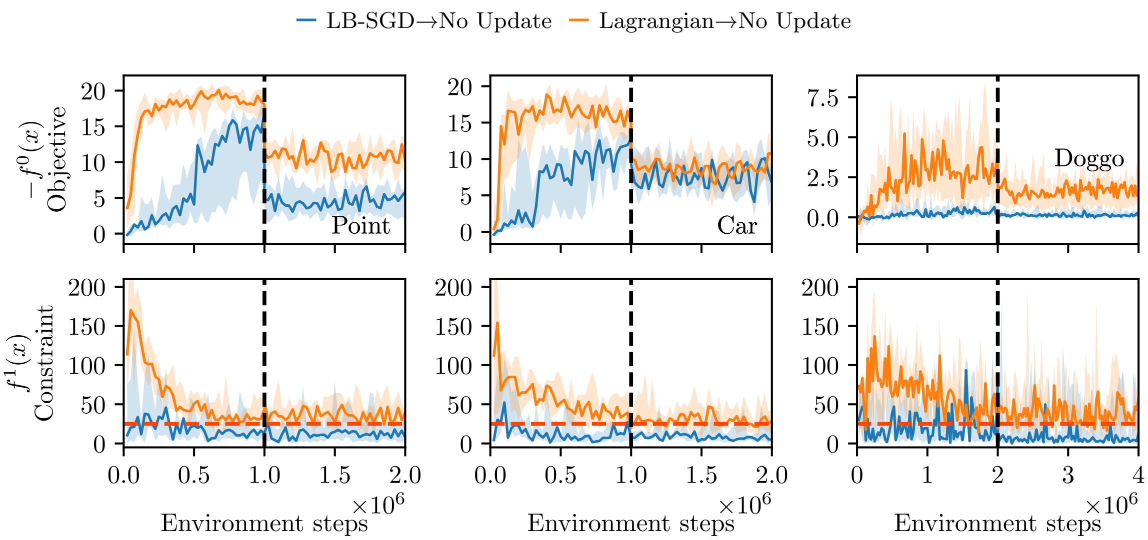

Figure 9: Across all different robots of the Safety Gym suite, LAMBDA with LB-SGD and the Augmented Lagrangian transfer well to the second stage in terms of safety. Since we do not update the policy, LAMBDA fails to reach the same task performance on the second stage, as expected. Shaded areas represent the the 5% and 95% percentiles of 5 different random seeds.

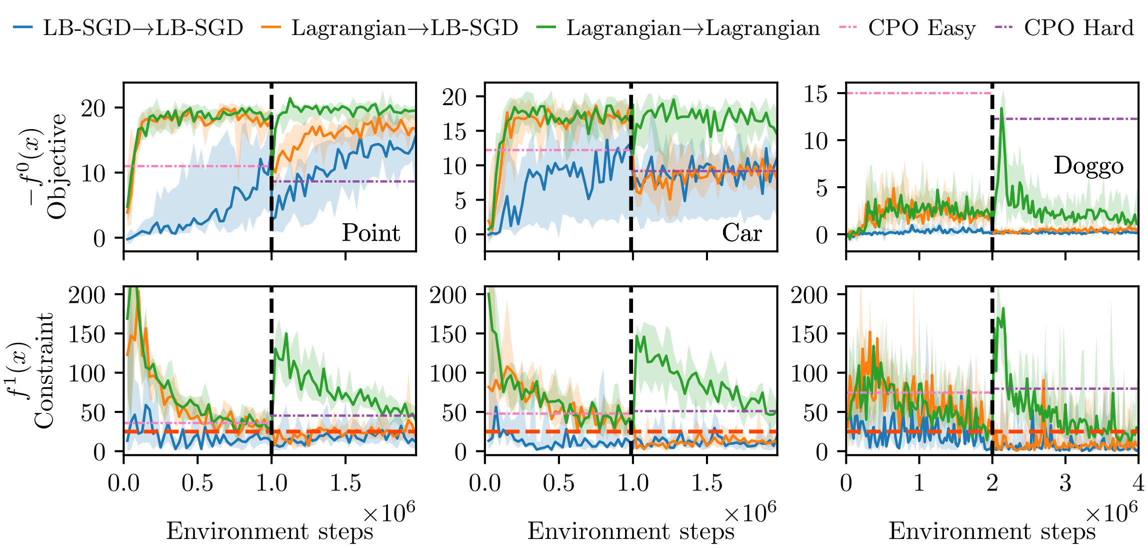

Figure 10: LAMBDA with LB-SGD transfers safely to the second task. The main trade-off, however, is the lower asymptotic performance of LB-SGD. Conversely, Lagrangian → Lagrangian fails transfer safely as the constraints with all robots rapidly grow when the new task is revealed (as shown by the vertical dashed black line). CPO Easy and CPO Hard correspond to CPO's performance on the easier task (larger goal size) and the harder task (smaller goal size). We train CPO on each task separately and show its performance at convergence, i.e., after 10×10^6 interaction steps for the "Point" and "Car" robots and 100×10^6 for the "Doggo" robot. As shown, while CPO achieves reasonable performance, it fails to satisfy the constraints in all tasks and all robots, as previously reported by Ray et al. (2019).

we demonstrate the efficiency of the log barrier approach for high-dimensional constrained reinforcement learning problems.

While not requiring to explicitly specify a prior (in the Bayesian sense, as considered in safe Bayesian optimization), our method does involve hyper-parameters such as η_0, η-decrease rate parameter ω, amount of steps per episode T_k, and exhibits sensitivity to the noise. Also, in the non-convex case, it can converge only to a local minimum, as any other descent optimization approach. However, it is easy to implement and has efficient computational performance due to cheap updates. Therefore, LB-SGD is better suited to problems of high scale.

For future work, it would be exciting to take the best of both worlds and combine the BO approaches that allow us to build and use a global model with our simple and cheap safe descent approach based on log barriers.

Acknowledgments

We thank Tingting Ni for very helpful discussions, and for helping to improve the sample complexity bound for the strongly-convex case. With modified analysis that uses similar ideas to Ni and Kamgarpour (2023), we could improve our previous bound from $\tilde{O}(d^2/\varepsilon^5)$ to $\tilde{O}(d^2/\varepsilon^4)$. We also thank the support of Swiss National Science Foundation, under the grant SNSF 200021_172781, the Swiss National Science Foundation under NCCR Automation, grant agreement 51NF40 180545, and European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme grant agreement No 815943.

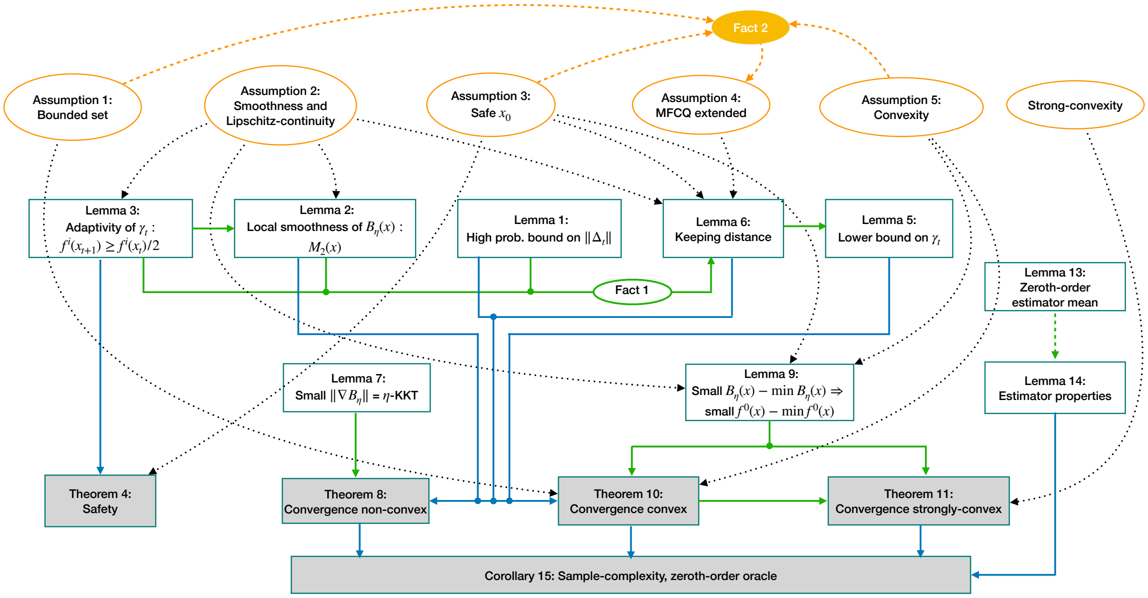

Figure 11: Relations of our theoretical results.

Appendix A. Additional proofs

First, on Figure 11 we show how various theorems and lemmas relate to each other throughout our paper.

A.1 MFCQ

Let x^* be a local minimizer of constrained problem (P), and let $\mathcal{I}(x^*) := \{i \in [m] : f^i(x^*) = 0\}$ denote the set of active constraints at x^*. Then, the classic MFCQ is satisfied at x^* if there exists $s \in \mathbb{R}^d$ such that $\langle s, \nabla f^i(x^*)\rangle < 0$ for all $i \in \mathcal{I}(x^*)$.

On Figure 12, for the point in the middle, both constraints are ρ-almost active. Note that at this point, no descent direction exists for both constraints since their gradients are pointing to the opposite directions. That is, this set does not satisfy the extended MFCQ with the given $\rho > 0$.

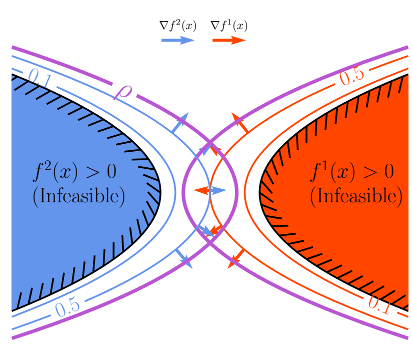

Figure 12: Illustration of extended MFCQ.

A.2 Proof of Lemma 1

Proof Assume x_t is strictly feasible, that is $\alpha_t^i > 0$. Using the triangle inequality, we get

$$
\begin{aligned}
\|\Delta_t\| &= \|g_t - \nabla B_\eta(x_t)\| \\
&= \left\| G_n^0(x_t) - \nabla f^0(x_t) + \sum_{i=1}^m \left[\eta G_n^i(x_t) \left(\frac{1}{\bar{\alpha}_t^i} - \frac{1}{\alpha_t^i} \right) + \eta (G_n^i(x_t) - \nabla f^i(x_t)) \frac{1}{\alpha_t^i} \right] \right\| \\
&\leq \left\| G_n^0(x_t) - \nabla f^0(x_t) \right\| + \sum_{i=1}^m \left[\eta \left\| G_n^i(x_t) \right\| \underbrace{\left| \frac{1}{\bar{\alpha}_t^i} - \frac{1}{\alpha_t^i} \right|}_{(1)} + \frac{\eta}{\alpha_t^i} \left\| G_n^i(x_t) - \nabla f^i(x_t)) \right\| \right].
\end{aligned}
$$

In the above, (1) we can bound as follows:

$$
\left| \frac{1}{\bar{\alpha}_t^i} - \frac{1}{\alpha_t^i} \right| = \left| \frac{\alpha_t^i - \bar{\alpha}_t^i}{\bar{\alpha}_t^i \alpha_t^i} \right| = \frac{|\alpha_t^i - \bar{\alpha}_t^i|}{|\bar{\alpha}_t^i| \alpha_t^i} \leq \frac{|\alpha_t^i - \bar{\alpha}_t^i|}{[\bar{\alpha}_t^i]_+ \alpha_t^i}
$$

From the sub-Gaussian property, for every $i \in [m]$ we have:

$$
\mathbb{P}\left\{ \left\| G_n^i(x_t) - \nabla f^i(x_t) \right\| \leq \hat{b}_i + \hat{\sigma}_i(n) \sqrt{\ln \frac{1}{\delta}} \right\} \geq 1 - \delta
$$

$$
\mathbb{P}\left\{ \left| \alpha_t^i - \bar{\alpha}_t^i \right| \leq \sigma_i(n) \sqrt{\ln \frac{1}{\delta}} \right\} \geq 1 - \delta,
$$

With high probability, we know $\|G_n^i(x_t)\| \leq L_i + \hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln \frac{1}{\delta}} =: \hat{L}_i$ if $\hat{b}_i + \hat{\sigma}_i(n)\sqrt{\ln \frac{1}{\delta}} \leq L_i$, and from what we conclude:

$$
\mathbb{P}\left\{ \|\Delta_t\| \leq \hat{b}_0 + \hat{\sigma}_0(n)\sqrt{\ln \frac{1}{\delta}} + \sum_{i=1}^m \frac{\eta}{\alpha_t^i} \left(\hat{b}_i + \hat{\sigma}_i(n) \sqrt{\ln \frac{1}{\delta}} \right) + \sum_{i=1}^m \hat{L}_i \frac{\eta}{\alpha_t^i [\bar{\alpha}_t^i]_+} \sigma_i(n) \sqrt{\ln \frac{1}{\delta}} \right\} \geq 1 - 2m\delta.
$$

■

A.2.1 Bias

Proof Using $\mathbb{E}[XY] \leq \sqrt{\mathbb{E}[X^2]\mathbb{E}[Y^2]}$, we get

$$\begin{aligned}
\|\mathbb{E}\Delta_t\| &= \|\mathbb{E}[g_t - \nabla B_\eta(x_t)]\| \\
&\leq \left\| \mathbb{E}[G^0(x_t) - \nabla f^0(x_t)] + \sum_{i=1}^m \mathbb{E}\left[\eta G^i(x_t)\left(\frac{1}{\bar{\alpha}_t} - \frac{1}{\alpha_t}\right) + \eta(G^i(x_t) - \nabla f^i(x_t))\frac{1}{\alpha_t}\right]\right\| \\
&= \left\| \sum_{i=1}^m \mathbb{E}\left[\eta G^i(x_t)\left(\frac{1}{\bar{\alpha}_t} - \frac{1}{\alpha_t}\right)\right]\right\| + \hat{b}_t^0 + \sum_{i=1}^m \frac{\eta}{\alpha_t^i}\hat{b}_t^i \\
&\leq \sum_{i=1}^m \mathbb{E}\left[\left\|\eta G^i(x_t)\left(\frac{1}{\bar{\alpha}_t^i} - \frac{1}{\alpha_t^i}\right)\right\|\right] + \hat{b}_t^0 + \sum_{i=1}^m \frac{\eta}{\alpha_t^i}\hat{b}_t^i \\
&\leq \sum_{i=1}^m \sqrt{\mathbb{E}[\eta^2\|G^i(x_t)\|^2]\mathbb{E}\left[\frac{1}{(\bar{\alpha}_t^i)^4}\|\alpha_t^i - \bar{\alpha}_t^i\|^2\right]} + \hat{b}_t^0 + \sum_{i=1}^m \frac{\eta}{\alpha_t^i}\hat{b}_t^i \\
&\leq \sum_{i=1}^m \frac{\eta L_i \sigma_t}{(\alpha_t^i)^2} + \hat{b}_t^0 + \sum_{i=1}^m \frac{\eta}{\alpha_t^i}\hat{b}_t^i,
\end{aligned}$$

■

A.3 Proof of the adaptivity

Proof For adaptivity, we require

$$f^i(x_{t+1}) \leq \frac{f^i(x_t)}{2}.$$

Using M_i-smoothness, we can bound the i-th constraint growth:

$$\begin{aligned}
f^i(x_{t+1}) &\leq f^i(x_t) + \langle \nabla f^i(x_t), x_{t+1} - x_t\rangle + \frac{M_i}{2}\|x_t - x_{t+1}\|^2 \\
&= f^i(x_t) - \gamma_t\langle \nabla f^i(x_t), g_t\rangle + \gamma_t^2\frac{M_i}{2}\|g_t\|^2
\end{aligned}$$

That is, the condition on γ_t for adaptivity (and safety) we can formulate by

$$-\gamma_t\langle \nabla f^i(x_t), g_t\rangle + \gamma_t^2\frac{M_i}{2}\|g_t\|^2 \leq \frac{-f^i(x_t)}{2} = \frac{\alpha_t^i}{2}.$$

By carefully rewriting the above inequality without strengthening it, we get

$$\begin{aligned}
\gamma_t^2\frac{M_i}{2}\|g_t\|^2 - \gamma_t\langle \nabla f^i(x_t), g_t\rangle \pm \frac{1}{2M_i}\frac{\langle \nabla f^i(x_t), g_t\rangle^2}{\|g_t\|^2} &\leq \frac{\alpha_t}{2} \\
\left(\gamma_t\|g_t\| - \frac{\langle \nabla f^i(x_t), g_t\rangle}{M_i\|g_t\|}\right)^2 &\leq \frac{\alpha_t}{M_i} + \frac{\langle \nabla f^i(x_t), g_t\rangle^2}{M_i^2\|g_t\|^2}
\end{aligned}$$

Using the quadratic inequality solution, we obtain the following sufficient bound on the adaptive γ_t :

$$\gamma_t\|g_t\| \leq \frac{\langle \nabla f^i(x_t), g_t\rangle}{M_i\|g_t\|} + \sqrt{\frac{\alpha_t}{M_i} + \frac{\langle \nabla f^i(x_t), g_t\rangle^2}{M_i^2\|g_t\|^2}} = (*)$$

Then, we can rewrite this expression of the right part as follows:

$$\begin{aligned}
(*) &= \sqrt{\frac{\langle \nabla f^i(x_t), g_t\rangle^2}{M_i^2\|g_t\|^2} + \frac{\alpha_t^i}{M_i}} + \frac{\langle \nabla f^i(x_t), g_t\rangle}{M_i\|g_t\|} = \sqrt{\frac{\alpha_t^i}{M_i}}\left(\sqrt{\frac{\langle \nabla f^i(x_t), g_t\rangle^2}{M_i\alpha_t^i\|g_t\|^2}+1} + \frac{\langle \nabla f^i(x_t), g_t\rangle}{\sqrt{M_i\alpha_t^i}\|g_t\|}\right)\\
&= \sqrt{\frac{\alpha_t^i}{M_i}}\frac{1}{\sqrt{\frac{\langle \nabla f^i(x_t), g_t\rangle^2}{M_i\alpha_t^i\|g_t\|^2}+1} - \frac{\langle \nabla f^i(x_t), g_t\rangle}{\sqrt{M_i\alpha_t^i}\|g_t\|}} = \frac{\alpha_t^i}{\sqrt{\frac{\langle \nabla f^i(x_t), g_t\rangle^2}{\|g_t\|^2} + M_i\alpha_t^i} - \frac{\langle \nabla f^i(x_t), g_t\rangle}{\|g_t\|}}\\
&= \frac{\alpha_t^i}{\sqrt{(\theta_t^i)^2 + M_i\alpha_t^i} - \theta_t^i}
\end{aligned}$$

Therefore, the condition $\gamma_t \leq \min_{i\in[1,m]}\left\{\frac{\alpha_t^i}{\sqrt{(\theta_t^i)^2+M_i\alpha_t^i}-\theta_t^i}\right\}\frac{1}{\|g_t\|}$ is sufficient for $f^i(x_{t+1}) \leq \frac{f^i(x_t)}{2}$. Using the Cauchy-Schwartz inequality, we can simplify this condition (but making it more conservative):

$$(*) \geq \frac{\alpha_t^i}{\sqrt{(\theta_t^i)^2 + \alpha_t M_i} + |\theta_t^i|} \geq \frac{\alpha_t^i}{2|\theta_t^i| + \sqrt{\alpha_t^i M_i}}.$$

■

A.4 Proof of the local smoothness

Proof Let us define the hessian of the log-barrier $B_\eta(x)$ by $H_B(y)$ at the region $\mathcal{S}(x_t)$ around x_t such that $\mathcal{S}(x_t) := \{y \in \mathbb{R}^d | y = x_t - ug_t, u \in [0, \gamma_t], f^i(y) \leq \frac{f^i(x_t)}{2} \forall i \in [m]\}$. Note that by definition of the log barrier, the Hessian of it at the point $y \in \mathcal{S}(x_t)$ is given by

$$H_B(y) = \nabla^2 B_\eta(y) = \nabla^2 f^0(y) + \sum_{i=1}^m \eta\frac{\nabla^2 f^i(y)}{-f^i(y)} + \sum_{i=1}^m \eta\frac{\nabla f^i(y)\nabla f^i(y)^T}{(-f^i(y))^2}.$$

From the above and the fact that for any $y \in Y_t$ we have $-f^i(y) \geq 0.5\alpha_t$, we get

$$\begin{aligned}
|g_t^T H_B(y)g_t| &\leq M_0\|g_t\|^2 + \eta\sum_{i=1}^m \frac{M_i}{0.5\alpha_t^i}\|g_t\|^2 + \eta\sum_{i=1}^m \frac{(\nabla f^i(y)^T g_t/\|g_t\|)^2}{(0.5\alpha_t^i)^2}\|g_t\|^2\\
&\leq \|g_t\|^2\left(M_0 + \eta\sum_{i=1}^m\frac{M_i}{0.5\alpha_t^i} + \eta\sum_{i=1}^m\frac{\langle \nabla f^i(y), g_t\rangle^2/\|g_t\|^2}{(0.5\alpha_t^i)^2}\right)\\
&\leq \|g_t\|^2\left(M_0 + 2\eta\sum_{i=1}^m\frac{M_i}{\alpha_t^i} + 4\eta\sum_{i=1}^m\frac{\langle \nabla f^i(y), g_t\rangle^2}{(\alpha_t^i)^2\|g_t\|^2}\right).
\end{aligned}$$

Thus,

$$M_2(x_t) = M_0 + 2\eta \sum_{i=1}^m \frac{M_i}{\alpha_t^i} + 4\eta \sum_{i=1}^m \frac{\langle \nabla f^i(y), g_t\rangle^2}{(\alpha_t^i)^2 \|g_t\|^2}.$$

Note that $\nabla f^i(y)$ is unknown, therefore we have to bound it based on $\nabla f^i(x_t)$. Observe that

$$\begin{aligned}&\langle \nabla f^i(y), g_t\rangle^2 \\ &= \langle \nabla f^i(x_t), g_t\rangle^2 + \langle \nabla f^i(y) - \nabla f^i(x_t), g_t\rangle^2 + 2\langle \nabla f^i(x_t), g_t\rangle \langle \nabla f^i(y) - \nabla f^i(x_t), g_t\rangle.\end{aligned} \tag{43}$$

Using the M_i-smoothness of $f^i(x)$ we get:

$$\|\nabla f^i(x_t) - \nabla f^i(y)\| \leq M_i \|x_t - y\| \leq M_i \gamma_t \|g_t\|.$$

Recall that we choose the step-size such that:

$$\gamma_t \leq \min_{i\in[m]} \left\{ \frac{\alpha_t^i}{2|\theta_t| + \sqrt{\alpha_t^i M^i}} \right\} \frac{1}{\|g_t\|_2}.$$

Hence,

$$\|\nabla f^i(x_t) - \nabla f^i(y)\| \leq M_i \|x_t - y\| \leq \frac{M_i \alpha_t^i}{2|\theta_t| + \sqrt{\alpha_t^i M_i}} \leq \frac{M_i \alpha_t^i}{\sqrt{\alpha_t^i M_i}} = \sqrt{\alpha_t^i M_i}.$$

Then, using Equation (43) and $\langle \nabla f^i(x_t), g_t\rangle^2 = (\theta_t^i)^2 \|g_t\|^2$ we get:

$$\begin{aligned}\langle \nabla f^i(y), g_t\rangle^2 &\leq |\theta_t^i|^2 \|g_t^i\|^2 + \alpha_t^i M_i \|g_t^i\|^2 + 2|\theta_t^i| \sqrt{\alpha_t^i M^i} \|g_t^i\|^2 \qquad (44)\\ &\leq 2|\theta_t^i|^2 \|g_t^i\|^2 + 2\alpha_t^i M^i \|g_t^i\|^2. \qquad (45)\end{aligned}$$

The above inequality is due to the fact that $a^2 + 2ab + b^2 \leq 2a^2 + 2b^2$. Thus, we finally obtain

$$\begin{aligned}M_2(x_t) &= M_0 + 2\eta \sum_{i=1}^m \frac{M_i}{\alpha_t^i} + 4\eta \sum_{i=1}^m \frac{2|\theta_t^i|^2 \|g_t^i\|^2 + 2\alpha_t^i M^i \|g_t^i\|^2}{(\alpha_t^i)^2 \|g_t\|^2} \\ &= M_0 + 10\eta \sum_{i=1}^m \frac{M_i}{\alpha_t^i} + 8\eta \sum_{i=1}^m \frac{(\theta_t^i)^2}{(\alpha_t^i)^2}.\end{aligned}$$

■

A.5 Proof of Fact 1

Fact 1 *Let Assumptions 2, 3 hold, and Assumption 4 hold with $\rho \geq \eta$, and let $\hat{\sigma}_i(n) \leq \frac{(\alpha_t^i)^2 L}{2\eta\sqrt{\ln\frac{1}{\delta}}}$, $\hat{b}_i \leq \frac{(\alpha_t^i)^2 L}{2\eta}$, and $\sigma_i(n) \leq \frac{\alpha_t^i}{2\sqrt{\ln\frac{1}{\delta}}}$. If at iteration t we have $\min_{i\in[m]} \alpha_t^i \leq \bar{c}\eta$ with $\bar{c} := \frac{l}{L}\frac{1}{2m+1}$, then, for the next iteration $t+1$ we get $\prod_{i\in\mathcal{I}} \alpha_{t+1}^i \geq \prod_{i\in\mathcal{I}} \alpha_t^i$ for any $\mathcal{I} : \mathcal{I}_t \subseteq \mathcal{I}$ with $\mathcal{I}_t := \{i \in [m] : \alpha_t^i \leq \eta\}$.*

Proof Using the local smoothness of the log barrier, we can see:

$$
\begin{aligned}
\eta \sum_{i\in\mathcal{I}_t} -\log \alpha^i_{t+1} &\leq \eta \sum_{i\in\mathcal{I}_t} -\log \alpha^i_t - \gamma_t\langle \eta \sum_{i\in\mathcal{I}_t} \frac{\nabla f^i(x_t)}{\alpha^i_t}, g_t\rangle + \frac{M_2(x_t)}{2}\gamma_t^2\|g_t\|^2 \\
&\leq \eta \sum_{i\in\mathcal{I}_t} -\log \alpha^i_t + \gamma_t \left(-\langle \eta \sum_{i\in\mathcal{I}_t} \frac{\nabla f^i(x_t)}{\alpha^i_t}, g_t\rangle + \frac{1}{2}\|g_t\|^2\right) \\
&= \eta \sum_{i\in\mathcal{I}_t} -\log \alpha^i_t + \frac{\gamma_t\eta^2}{2}\left(2\langle A, A+B\rangle + \|A+B\|^2\right) \\
&= \eta \sum_{i\in\mathcal{I}_t} -\log \alpha^i_t + \frac{\gamma_t\eta^2}{2}\left(\|B\|^2 - \|A\|^2\right),
\end{aligned} \tag{46}
$$

where $g_t = A + B$, with $A := \sum_{i\in\mathcal{I}_t} \frac{\nabla f^i(x_t)}{\alpha^i_t}$ and $B := \frac{g_t}{\eta} - \sum_{i\in\mathcal{I}_t} \frac{\nabla f^i(x_t)}{\alpha^i_t}$. Using Assumption 4 we obtain a lower bound on $\|A\|$:

$$
\|A\| = \left\| \sum_{i\in\mathcal{I}_t} \frac{\nabla f^i(x_t)}{\alpha^i_t} \right\| \geq \langle \sum_{i\in\mathcal{I}_t} \frac{\nabla f^i(x_t)}{\alpha^i_t}, s_x\rangle \geq \sum_{i\in\mathcal{I}_t} \frac{\langle \nabla f^i(x_t), s_x\rangle}{\alpha^i_t} \geq \sum_{i\in\mathcal{I}_t} \frac{l}{\alpha^i_t}. \tag{47}
$$

The second part $\|B\|$ we can upper bound with high probability $1-\delta$ using the definition of $\mathcal{I}_t$ as follows (since $\forall \alpha^j_t \notin \mathcal{I}_t$ we have $\alpha^j_t \geq \eta$, therefore $\bar{\alpha}^j_t \geq \eta/2$ for $\sigma_i(n) \leq \frac{\alpha^i_t}{2\sqrt{\ln\frac{1}{\delta}}}$):

$$
\begin{aligned}
\|B\| &= \left\| \frac{G^0_n(x_t,\xi_t)}{\eta} + \sum_{j\notin\mathcal{I}_t} \frac{G^j_n(x_t,\xi_t)}{\bar{\alpha}^j_t} + \sum_{i\in\mathcal{I}_t} \frac{G^i_n(x_t,\xi_t)}{\bar{\alpha}^i_t} - \sum_{i\in\mathcal{I}_t} \frac{\nabla f^i(x_t)}{\alpha^i_t} \right\| \\
&\leq \frac{\max_i \|G^i_n(x_t,\xi_t)\|}{\eta}\left(1 + 2(m - |\mathcal{I}_t|)\right) + \sum_{i\in\mathcal{I}_t}\left(\frac{\hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}} + \hat{b}_i}{\alpha^i_t} + \|G^i_n(x_t,\xi_t)\| \left| \frac{1}{\bar{\alpha}^i_t} - \frac{1}{\alpha^i_t}\right| \right) \\
&\leq \frac{\max_i \|G^i_n(x_t,\xi_t)\|}{\eta}\left(1 + 2(m - |\mathcal{I}_t|)\right) + \sum_{i\in\mathcal{I}_t}\left(\frac{\hat{\sigma}_i(n)\sqrt{\ln\frac{1}{\delta}} + \hat{b}_i}{\alpha^i_t} + \|G^i_n(x_t,\xi_t)\| \frac{\sigma_i(n)\sqrt{\ln\frac{1}{\delta}}}{\bar{\alpha}^i_t\alpha^i_t} \right) \\
&\leq \frac{L}{\eta}\left(2m+1\right),
\end{aligned} \tag{48}
$$

for $\hat{\sigma}_i(n) \leq \frac{\alpha^i_t \max\|G^i_n(x_t,\xi_t)\|}{2\eta\sqrt{\ln\frac{1}{\delta}}}$, $\hat{b}_i \leq \frac{\alpha^i_t \max\|G^i_n(x_t,\xi_t)\|}{2\eta}$, and $\sigma_i(n) \leq \frac{(\alpha^i_t)^2}{2\eta\sqrt{\ln\frac{1}{\delta}}}$, implying $\bar{\alpha}^i_t \geq \alpha^i_t/2$ and using $\|G^i_n(x_t,\xi_t)\| \leq L$. Then, if $\min \alpha^i_t \leq \bar{c}\eta$, we have $\sum_{i\in\mathcal{I}_t} \frac{1}{\alpha^i_t} \geq \frac{1}{\bar{c}\eta} = \frac{L}{l\eta}\left(2m+1\right)$, and therefore with high probability $\|B\| \leq \|A\|$. Then we get (46), that implies

$$
\prod_{i\in\mathcal{I}_t} \alpha^i_{t+1} \geq \prod_{i\in\mathcal{I}_t} \alpha^i_t. \tag{49}
$$

Moreover, using the same reasoning, we can prove that

$$
\prod_{i\in\mathcal{I}} \alpha^i_{t+1} \geq \prod_{i\in\mathcal{I}} \alpha^i_t. \tag{50}
$$

for any subset of indices $\mathcal{I} \subseteq [m]$ such that $\mathcal{I}_t \subseteq \mathcal{I}$. ■

A.6 Lower bound on γ_t

Here we assume $\underline{\alpha}_t^i \geq c\eta$. Recall that

$$\gamma_t = \min\left\{ \min_{i\in[m]} \left\{ \frac{\underline{\alpha}_t^i}{2|\hat{\theta}_t^i| + \sqrt{\underline{\alpha}_t^i M_i}} \right\} \frac{1}{\|g_t\|}, \frac{1}{\hat{M}_2(x_t)} \right\}.$$

where

$$\hat{M}_2(x_t) = M_0 + 10\eta \sum_{i=1}^m \frac{M_i}{\underline{\alpha}_t^i} + 8\eta \sum_{i=1}^m \frac{(\hat{\theta}_t^i)^2}{(\underline{\alpha}_t^i)^2}.$$

We get the lower bound by constructing a bound on both of the terms inside the minimum.
1) We have $\hat{M}_2(x_t) \leq \left(1 + 10\frac{m}{c}\right) M + 8\frac{mL^2}{\eta c^2}$, by definition of $\hat{M}_2(x_t)$, which implies

$$\frac{1}{\hat{M}_2(x_t)} \geq \eta \left(\frac{1}{\frac{8m}{c^2}L^2 + \eta(1 + 10\frac{m}{c})M} \right)$$

2) Using definition of g_t, we get $\|g_t\| \leq L_0 + \sum_{i=1}^m \frac{L_i}{c}$. Hence, we can bound

$$\min_{i\in[m]} \left\{ \frac{\underline{\alpha}_t^i}{2|\hat{\theta}_t^i| + \sqrt{\underline{\alpha}_t^i M_i}} \right\} \frac{1}{\|g_t\|} \geq \frac{c\eta}{(2L + \sqrt{Mc\eta})L(1 + \frac{m}{c})}.$$

Therefore,

$$\gamma_t \geq \frac{\eta}{2} \min\left\{ \frac{1}{\frac{4m}{c^2}L^2 + \eta(0.5 + 5\frac{m}{c})M}, \frac{1}{L^2(\frac{1}{c} + \frac{m}{c^2}) + 0.5\sqrt{\frac{M\eta}{cL^2}}L^2(1 + \frac{m}{c})} \right\},$$

$$\gamma_t \geq \frac{\eta}{2L^2(1 + \frac{m}{c})} \min\left\{ \frac{1}{\frac{4}{c} + \frac{5M\eta}{L^2}}, \frac{1}{\frac{1}{c} + \sqrt{\frac{M\eta}{4cL^2}}} \right\}.$$

$$\gamma_t \geq \frac{c\eta}{2L^2(1 + \frac{m}{c})} \min\left\{ \frac{1}{4 + \frac{5Mc\eta}{L^2}}, \frac{1}{1 + \sqrt{\frac{Mc\eta}{4L^2}}} \right\}.$$

Finally, the bound is

$$\gamma_t \geq \eta C$$

with

$$C := \frac{c\eta}{2L^2(1 + \frac{m}{c})} \min\left\{ \frac{1}{4 + \frac{5Mc\eta}{L^2}}, \frac{1}{1 + \sqrt{\frac{Mc\eta}{4L^2}}} \right\}.$$

A.7 Proof of Lemma 9

Proof From Fact 2 it follows that

$$\forall x \in \mathcal{X}\ \exists s_x = \frac{x - x_0}{\|x - x_0\|} \in \mathbb{R}^d: \ \langle s_x, \nabla f^i(x)\rangle \geq \frac{\beta}{2R} \quad \forall i \in \mathcal{I}_{\beta/2}(x).$$

Let $\hat{x}$ be an approximately optimal point for the log barrier: $B_\eta(\hat{x}) - B_\eta(x^*_\eta) \leq \eta$, that is equivalent to:

$$f^0(\hat{x}) + \eta \sum_{i=1}^m - \log(-f^i(\hat{x})) - f^0(x^*_\eta) - \eta \sum_{i=1}^m - \log(-f^i(x^*_\eta)) \leq \eta.$$

Then, for the objective function we have the following bound:

$$f^0(\hat{x}) - f^0(x^*_\eta) \leq \eta + \eta \sum_{i=1}^m - \log \frac{-f^i(x^*_\eta)}{-f^i(\hat{x})}. \tag{51}$$

The optimal point for the log barrier x^*_η must satisfy the stationarity condition

$$\nabla B_\eta(x^*_\eta) = \nabla f^0(x^*_\eta) + \eta \sum_{i=1}^m \frac{\nabla f^i(x^*_\eta)}{-f^i(x^*_\eta)} = 0.$$

By carefully rearranging the above, we obtain

$$\sum_{i \in \mathcal{I}_{\beta/2}(x^*_\eta)} \frac{\nabla f^i(x^*_\eta)}{-f^i(x^*_\eta)} + \sum_{i \notin \mathcal{I}_{\beta/2}(x^*_\eta)} \frac{\nabla f^i(x^*_\eta)}{-f^i(x^*_\eta)} = \frac{-\nabla f^0(x^*_\eta)}{\eta}.$$

By taking a dot product of both sides of the above equation with $s_x = \frac{x^*_\eta - x_0}{\|x^*_\eta - x_0\|}$, using the Lipschitz continuity we get for x^*_η:

$$\frac{1}{\min_i\{-f^i(x^*_\eta)\}} \sum_{i \in \mathcal{I}_{\beta/2}(x^*_\eta)} \langle \nabla f^i(x^*_\eta), s_x\rangle \frac{\min_i\{-f^i(x^*_\eta)\}}{-f^i(x^*_\eta)} \tag{52}$$

$$= \frac{\langle -\nabla f^0(x^*_\eta), s_x\rangle}{\eta} - \sum_{i \notin \mathcal{I}_{\beta/2}(x^*_\eta)} \frac{\langle \nabla f^i(x^*_\eta), s_x\rangle}{-f^i(x^*_\eta)} \leq \frac{mL}{\eta}. \tag{53}$$

From the above, using Fact 2, we get

$$\min\{-f^i(x^*_\eta)\} \geq \frac{\eta\beta}{2mLR}.$$

Hence, combining the above with (51) we get the following relation of point $\hat{x}$ and point x^*_η optimal for the log barrier:

$$f^0(\hat{x}) - f^0(x^*_\eta) \leq \eta + \eta \sum_{i=1}^m \log \frac{-f^i(\hat{x})}{-f^i(x^*_\eta)} \leq \eta \left(1 + m \log\left(\frac{2mLR\hat{\beta}}{\eta\beta}\right)\right). \tag{54}$$

Next, note that the Lagrangian $\mathcal{L}(x,\lambda)$ is a convex function over x and concave over λ. Hence, for $(x^*_\eta,\lambda^*_\eta) := \left(x^*_\eta, \left[\frac{\eta}{-f^1(x^*_\eta)},\dots,\frac{\eta}{-f^m(x^*_\eta)}\right]^T\right)$ we have

$$\mathcal{L}(x^*_\eta,\lambda^*_\eta) - \mathcal{L}(x^*,\lambda^*) \le \mathcal{L}(x^*_\eta,\lambda^*_\eta) - \mathcal{L}(x^*,\lambda^*_\eta) \le \langle \nabla_x \mathcal{L}(x^*_\eta,\lambda^*_\eta), \lambda^*_\eta - x^* \rangle \le 0.$$

Expressing $\mathcal{L}(x^*_\eta,\lambda^*_\eta)$ and $\mathcal{L}(x^*,\lambda^*)$ and exploiting the fact that $\nabla B_\eta(x^*_\eta) = \nabla_x \mathcal{L}(x^*_\eta,\lambda^*_\eta) = 0$, we obtain $\mathcal{L}(x^*_\eta,\lambda^*_\eta) - \mathcal{L}(x^*,\lambda^*) = f^0(x^*_\eta) - f^0(x^*) - m\eta \le 0$. Consequently, we have $f^0(x^*_\eta) - f^0(x^*) \le m\eta$. Combining the above and (54), we get

$$f^0(\hat{x}) - \min_{x\in\mathcal{X}} f^0(x) \le \eta + \eta m \log\left(\frac{2mLR\hat{\beta}}{\eta\beta}\right) + m\eta.$$

■

A.8 Zeroth-order estimator properties proof

The deviation of the gradient estimators $G^i(x_t,\nu) - \nabla f^i_\nu(x_t)$, by definition can be expressed as follows for $i = 0,\dots,m$

$$G^i(x_t,\nu) - \nabla f^i_\nu(x_t) = \frac{1}{n_t}\sum_{j=1}^{n_t}\left[\underbrace{\left(d\frac{f^i(x_k+\nu s_{tj}) - f^i(x_t)}{\nu}s_{tj} - \nabla f^i_\nu(x_t)\right)}_{v^i_j} + \underbrace{d\frac{\xi^{i+}_{tj} - \xi^{i-}_{tj}}{\nu}s_{tj}}_{u^i_j}\right], \tag{55}$$

where the first term under the summation v^i_j is dependent only on random s_{tj}, however the second term is dependent on both random variables coming from the noise $\xi^{i\pm}_{tj}$ and from the direction s_{tj}.

Then, using the fact that the additive noise $\xi^{i\pm}_{tj}$ is zero-mean and independent on s_{tj}, we get:

$$\mathbb{E}\left\|G^i_{\nu,n}(x_t,\xi) - \nabla f^i_\nu(x_t)\right\|^2 = \mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^n v^i_j\right\|^2 + \mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^n u^i_j\right\|^2 \tag{56}$$

Using the result of Lemma 2.10 (Berahas et al., 2021), we can bound the first part of the above expression $\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^n v^i_j\right\|^2$:

$$\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^n v^i_j\right\|^2 \le \frac{3d^2}{n}\left(\frac{\|\nabla f^i(x)\|^2}{d} + \frac{M_i^2\nu^2}{4}\right). \ \forall i \in \{0,\dots,m\}. \tag{57}$$

The second part u_j^i is zero-mean, hence does not influence the bias. Indeed, using the independence of $\xi_{tj}^{j\pm}$ and s_{tj} we derive

$$\mathbb{E}\sum_{j=1}^{n_t} u_j^i = \frac{d}{\nu}\mathbb{E}\left(\sum_{j=1}^{n_t}(\xi_{tj}^{i+} - \xi_{tj}^{i-})s_{tj}\right) = 0. \tag{58}$$

Its variance can be bounded as follows, using $\|s_{tj}\| = 1$:

$$\mathbb{E}\left\|\frac{1}{n}\sum_{j=1}^{n} u_j^i\right\|^2 = \mathbb{E}\frac{d^2}{\nu^2 n^2}\left\|\sum_{j=1}^{n}(\xi_{tj}^{i+} - \xi_{tj}^{i-})s_{tj}\right\|^2 \leq 4\frac{d^2}{\nu^2 n^2}\sum_{j=1}^{n}\mathbb{E}\|\xi_{tj}^{i+}\|^2\|s_{tj}\|^2 \leq 4\frac{d^2\sigma^2}{\nu^2 n}. \tag{59}$$

From the above, and Lemma 2.10 (Berahas et al., 2021) the statement of the Lemma follows directly.

References

Joshua Achiam, David Held, Aviv Tamar, and Pieter Abbeel. Constrained policy optimization, 2017.

E. Altman. *Constrained Markov Decision Processes.* Chapman and Hall, 1999.

Sanae Amani, Mahnoosh Alizadeh, and Christos Thrampoulidis. Linear stochastic bandits under safety constraints. *arXiv preprint arXiv:1908.05814*, 2019.

Yossi Arjevani, Yair Carmon, John C Duchi, Dylan J Foster, Nathan Srebro, and Blake Woodworth. Lower bounds for non-convex stochastic optimization. *arXiv preprint arXiv:1912.02365*, 2019.

Yarden As, Ilnura Usmanova, Sebastian Curi, and Andreas Krause. Constrained policy optimization via bayesian world models. *ArXiv*, 2022. URL `https://arxiv.org/abs/2201.09802`.

Francis Bach and Vianney Perchet. Highly-smooth zero-th order online optimization. In *Conference on Learning Theory*, pages 257–283, 2016.

Krishnakumar Balasubramanian and Saeed Ghadimi. Zeroth-order (non)-convex stochastic optimization via conditional gradient and gradient updates. In *Advances in Neural Information Processing Systems*, pages 3455–3464, 2018.

Albert Berahas, Liyuan Cao, Krzysztof Choromanski, and Katya Scheinberg. A theoretical and empirical comparison of gradient approximations in derivative-free optimization. *Foundations of Computational Mathematics*, 05 2021. doi: 10.1007/s10208-021-09513-z.

Felix Berkenkamp, Andreas Krause, and Angela P Schoellig. Bayesian optimization with safety constraints: safe and automatic parameter tuning in robotics. *arXiv preprint arXiv:1602.04450*, 2016a.

Felix Berkenkamp, Angela P. Schoellig, and Andreas Krause. Safe controller optimization for quadrotors with gaussian processes. In *2016 IEEE International Conference on Robotics and Automation (ICRA)*, pages 491–496, 2016b. doi: 10.1109/ICRA.2016.7487170.

Felix Berkenkamp, Matteo Turchetta, Angela P. Schoellig, and Andreas Krause. Safe model-based reinforcement learning with stability guarantees, 2017.

Felix Berkenkamp, Andreas Krause, and Angela P. Schoellig. Bayesian optimization with safety constraints: Safe and automatic parameter tuning in robotics, 2020.

Sébastien Bubeck, Yin Tat Lee, and Ronen Eldan. Kernel-based methods for bandit convex optimization. In *Proceedings of the 49th Annual ACM SIGACT Symposium on Theory of Computing*, pages 72–85, 2017.

Lin Chen, Mingrui Zhang, and Amin Karbasi. Projection-free bandit convex optimization. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2047–2056. PMLR, 2019.

Yinlam Chow, Mohammad Ghavamzadeh, Lucas Janson, and Marco Pavone. Risk-constrained reinforcement learning with percentile risk criteria. *CoRR*, abs/1512.01629, 2015. URL `http://arxiv.org/abs/1512.01629`.

Kurtland Chua, Roberto Calandra, Rowan McAllister, and Sergey Levine. Deep reinforcement learning in a handful of trials using probabilistic dynamics models. *CoRR*, abs/1805.12114, 2018. URL `http://arxiv.org/abs/1805.12114`.

Djork-Arné Clevert, Thomas Unterthiner, and Sepp Hochreiter. Fast and accurate deep network learning by exponential linear units (elus), 2015. URL `https://arxiv.org/abs/1511.07289`.

Gal Dalal, Krishnamurthy Dvijotham, Matej Vecerik, Todd Hester, Cosmin Paduraru, and Yuval Tassa. Safe exploration in continuous action spaces, 2018.

Marc Peter Deisenroth and Carl Edward Rasmussen. Pilco: A model-based and data-efficient approach to policy search. In *Proceedings of the 28th International Conference on International Conference on Machine Learning*, ICML'11, page 465–472, Madison, WI, USA, 2011. Omnipress. ISBN 9781450306195.

John C. Duchi, Michael I. Jordan, Martin J. Wainwright, and Andre Wibisono. Optimal rates for zero-order convex optimization: The power of two function evaluations. *IEEE Transactions on Information Theory*, 61(5):2788–2806, 2015. doi: 10.1109/TIT.2015.2409256.

Eloísa Díaz-Francés and Francisco Rubio. On the existence of a normal approximation to the distribution of the ratio of two independent normal random variables. *Statistical Papers*, 54(2):309–323, May 2013. doi: 10.1007/s00362-012-0429-2. URL `https://ideas.repec.org/a/spr/stpapr/v54y2013i2p309-323.html`.

David Eriksson and Martin Jankowiak. High-dimensional bayesian optimization with sparse axis-aligned subspaces, 2021. URL `https://arxiv.org/abs/2103.00349`.

David Eriksson and Matthias Poloczek. Scalable constrained bayesian optimization. In Arindam Banerjee and Kenji Fukumizu, editors, *Proceedings of The 24th International Conference on Artificial Intelligence and Statistics*, volume 130 of *Proceedings of Machine Learning Research*, pages 730–738. PMLR, 13–15 Apr 2021. URL `https://proceedings.mlr.press/v130/eriksson21a.html`.

Mahyar Fazlyab, Alexander Robey, Hamed Hassani, Manfred Morari, and George Pappas. Efficient and accurate estimation of lipschitz constants for deep neural networks. *Advances in Neural Information Processing Systems*, 32, 2019.

Mohammad Fereydounian, Zebang Shen, Aryan Mokhtari, Amin Karbasi, and Hamed Hassani. Safe learning under uncertain objectives and constraints. *arXiv preprint arXiv:2006.13326*, 2020.

Abraham D Flaxman, Adam Tauman Kalai, and H Brendan McMahan. Online convex optimization in the bandit setting: gradient descent without a gradient. In *Proceedings* *of the sixteenth annual ACM-SIAM symposium on Discrete algorithms*, pages 385–394. Society for Industrial and Applied Mathematics, 2005.

Peter I. Frazier. A tutorial on bayesian optimization, 2018. URL `https://arxiv.org/abs/1807.02811`.

Dan Garber and Ben Kretzu. Improved regret bounds for projection-free bandit convex optimization. In *International Conference on Artificial Intelligence and Statistics*, pages 2196–2206. PMLR, 2020.

Danijar Hafner, Timothy Lillicrap, Mohammad Norouzi, and Jimmy Ba. Mastering atari with discrete world models, 2021.

Nikolaus Hansen and Andreas Ostermeier. Completely derandomized self-adaptation in evolution strategies. *Evol. Comput.*, 9(2):159–195, 2001. URL `http://dblp.uni-trier.de/db/journals/ec/ec9.html#HansenO01`.

Elad Hazan and Haipeng Luo. Variance-reduced and projection-free stochastic optimization. In *International Conference on Machine Learning*, pages 1263–1271, 2016.

Oliver Hinder and Yinyu Ye. A one-phase interior point method for nonconvex optimization. *arXiv preprint arXiv:1801.03072*, 2018.

Oliver Hinder and Yinyu Ye. A polynomial time log barrier method for problems with nonconvex constraints. *arXiv preprint: https://arxiv.org/pdf/1807.00404.pdf*, 2019.

Anatoli B. Juditsky, Guanghui Lan, Arkadii S. Nemirovski, and Alexander Shapiro. Stochastic Approximation approach to Stochastic Programming. Research report, LJK, 2013. URL `https://hal.archives-ouvertes.fr/hal-00853911`. http://www.optimization-online.org/DB_HTML/2007/09/1787.html.

James Kennedy and Russell C. Eberhart. Particle swarm optimization. In *Proceedings of the IEEE International Conference on Neural Networks*, pages 1942–1948, 1995.

Johannes Kirschner, Mojmir Mutnỳ, Nicole Hiller, Rasmus Ischebeck, and Andreas Krause. Adaptive and safe bayesian optimization in high dimensions via one-dimensional subspaces. *arXiv preprint arXiv:1902.03229*, 2019.

Torsten Koller, Felix Berkenkamp, Matteo Turchetta, and Andreas Krause. Learning-based model predictive control for safe exploration. In *2018 IEEE Conference on Decision and Control (CDC)*, pages 6059–6066. IEEE, 2018.

Guanghui Lan. *First-order and Stochastic Optimization Methods for Machine Learning.* 01 2020. ISBN 978-3-030-39567-4. doi: 10.1007/978-3-030-39568-1.

Timothy P. Lillicrap, Jonathan J. Hunt, Alexander Pritzel, Nicolas Heess, Tom Erez, Yuval Tassa, David Silver, and Daan Wierstra. Continuous control with deep reinforcement learning, 2015. URL `https://arxiv.org/abs/1509.02971`.

M. Luersen, R. Le Riche, and F. A Guyon. Constrained, globalized, and bounded nelder–mead method for engineering optimization. *Struct Multidisc Optim 27*, page 43–54, 2004. doi: 10.1007/s00158-003-0320-9.

O.L Mangasarian and S Fromovitz. The fritz john necessary optimality conditions in the presence of equality and inequality constraints. *Journal of Mathematical Analysis and Applications*, 17(1):37–47, 1967. ISSN 0022-247X. doi: https://doi.org/10.1016/0022-247X(67)90163-1. URL `https://www.sciencedirect.com/science/article/pii/0022247X67901631`.

Riccardo Moriconi, Marc P. Deisenroth, and K. S. Sesh Kumar. High-dimensional bayesian optimization using low-dimensional feature spaces, 2019. URL `https://arxiv.org/abs/1902.10675`.

J. A. Nelder and R. Mead. A Simplex Method for Function Minimization. *The Computer Journal*, 7(4):308–313, 01 1965. ISSN 0010-4620. doi: 10.1093/comjnl/7.4.308. URL `https://doi.org/10.1093/comjnl/7.4.308`.

A. S. Nemirovsky and D. B. Yudin. Problem complexity and method efficiency in optimization. *SIAM Review*, 27(2):264–265, 1985. doi: 10.1137/1027074. URL `https://doi.org/10.1137/1027074`.

Tingting Ni and Maryam Kamgarpour. Interior point constrained reinforcement learning with global convergence guarantees, 2023.

Jorge Nocedal and Stephen Wright. *Numerical optimization*. Springer Science & Business Media, 2006.

Christopher Price. A modified nelder-mead barrier method for constrained optimization. *Numerical Algebra, Control, and Optimization*, 11, 01 2019. doi: 10.3934/naco.2020058.

Carl Edward Rasmussen and Christopher K. I. Williams. *Gaussian Processes for Machine Learning (Adaptive Computation and Machine Learning)*. The MIT Press, 2005. ISBN 026218253X.

Alex Ray, Joshua Achiam, and Dario Amodei. Benchmarking Safe Exploration in Deep Reinforcement Learning. 2019.

Ingo Rechenberg. Evolution strategy: Nature's way of optimization. In H. W. Bergmann, editor, *Optimization: Methods and Applications, Possibilities and Limitations*, pages 106–126, Berlin, Heidelberg, 1989. Springer Berlin Heidelberg. ISBN 978-3-642-83814-9.

John Schulman, Philipp Moritz, Sergey Levine, Michael Jordan, and Pieter Abbeel. High-dimensional continuous control using generalized advantage estimation, 2015. URL `https://arxiv.org/abs/1506.02438`.

Ohad Shamir. On the complexity of bandit and derivative-free stochastic convex optimization. In *Conference on Learning Theory*, pages 3–24, 2013.

Alexander Shapiro, Darinka Dentcheva, and Andrzej Ruszczyński. *Lectures on Stochastic Programming*. SIAM, Philadelphia, PA, USA, 2009.

Jasper Snoek, Oren Rippel, Kevin Swersky, Ryan Kiros, Nadathur Satish, Narayanan Sundaram, Md. Mostofa Ali Patwary, Prabhat, and Ryan P. Adams. Scalable bayesian optimization using deep neural networks, 2015. URL https://arxiv.org/abs/1502.05700.

Niranjan Srinivas, Andreas Krause, Sham Kakade, and Matthias Seeger. Information-theoretic regret bounds for gaussian process optimization in the bandit setting. *IEEE Transactions on Information Theory*, 58(5):3250–3265, May 2012. doi: 10.1109/TIT.2011.2182033.

Rainer Storn and Kenneth V. Price. Differential evolution - a simple and efficient heuristic for global optimization over continuous spaces. *J. Glob. Optim.*, 11(4):341–359, 1997. URL http://dblp.uni-trier.de/db/journals/jgo/jgo11.html#StornP97.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In Francis Bach and David Blei, editors, *Proceedings of the 32nd International Conference on Machine Learning*, volume 37 of *Proceedings of Machine Learning Research*, pages 997–1005, Lille, France, 07–09 Jul 2015a. PMLR. URL https://proceedings.mlr.press/v37/sui15.html.

Yanan Sui, Alkis Gotovos, Joel Burdick, and Andreas Krause. Safe exploration for optimization with gaussian processes. In *International Conference on Machine Learning*, pages 997–1005, 2015b.

Richard S Sutton, David McAllester, Satinder Singh, and Yishay Mansour. Policy gradient methods for reinforcement learning with function approximation. In S. Solla, T. Leen, and K. Müller, editors, *Advances in Neural Information Processing Systems*, volume 12. MIT Press, 2000. URL https://proceedings.neurips.cc/paper/1999/file/464d828b85b0bed98e80ade0a5c43b0f-Paper.pdf.

Ilnura Usmanova, Andreas Krause, and Maryam Kamgarpour. Safe convex learning under uncertain constraints. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 2106–2114, 2019.

Ilnura Usmanova, Andreas Krause, and Maryam Kamgarpour. Safe non-smooth black-box optimization with application to policy search. In *Learning for Dynamics and Control*, pages 980–989, 2020.

Sharan Vaswani, Benjamin Dubois-Taine, and Reza Babanezhad. Towards Noise-adaptive, Problem-adaptive Stochastic Gradient Descent. working paper or preprint, November 2021. URL https://hal.archives-ouvertes.fr/hal-03456663.

Kim P. Wabersich and Melanie N. Zeilinger. A predictive safety filter for learning-based control of constrained nonlinear dynamical systems, 2021.