Proceedings of Machine Learning Research vol 120:1–10, 2020 2nd Annual Conference on Learning for Dynamics and Control

Safe non-smooth black-box optimization with application to policy search

Ilnura Usmanova ILNURAU@CONTROL.EE.ETHZ.CH
Automatic Control Laboratory, ETH Zürich, Switzerland

Andreas Krause KRAUSEA@ETHZ.CH
Machine Learning Institute, ETH Zürich, Switzerland

Maryam Kamgarpour MKAMGAR@CONTROL.EE.ETHZ.CH
Automatic Control Laboratory, ETH Zürich, Switzerland

Editors: A. Bayen, A. Jadbabaie, G. J. Pappas, P. Parrilo, B. Recht, C. Tomlin, M.Zeilinger

Abstract

For safety-critical black-box optimization tasks, observations of the constraints and the objective are often noisy and available only for the feasible points. We propose an approach based on log barriers to find a local solution of a non-convex non-smooth black-box optimization problem $\min f^0(x)$ subject to $f^i(x) \leq 0,\ i = 1, \ldots, m$, guaranteeing constraint satisfaction while learning an optimal solution with high probability. Our proposed algorithm exploits noisy observations to iteratively improve on an initial safe point until convergence. We derive the convergence rate and prove safety of our algorithm. We demonstrate its performance in an application to an iterative control design problem. [1] [2]

1. Introduction

Motivation Machine learning algorithms are increasingly being deployed for safety-critical emerging applications such as autonomous driving, personalized medicine and robotics. In such scenarios, safety and reliability of these algorithms is crucial. When the model is unknown, too complex or unreliable, it is common to adopt a black-box *bandit* setup; our goal is to include safety in these learning techniques.

Related work In the optimization literature, several constrained optimization algorithms exist guaranteeing feasibility of the iterates given just local information about the constraints. These include Feasible Sequential Quadratic Programming (FSQP) (Jian et al., 2005; Luo et al., 2012; Tang et al., 2014), the Method of Feasible Directions (MFD) (Zoutendijk, 1960), and their variations. However, all these methods require first and/or second order information and do not consider the presence of noise. On the other hand, there are many works on derivative-free optimization, including non-convex and non-smooth problems (Balasubramanian and Ghadimi, 2018; Nesterov and Spokoiny, 2017; Ghadimi and Lan, 2013; Lan, 2013), based on finite difference gradient estimation techniques. However, these methods do not guarantee the feasibility of the points where measurements are taken with respect to unknown constraints. This issue can be addressed by interior point methods, where a barrier function is optimized. However, existing work on interior point methods typically require second order information.

Safe learning with zero-th order (bandit) information has been considered in Bayesian Optimization (Berkenkamp et al., 2016; Sui et al., 2015) for non-convex constrained problems. As these works aim to compute a global optimum, they have to solve a nontrivial non-convex subproblem at each iteration.

1. We thank the support of Swiss National Science Foundation, under the grant SNSF 200021_172781, and ERC under the European Union's Horizon 2020 research and innovation programme grant agreement No 815943.
2. In the interest of space, we refer to a companion ArXiv preprint https://arxiv.org/abs/1912.09466 for the proofs.

Problem	Unconstrained & smooth objective	Known linear constraints & smooth objective	Single unknown smooth constraint & smooth objective	Several unknown non-smooth constraints & non-smooth objective
Feedback	2-point bandit feedback	2-point bandit feedback	1-point bandit feedback	1-point bandit feedback
Safety	-	Yes *(known constraints)*	Yes	Yes
Optimality condition	Stationary point: $\mathbb{E}\|\nabla f(x_T)\|_2 \leq \eta$	η-stationary point: $\forall u \in D$ $\mathbb{E}\langle \nabla f(x_T), x_T - u\rangle \leq \eta$	η-approximate KKT point	η-approximate KKT point of the smoothed problem
Number of measurements	$O\left(\frac{d}{\eta^4}\right)$ or $O\left(\frac{d}{\eta^{3.5}}\right) + \tilde{O}\left(\frac{d^4}{\eta^{2.5}}\right)$ (Balasubramanian and Ghadimi, 2018)	$O\left(\frac{d}{\eta^4}\right)$ (Balasubramanian and Ghadimi, 2018)	$\tilde{O}\left(\frac{d^3}{\eta^7}\right)$ (Usmanova et al., 2019)	$\tilde{O}\left(\frac{d^2}{\eta^9}\right)$ **(this work)**

Table 1: Upper bounds on number of zero-th order oracle calls for non-convex smooth optimization algorithms.

Moreover, for most common kernel functions, these algorithms require a number of measurements that is exponential on the dimensionality. This makes Safe Bayesian Optimization methods not always applicable to high dimensional problems. Moreover, appropriately choosing a prior distribution and the kernel parameters might not be a trivial task. Gradient based local methods usually do not suffer from these drawbacks.

First order methods in application to barrier functions in the recent past were considered to have exponential runtime bounds due to the bad behavior of any barrier on the boundary of the feasible set. However, in the recent work (Hinder and Ye, 2019) the authors demonstrated that for smooth problems a gradient descent algorithm with adaptive step size on a log barrier function can be tractable, i.e., present attractive polynomial runtime convergence. Motivated by safe learning problems, the recent work (Usmanova et al., 2019) extended this approach by (Hinder and Ye, 2019) to smooth non-convex optimization problems with zero-th order noisy measurements. However, the bound on the number of measurements was valid only in the case of *a single* smooth constraint function. In this paper, generalizing the approach of (Usmanova et al., 2019), we develop a safe algorithm for the *non-smooth* non-convex constrained optimization problems subject to an arbitrary number of constraint functions. In Table 1, a comparison of our algorithm with existing methods for unconstrained and constrained zero-th order non-convex optimization is provided. In the first two algorithms a 2-point bandit feedback is assumed, i.e., it is possible to measure at several points with the same noise realization. In our algorithm we assume a more realistic and more challenging setup with changing noise at each measurement.

Safe learning is widely used in control of unknown dynamical systems. For example, the work by (Dean et al., 2019) exploited system identification and robust optimal control to learn the safe linear quadratic regulator (LQR) subject to constraints on the state and input trajectories. There are many works aiming at guaraneeing safety while learning the optimal policy and dynamics in non-linear control, such as Fisac et al. (2018); Berkenkamp et al. (2016); Gillula and Tomlin (2012). Often Bayesian optimization approach is used to solve the above problems. However, Bayesian optimization might not enjoy acceptable scalability with dimensionality, thus limiting its applicability to control. Non-smoothness can also appear in some control problems such as bipedal walking, etc (Ames, 2014). In this paper, we consider an application of our method to safe learning for model-free control. We test our algorithm on a low dimensional control system, but theoretically the dependence on the dimensionality is only polynomial and the algorithm can be applied for higher dimensional problems.

Our contributions Our contribution is to propose an algorithm to find an approximate local solution to non-convex non-smooth cost functions subject to non-convex non-smooth constraints. Furthermore, we prove the safety of the approach and derive its convergence rate in expectation in terms of the variance of the noise. Our algorithm is based on the log barrier gradient descent approach. Our convergence is with respect to an approximate stationary point of the smooth approximation of the problem. In the special case, where both the cost function and the constraints are smooth, we establish convergence to an approximate KKT point of the initial problem. We validate the performance of our algorithm in application to a simple model-free control problem.

2. Problem statement

Notations and definitions. Let $\|\cdot\|$ denote the l_2-norm on $\mathbb{R}^d$. A function $f:\mathbb{R}^d \to \mathbb{R}$ is called *L-Lipschitz continuous* if $|f(x)-f(y)| \leq L\|x-y\|_2$. It is called *$M$-smooth* if the gradients $\nabla f(x)$ are M-Lipschitz continuous, i.e., $\|\nabla f(x) - \nabla f(y)\|_2 \leq M\|x-y\|_2$. A random variable ξ is zero-mean σ^2-sub-Gaussian if $\forall \lambda \in \mathbb{R}\ \mathbb{E}\left[e^{\lambda\xi}\right] \leq \exp\left(\frac{\lambda^2\sigma^2}{2}\right)$, which implies that $\mathrm{Var}\left[\xi\right] \leq \sigma^2$ (this can be shown using Tailor expansion). By $\mathbb{S}^d$ and $\mathbb{B}^d$ we denote the unit sphere and the unit ball in $\mathbb{R}^d$, respectively. We denote the characteristic function of a set $\mathcal{X} \subseteq \mathbb{R}^d$ by $\mathbb{I}_{\mathcal{X}} = \begin{cases} 0, & x \in \mathcal{X} \\ +\infty, & x \notin \mathcal{X} \end{cases}$.

Problem formulation We consider safe non-convex non-smooth constrained optimization problem

$$\begin{aligned} &\min_{x\in\mathbb{R}^d} f^0(x) \\ &\text{subject to } \ f^i(x) \leq 0, \ \ i = 1,\ldots,m, \end{aligned} \tag{1}$$

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 L-Lipschitz continuous functions, and can only be accessed at feasible points x. We denote by D the feasible set $D := \{x \in \mathbb{R}^d : f^i(x) \leq 0, i = 1,\ldots,m\}$.

Assumption 1 *The set D has a non-empty interior, and there exists a known starting point x_0 for which $f^i(x_0) < 0$ for $i = 1,\ldots,m$.*

This assumption is common in works on safe learning (Berkenkamp et al., 2016; Sui et al., 2015) or on model-free LQR problems (Fazel et al., 2018; Abbasi-Yadkori et al., 2019).

Information. We assume access to noisy measurements of all cost and constraint function values for any requested feasible point $x \in D$. In particular, the measurements are given by $F^i(x,\xi^i) = f^i(x) + \xi^i,\ \forall i = 0,\ldots,m$ with zero-mean sub-Gaussian noise ξ^i. The ξ^i's are i.i.d. across different measurements.

Goal. The goal of the algorithm is to find an approximate local optimum, using only noisy zeroth-order information. Moreover, it has to guarantee safety, i.e., constraint satisfaction with high probability for all points at which measurements are taken. For differentiable non-convex objective and constraints, the notion of local optimality is captured by KKT condition. In this setting, we show that our algorithm converges to an η-approximate KKT point for any $\eta > 0$ with constants $\tau_1, \tau_2 > 0$, that are fixed and independent on η:

$$\lambda^i, -f^i(x) \geq 0,\ \forall i = 1,\ldots,m \tag{η-KKT.1}$$

$$\lambda^i(-f^i(x)) \leq \tau_1\eta,\ \forall i = 1,\ldots,m \tag{η-KKT.2}$$

$$\|\nabla_x L(x,\lambda)\|_2 \leq \tau_2\eta, \tag{η-KKT.3}$$

where $L(x,\lambda) := f^0(x) + \sum_{i=1}^m \lambda^i f^i(x)$ is the corresponding Lagrangian function. For non-differentiable non-convex objective and constraints, local optimality conditions are less understood. In this case, we show convergence to an approximate KKT point of a corresponding smoothed problem. The smoothing will be described in the approach below. In Corollary 4, we connect the solution of the smoothed problem with an approximate KKT point of the initial problem for the case of differentiable cost function and constraints.

3. Proposed approach

We propose to construct a log barrier for the smooth approximation of problem (1), and then apply the zero-th order stochastic gradient descent with an adaptive step size to minimize it. To estimate the gradient of the smoothed function we sample points around the current iterate, and take measurements at these points. A measurement is denoted as *safe* if the point at which it is taken is feasible with high probability.

3.1. Zero-th order gradient estimation.

Our algorithm uses a randomized zero-th order gradient estimator for cost and constraint functions. For a point x_k the gradient is estimated taking n_k samples uniformly at random on the unit sphere $\mathbb{S}^d$.

$$G^i(x_k,\nu) := \sum_{j=1}^{n_k} \frac{\hat{G}^i_j(x_k,\nu)}{n_k},\ \hat{G}^i_j(x_k,\nu) := d\frac{F^i(x_k+\nu s_{kj},\xi^{i+}_{kj}) - F^i(x_k,\xi^{i-}_{kj})}{\nu} s_{kj} \tag{2}$$

for $i = 0,\dots,m$, where all $\{[\xi^{i+}_{kj},\xi^{i-}_{kj}]\}_{j=1\dots,n_k}$ are i.i.d. sub-Gaussian random variables, $\nu > 0$ is the sampling radius, s_{kj} are the sampled unit vectors. For the sampling radius $\nu \geq 0$ define the ν-*smoothed* estimate of the function $f(x)$ by $f_\nu(x) := \mathbb{E}_b f(x+\nu b)$, where b is uniformly distributed on the unit ball $\mathbb{B}^d$ Then $\mathbb{E}_{s_k,\xi_k} G^i(x_k,\nu) = \nabla f^i_\nu(x_k)$ (Flaxman et al., 2005; Nesterov and Spokoiny, 2017; Balasubramanian and Ghadimi, 2018). This shows that in expectation we can get a gradient of the smooth estimate $f_\nu(x)$ of non-smooth function $f(x)$ using randomized gradient estimator. Hazan et al. (2016) showed the first two of the following properties of $f_\nu(x)$:

1) The gradient $\nabla f_\nu(x)$ is Lipschitz continuous with constant M_ν satisfying $M_\nu \leq \frac{L}{\nu}$.
2) $\forall x \in \mathbb{R}^d\ |f_\nu(x) - f(x)| \leq \nu L$.
3) The function $f_\nu(x)$ is Lipschitz continuous with $L_\nu \leq L$, which implies $\|\nabla f_\nu(x)\| \leq L$ for differentiable $f_\nu(x)$. (For the proof of this property see Appendix A.)

3.2. Smoothed log barrier function.

We address the safe learning problem using the log barriers approach. Define $f^c(x) = \max_{i=1,\dots,m} f^i(x)$, which is in general non-smooth and non-convex. The logarithmic barrier with parameter $\eta > 0$ of the initial problem with the constraints replaced with $f^c(x)$ above is defined as $B_\eta(x) = f^0(x) - \eta\log(-f^c(x))$. We define the *locally smooth* barrier function and its gradient using *smoothed* functions $f^0_\nu(x)$ and $f^c_\nu(x)$:

$$B_{\eta,\nu}(x) = f^0_\nu(x) - \eta\log(-f^c_\nu(x)), \tag{3}$$

$$\nabla B_{\eta,\nu}(x) = \nabla f^0_\nu(x) + \eta\frac{\nabla f^c_\nu(x)}{-f^c_\nu(x)}, \tag{4}$$

It is evident that the gradient of the barrier grows to infinity while converging towards the boundary and hence, the barrier function cannot be smooth. *Local* smoothness of barrier function refers to existence of a value $M_{2,\nu}(x)$ that bounds the change in barrier gradient $\frac{\|\nabla B_{\nu,\eta}(x)-\nabla B_{\nu,\eta}(y)\|}{\|x-y\|}$ for a ball around point x, where $M_{2,\nu}(x)$ is determined later in Appendix H in (26). Our goal is to design an algorithm that converges to a locally optimal point x^* of the smoothed log barrier $B_{\eta,\nu}(x)$, which is basically an unconstrained approximation of constrained smoothed problem: $f^0_\nu(x) + \mathbb{I}_{f^c_\nu(x)\leq 0}$. Then, we show that this point x^* satisfies η-approximate KKT conditions for the smoothed problem $\min_{f^c_\nu(x)\leq 0} f^0_\nu(x)$.

3.3. Log barrier gradient estimator.

First, we need to propose a way to estimate $\nabla B_{\eta,\nu}(x_k) = \nabla f_\nu^0(x_k) + \eta \frac{\nabla f_\nu^c(x_k)}{-f_\nu^c(x)}$. To estimate $\nabla f_\nu^0(x_k)$ and $\nabla f_\nu^c(x_k)$ we can use $G^0(x_k,\nu)$ defined by (2), and $G^c(x_k,\nu)$ defined similarly but with $F^c(x_k,\xi_k) = \max_i F^i(x_k,\xi_k^i)$ instead of $F^i(x_k,\xi_k^i)$. However, to estimate the denominator we propose to use a lower confidence bound on $-f_\nu^c(x_k)$, constructed as follows. Given $\delta > 0$ for $i = 1,\dots,m$ define $\hat{F}^i(x_k) := \frac{\sum_{j=1}^{n_k} F^i(x_k,\xi_{jk}^{i-})}{n_k} + \frac{\sigma}{\sqrt{n_k}}\sqrt{\ln\frac{1}{\delta}}$. We show in Appendix C that $\mathbb{P}\{f^i(x_k) \le \hat{F}^i(x_k)\} \ge 1-\delta$. We define by $\hat{F}_\nu^c(x_k) := \max_{i=1,\dots,m} \hat{F}^i(x_k) + \nu L$ an upper confidence bound on $f_\nu^c(x_k)$ and by $\hat{\alpha}_k := |\hat{F}_\nu^c(x_k)|$ a lower confidence bound on both $|f^c(x_k)|$ and $|f(x_k)|$. More precisely, $\mathbb{P}\left\{\hat{\alpha}_k \le \min\{|f_\nu^c(x_k)|, |f^c(x_k)|\}\right\} \ge 1-\delta$. The proofs of the above statements are shown in Appendix C. Then, we propose to estimate $\nabla B_{\eta,\nu}(x_k)$ by

$$g_k := G^0(x_k,\nu) + \eta \frac{G^c(x_k,\nu)}{\hat{\alpha}_k}. \tag{5}$$

Later in Fact 2, we bound the deviation $\|g_k - \nabla B_{\eta,\nu}(x_k)\|$ with high probability. Next, to define our algorithm we need to make a second assumption.

Assumption 2 *Let $D' \subseteq D$ be the subset defined by $D' = \left\{x \in \mathbb{R}^d : f_\nu^c(x) + \eta \le 0\right\}$. There exists $l > 0$ such that the norm of the gradient $\nabla f_\nu^c(x)$ is lower bounded on $D \setminus D'$ by l, i.e, $0 < l \le \|\nabla f_\nu^c(x)\| \le L$.*

Assumption 2 is needed to demonstrate that close to the boundary of the constraint set the term in the barrier gradient related to constraints becomes large enough to push the step direction away from the boundary back to the feasible set. A slightly modified Margensian Fromowitz Constraint Qualification (MFCQ) that holds for all points leads to the satisfaction of Assumption 2. We show this in Appendix B.

The proposed stochastic zero-th order algorithm is defined in Algorithm 1 below:

Algorithm 1 Stochastic Zero-th Order Logarithmic Barrier Method **(ZeLoBa)**

Input: $x_0 \in D$, number of iterations K, $\eta > 0$, $L, l > 0$, $C = \frac{7l}{54L}$, $\nu = \frac{C\eta}{L}$, $\{n_k\}_{k=1}^K$ defined in Lemma 1
while $k \le K$ **do**
  Sample n_k vectors $s_{kj}, j = 1,\dots,n_k$ independently from the uniform distribution on $\mathbb{S}^d$;
  Take n_k noisy measurements of each function $f^i(x),\ i = 0,\dots,m$ at points $F^i(x_k,\xi_{kj}^{i-}), F^i(x_k+\nu s_{kj},\xi_{kj}^{i+})$;
  Compute an estimator g_k of $\nabla B_{\eta,\nu}(x_k)$ using (5);
  Compute $\gamma_k = \frac{1}{\|g_k\|}\min\left\{\frac{\hat{\alpha}_k}{2Lk^{2/5}}, \frac{1}{k^{3/5}}\right\}$;
  $x_{k+1} = x_k - \gamma_k g_k$, $\lambda_{k+1} = \frac{\eta}{\hat{\alpha}_{k+1}}$;
end
Sample R from a discrete random distribution $\mathbb{P}\{R = k\} = \frac{\gamma_k\|g_k\|}{\sum_{k=1}^K \gamma_k\|g_k\|}$
Output: x_R

In the above, l is the constant defined in Assumption 2. Our algorithm is defined for fixed η. In practice interior point methods often use decreasing η. Our algorithm can be used for inner iterations of the classical log barrier method with decreasing η.

4. Safety and convergence analysis

From the algorithm we require the safety of the iterates, $f^c(x_k) \le 0$, and the safety of the measurements, $f^c(x_k + \nu s_{kj}) \le 0$, with high probability. Also, we require convergence to a stationary point of the smoothed function in expectation. Here, we show that these properties hold for ZeLoBa algorithm.

4.1. Safety.

Given the required accuracy η, the smoothing parameter ν (which is also the sampling radius) has to be conservative enough to guarantee constraint satisfaction at any measured point $x_k + \nu s_{kj}$ of ZeLoBa algorithm. Thus, we need to show that the iterates x_k always keep a sufficient distance $\Lambda > 0$ from the boundary, namely $-f^c_\nu(x_k) \geq \hat{\alpha}_k \geq \Lambda$. To show the above, we first need to bound the deviation of g_k from $\nabla B_{\eta,\nu}(x_k)$. Define the deviation by $\zeta_k := g_k - \nabla B_{\eta,\nu}(x_k)$. The deviation ζ_k is dependent on deviations $\Delta^0_k := G^0(x_k, \nu) - \nabla f^0_\nu(x_k), \Delta^c_k := G^c(x_k, \nu) - \nabla f^c_\nu(x_k)$, thus, we bound these latter terms first. We denote $\Sigma := (d+1)\sqrt{\ln\frac{1}{\delta} + \ln(2K+1)}\left(\sqrt{2}\sigma + L\nu\right)$. From the sub-Gaussian property of the noise $\xi^{i\pm}_{kj}$ and L-Lipschitz continuity of $f^i(x), i = 0, \ldots, m$, we have:

Fact 1 *For deviations $\Delta^j_k = G^j(x_k, \nu) - \nabla f^j_\nu(x_k), j = \{0, c\}$, we have $\mathbb{E}\|\Delta^j_k\|^2 \leq \frac{d^2}{n_k}\left(L^2 + \frac{2\sigma^2}{\nu^2}\right)$. For all points x_k with $k \leq K$ we have $\mathbb{P}\left\{\forall k = 1, \ldots, K \ \|\Delta^j_k\| \leq \frac{\Sigma}{\nu\sqrt{n_k}}\right\} \geq 1 - \delta$.*

For the proof see Appendix D. Using this result, we can get the following bound on ζ_k:

Fact 2 *For deviation $\zeta_k = g_k - \nabla B_{\eta,\nu}(x_k)$, we have $\mathbb{E}\|\zeta_k\| \leq \frac{(d+1)(\sqrt{2}\sigma + L\nu)}{\nu\sqrt{n_k}}\left(1 + \frac{2\eta}{\hat{\alpha}_k}\right)$. For all $k \leq K$ we have $\mathbb{P}\left\{\forall k = 1, \ldots, K \ \|\zeta_k\| \leq \frac{\Sigma}{\nu\sqrt{n_k}}\left(1 + \frac{2\eta}{\hat{\alpha}_k}\right)\right\} \geq 1 - \delta$.*

For the proof see Appendix E. From the above facts, observe that if we keep the iterates x_k away from the boundary, $\hat{\alpha}_k \geq \Lambda > 0$, we can bound the deviation ζ_k. Luckily, the Log Barrier gradient approach with sufficiently large number of measurements in ZeLoBa ensures this property, as shown in the following lemma.

Lemma 1 *Under Assumption 2, if the initial point satisfies $-f^c_\nu(x_0) \geq 2C\eta$ with $C = \frac{7l}{54L}$, then for all iterates x_k of ZeLoBa algorithm with $n_k \geq \frac{4\Sigma^2(C+1)^2}{\nu^2 C^2 L^2}$ we have $\mathbb{P}\{\hat{\alpha}_k \geq C\eta \ \forall k \leq K\} \geq 1 - \delta$.*

Proof sketch: The idea is to show that the satisfaction of $\hat{\alpha}_k \geq C\eta$ and $-f^c_\nu(x_k) \geq 2C\eta$ for iteration k implies the same bounds for the next iteration $k+1$ with high probability. To prove this, we divide the condition $-f^c_\nu(x_k) \geq 2C\eta$ into two following cases. **Case 1.** $-f^c_\nu(x_k) \geq 4C\eta$, i.e., x_k is far from the boundary of the constraint set. Then, in the next iteration $-f^c_\nu(x)$ cannot decrease more than twice due to the choice of the step size and L-Lipschitz continuity of $-f^c_\nu(x)$. Thus, for x_{k+1} the bound $-f^c_\nu(x_{k+1}) \geq 2C\eta$ holds. **Case 2.** $-f^c_\nu(x_k) \leq 4C\eta$, i.e., x_k is close enough to the boundary. In this case, we show that $-g_k$ pushes x_{k+1} away from the boundary. That is, $-g_k$ is the descent direction for f^c_ν: $\langle g_k, \nabla f^c_\nu(x_k)\rangle \geq 0$. This is because g_k defined in (5) can be expressed as a sum of $\frac{\eta}{\hat{\alpha}_k}\nabla f^c_\nu(x_k)$ and $\nabla f^0_\nu(x_k) + \zeta_k$, and the first term will be dominating. Indeed, close to the boundary the factor $\frac{\eta}{\hat{\alpha}_k}$ is large and $\|\nabla f^c_\nu(x_k)\|$ is lower bounded $\|\nabla f^c_\nu(x_k)\| \geq l > 0$ due to Assumption 2. Moreover, the step size γ_k is sufficiently small to guarantee that $f^c_\nu(x_{k+1})$ will not increase compared to $f^c_\nu(x_k)$ due to the M_ν-smoothness. Consequently, $-f^c_\nu(x_{k+1}) \geq 2C\eta$ holds for both cases. This implies $\hat{\alpha}_{k+1} \geq C\eta$ with high probability. Everything above holds conditioned on the previous iteration k. Carefully combining the conditional probabilities along $k = 1, \ldots, K$, we get the result of the lemma. The full proof can be found in Appendix F.

The above lemma implies that the sampling radius $\nu = \frac{C\eta}{L} \leq \frac{\hat{\alpha}_k}{L}$ is safe. Hence, our algorithm is safe:

Proposition 2 *Let Assumptions 1,2 hold and $n_k \geq \frac{4\Sigma^2(C+1)^2}{\nu^2 C^2 L^2}$. Then all iterations x_k and measurement points $x_k + \nu s_{kj}$ generated by ZeLoBa algorithm are safe, namely, $\mathbb{P}\{f^c(x_k) \leq 0 \ \forall k \leq K\} \geq 1 - \delta$ and $\mathbb{P}\{f^c(x_k + \nu s_{k,j}) \leq 0 \ \forall k \leq K\} \geq 1 - \delta$.*

For the proof see Appendix G.

4.2. Convergence.

Theorem 3 *Under Assumptions 1,2, for* $n_k \geq \frac{4\Sigma^2(C+1)^2}{\nu^2 C^2 L^2}$ *and* $K \geq \frac{1}{\eta^5}$ *iterations of ZeLoBa algorithm we have* $\mathbb{E}\|\nabla B_{\eta,\nu}(x_R)\| \leq \eta(C_1 + C_2 \ln K)$, *with* $C_1 = \frac{2L(B_{\eta,\nu}(x_0) - \min_x B_{\eta,\nu}(x))}{C} + \frac{L}{C}\left(1 + \frac{1}{C}\right)$ *and* $C_2 = \frac{1}{4C}\left(1 + \frac{1}{C}\right) + \frac{L}{2C}$. *This implies that for the pair* (x_R, λ_R) *in expectation* η*-approximate KKT condition holds:*

$$\mathbb{P}\{\lambda_R, -f_\nu^c(x_R) \geq 0\} \geq 1 - \delta, \tag{η-KKT.1}$$

$$\mathbb{P}\{\lambda_R(-f_\nu^c(x_R)) \leq 3\eta\} \geq 1 - \delta, \tag{η-KKT.2}$$

$$\mathbb{E}\|\nabla_x L(x_R, \lambda_R)\|_2 \leq (C_1 + C_2 \ln K)\eta. \tag{η-KKT.3}$$

The total number of measurements required is $N_K = n_k \cdot K = O(\frac{d^2}{\eta^9})$.

Proof sketch: The proof is based on standard non-convex analysis techniques. The log barrier $B_{\eta,\nu}(x)$ is only locally smooth with smoothness parameter $M_{2,\nu}(x_k) \leq M_\nu\left(1 + \frac{2\eta}{\hat{\alpha}_k}\right) + \frac{4L^2\eta}{\hat{\alpha}_k^2}$ for all the points within the ball with radius γ_k around x_k. Using the local smoothness, we bound the improvement in barrier value $B_{\eta,\nu}(x_{k+1}) - B_{\eta,\nu}(x_k)$ for each iteration k. Summing this together for all $k \leq K$ provides the bound on $\sum_{k=1}^K \frac{\gamma_k}{\|g_k\|}\|\nabla B_{\eta,\nu}(x_k)\|$. This expression represents scaled $\mathbb{E}_R\|\nabla B_{\eta,\nu}(x_R)\|$ for R defined at Step 8 of ZeLoBa algorithm. That is, we get $\mathbb{E}\|\nabla B_{\eta,\nu}(x_R)\| \leq (C_1 + C_2 \ln K)\eta$ for $K \geq \frac{1}{\eta^5}$. By construction, $\nabla B_{\eta,\nu}(x_R)$ equals to $\nabla_x L(x_R, \lambda_R)$ for the smoothed problem, that implies (η-KKT.3). (η-KKT.1) follows from Proposition 2. (η-KKT.2) follows from $\lambda_R = \frac{\eta}{\hat{\alpha}_R}$ and Lemma 1. The full proof is in Appendix H.

Remark: The obtained bound on the number of measurements, $O(\frac{d^2}{\eta^9})$, is $\frac{1}{\eta^2}$ times worse compared to Usmanova et al. (2019). This comes as a price for non-smoothness. This difference agrees with the difference $\frac{1}{\eta^2}$ in upper bounds in other works on zero-th order optimization such as Duchi et al. (2015).

Corollary 4 *If the initial objective and constraints are differentiable, then the result obtained in Theorem 3 entails satisfaction of the approximate KKT condition for the initial problem (1).*

Proof We define $\hat{\lambda}_R \in \mathbb{R}^m$, where $\hat{\lambda}_R^i = \begin{cases} 0 & , i \notin \arg\max_i \hat{F}^i(x_R), \\ \frac{\eta}{-\hat{F}_\nu^c(x_R)} & , i \in \arg\max_i \hat{F}^i(x_R) \end{cases}$. We can easily see that condition η-KKT.1 holds with high probability by construction: $-f_\nu^c(x_R) \geq \hat{\alpha}_R \geq 0$. Condition η-KKT.2 holds for all $i \notin \arg\max_i F^i(x_R)$ since $\hat{\lambda}_R^i$ is just equal to 0. For $i \in \arg\max_i \hat{F}^i(x_R)$, we have $\frac{\eta}{-\hat{F}^i(x_R)}(-f^i(x_R)) \leq \eta + \eta\frac{\hat{F}^i(x_R) - f^i(x_R)}{-\hat{F}^i(x_R)} \leq \eta + \eta\frac{\sigma\sqrt{\ln 1/\delta}/\sqrt{n_k}}{\hat{\alpha}_R} \leq 3\eta$. Finally, we can verify that condition η-KKT.3 holds as follows, using $\|\nabla f_\nu^i(x) - \nabla f^i(x)\| \leq \nu L d$ (Nesterov and Spokoiny, 2017):

$$\mathbb{E}\|L(x_R, \hat{\lambda}_R)\| = \mathbb{E}\|\nabla f^0(x_R) + \sum_{i=1}^m \hat{\lambda}^i \nabla f^i(x_R)\| = \mathbb{E}\left\|\nabla f^0(x_R) + \frac{\eta \nabla f^i(x_R)}{-\hat{F}_\nu^c(x_R)}\right\|$$

$$\leq \mathbb{E}\left\|\nabla f_\nu^0(x_R) + \frac{\eta \nabla f_\nu^i(x_R)}{-\hat{F}_\nu^c(x_R)}\right\| + \|\nabla f^0(x_R) - \nabla f_\nu^0(x_R)\| + \eta\left\|\frac{\nabla f^i(x_R)}{-\hat{F}_\nu^c(x_R)} - \frac{\nabla f_\nu^i(x_R)}{-\hat{F}_\nu^c(x_R)}\right\|$$

$$\leq \|\nabla B_{\eta,\nu}(x_R)\| + \eta\left(1 + \frac{\nu L d}{\hat{\alpha}_R}\right) \leq \|\nabla B_{\eta,\nu}(x_R)\| + \eta(d+1) \leq \eta(C_1 \ln K + C_2 + d + 1).$$

■

In case of non-smooth objective and constraints, we do not know yet how to relate directly the result of Theorem 3 to the original problem (1). This is a direction of the future research.

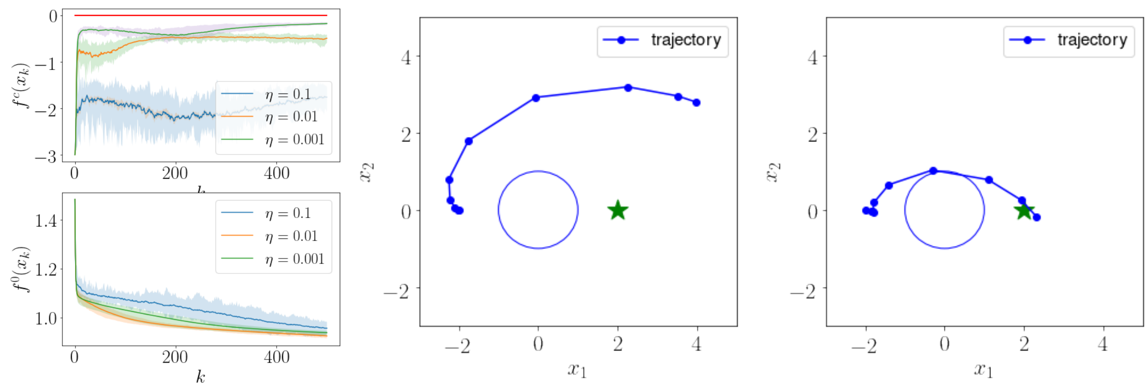

Figure 1: a) *Left.* Maximum constraint value (*top*) and the objective value (*bottom*) for 20 experiments; b) *Middle.* Control trajectory with U_0; c) *Right.* Final control trajectory obtained by ZeLoBa with $\eta = 0.001$.

5. Experiments

We consider the application to safe iterative controller design. Consider the basic unicycle dynamics $\dot{x} = v\cos\theta, \dot{y} = v\sin\theta, \dot{\theta} = \omega$. Here the states $q = [x, y, \theta]$ describe the spatial coordinates x, y and the direction angle θ. The control inputs $u = [v, \omega]$ describe the speed and the angular velocity. Since the simple Euler discretization is valid only when the sampling period dt is sufficiently small, we use a discretized model of the unicycle based on direct integration of the dynamics (Nino-Suarez et al., 2006; Adinandra et al., 2012):

$$q_{t+1} = \begin{bmatrix} x_{t+1} \\ y_{t+1} \\ \theta_{t+1} \end{bmatrix} = q_t + \begin{bmatrix} 2v_t + \gamma(\omega_t)\cos(\theta_t + \frac{dt}{2}\omega_t) \\ 2v_t + \gamma(\omega_t)\sin(\theta_t + \frac{dt}{2}\omega_t) \\ dt\omega_t \end{bmatrix}, \ \gamma(\omega_t) = \begin{cases} \sin(\frac{dt}{2}\omega_t), & \omega_t \neq 0 \\ \frac{dt}{2}, & \omega_t = 0 \end{cases}.$$

We choose a memoryless linear feedback law $u_{t+1} = Uq_t$, where $U \in \mathbb{R}^{3\times 2}$ is the optimizing parameter. The state sequence determined by U is denoted by $q_t(U), t = 1, \dots, T$ where T is the planning horizon. The goal is to lead the vehicle from a starting point q_A to a goal destination q_B while avoiding collision with high-probability. The cost function is defined as $\sum_{t=1}^{T} \|q_t(U) - q_B\|^2$. The constraints are formulated such that the trajectory does not collide with the the ball shaped obstacle placed at $(x_C, y_C)^T$ with radius 1. The resulting constrained optimization problem is as follows:

$$\min_{U\in\mathbb{R}^{3\times 2}} \frac{1}{T}\sum_{t=1}^{T} \|q_t(U) - q_B\|^2$$
$$\text{subject to } \ 1 - \left\|(x_t(U), y_t(U))^T - (x_C, y_C)^T\right\|^2 \leq 0, \ t = 1, \dots, T.$$

In the zero-th order oracle approach, we assume no knowledge of the dynamics, the constraints or the cost functions. We only assume noisy measurements of the cost function and the constraints. Thus, we address this problem using the ZeLoBa algorithm. We set the parameters of the algorithm to $\nu_k = \min\{\frac{\eta}{L}, \frac{\hat{\alpha}_k}{L}\}$ for safety, $L = 40$ set by trial, $n_k = 7$, $K = 500$, and initialize the algorithm with a safe control policy. The algorithm iteratively improves the controller while avoiding the constraints. The total number of measurements is $N_K = 3500$. In Figure 1 a) below we demonstrate the achieved results of 20 trials of the stochastic ZeLoBa algorithm with the fixed initialization. In none of the trials the constraints were violated. In Figure 1 b) we show the trajectory generated by U_0 controller. In Figure 1 c) we demonstrate an example of the trajectory generated by the final controller obtained during one of the trials of stochastic ZeLoBa algorithm.

References

Yasin Abbasi-Yadkori, Nevena Lazic, and Csaba Szepesvari. Model-free linear quadratic control via reduction to expert prediction. In *The 22nd International Conference on Artificial Intelligence and Statistics*, pages 3108–3117, 2019.

Sisdarmanto Adinandra, Erik Schreurs, and Henk Nijmeijer. A practical model predictive control for a group of unicycle mobile robots. *IFAC Proceedings Volumes*, 45(17):472–477, 2012.

Aaron D Ames. Human-inspired control of bipedal walking robots. *IEEE Transactions on Automatic Control*, 59(5):1115–1130, 2014.

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.

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*, 2016.

Sarah Dean, Stephen Tu, Nikolai Matni, and Benjamin Recht. Safely learning to control the constrained linear quadratic regulator. In *2019 American Control Conference (ACC)*, pages 5582–5588. IEEE, 2019.

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.

Maryam Fazel, Rong Ge, Sham Kakade, and Mehran Mesbahi. Global convergence of policy gradient methods for the linear quadratic regulator. In *International Conference on Machine Learning*, pages 1466–1475, 2018.

Jaime F Fisac, Anayo K Akametalu, Melanie N Zeilinger, Shahab Kaynama, Jeremy Gillula, and Claire J Tomlin. A general safety framework for learning-based control in uncertain robotic systems. *IEEE Transactions on Automatic Control*, 2018.

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.

Saeed Ghadimi and Guanghui Lan. Stochastic first-and zeroth-order methods for nonconvex stochastic programming. *SIAM Journal on Optimization*, 23(4):2341–2368, 2013.

Jeremy H Gillula and Claire J Tomlin. Guaranteed safe online learning via reachability: tracking a ground target using a quadrotor. In *2012 IEEE International Conference on Robotics and Automation*, pages 2723–2730. IEEE, 2012.

Elad Hazan, Kfir Yehuda Levy, and Shai Shalev-Shwartz. On graduated optimization for stochastic non-convex problems. In *International conference on machine learning*, pages 1833–1841, 2016.

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.

Jin-Bao Jian, Chun-Ming Tang, Qing-Jie Hu, and Hai-Yan Zheng. A feasible descent sqp algorithm for general constrained optimization without strict complementarity. *Journal of Computational and Applied Mathematics*, 180(2):391–412, 2005.

Guanghui Lan. The complexity of large-scale convex programming under a linear optimization oracle. *arXiv preprint arXiv:1309.5550*, 2013.

Zhijun Luo, Zhibin Zhu, and Guohua Chen. Modifying feasible sqp method for inequality constrained optimization. In *International Conference on Information Computing and Applications*, pages 323–330. Springer, 2012.

Yurii Nesterov and Vladimir Spokoiny. Random gradient-free minimization of convex functions. *Foundations of Computational Mathematics*, 17(2):527–566, 2017.

PA Nino-Suarez, E Aranda-Bricaire, and M Velasco-Villa. Discrete-time sliding mode path-tracking control for a wheeled mobile robot. In *Proceedings of the 45th IEEE Conference on Decision and Control*, pages 3052–3057. IEEE, 2006.

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, 2015.

Chun-ming Tang, Shuai Liu, Jin-bao Jian, and Jian-ling Li. A feasible sqp-gs algorithm for nonconvex, nonsmooth constrained optimization. *Numerical Algorithms*, 65(1):1–22, 2014.

Ilnura Usmanova, Andreas Krause, and Maryam Kamgarpour. Log barriers for safe non-convex black-box optimization. *arXiv preprint https://arxiv.org/abs/1912.09478*, 2019.

Guus Zoutendijk. *Methods of feasible directions: a study in linear and non-linear programming*. Elsevier, 1960.