A Stochastic Derivative Free Optimization Method with Momentum

Eduard Gorbunov
MIPT, Russia
eduard.gorbunov@phystech.edu

Adel Bibi
KAUST, Saudi Arabia
adel.bibi@kaust.edu.sa

Ozan Sener
Intel Labs
ozan.sener@intel.com

El Houcine Bergou
KAUST, Saudi Arabia and MaIAGE, INRA, France
elhoucine.bergou@inra.fr

Peter Richtárik
KAUST, Saudi Arabia and MIPT, Russia
peter.richtarik@kaust.edu.sa

Abstract

We consider the problem of unconstrained minimization of a smooth objective function in $\mathbb{R}^d$ in setting where only function evaluations are possible. We propose and analyze stochastic zeroth-order method with heavy ball momentum. In particular, we propose, `SMTP`, a momentum version of the stochastic three-point method (`STP`) [1]. We show new complexity results for non-convex, convex and strongly convex functions. We test our method on a collection of learning continuous control tasks on several MuJoCo [2] environments with varying difficulty and compare against `STP`, other state-of-the-art derivative-free optimization algorithms and against policy gradient methods. `SMTP` significantly outperforms `STP` and all other methods that we considered in our numerical experiments. Our second contribution is `SMTP` with importance sampling, dubbed `SMTP_IS`. We provide convergence analysis of this method for non-convex, convex and strongly convex objectives.

1 Introduction

In this paper, we consider the following minimization problem

$$\min_{x\in\mathbb{R}^d} f(x), \tag{1}$$

where $f : \mathbb{R}^d \to \mathbb{R}$ is "smooth" but not necessarily a convex function in a Derivative-Free Optimization (DFO) setting where only function evaluations are possible. The function f is bounded from below by $f(x^*)$ where x^* is a minimizer. Lastly and throughout the paper, we assume that f is L-smooth.

Assumption 1.1. *(L-smoothness) We say that f is L-smooth if $\|\nabla f(x) - \nabla f(y)\|_2 \leq L\|x - y\|_2$ for all $x, y \in \mathbb{R}^d$.*

DFO. In DFO setting [3, 4], the derivatives of the objective function f are not accessible. That is they are either impractical to evaluate, noisy (function f is noisy) [5] or they are simply not available at all. In standard applications of DFO, evaluations of f are only accessible through simulations of black-box engine or software as in reinforcement learning and continuous control environments [2]. This setting of optimization problems appears also in applications from computational medicine [6] and fluid dynamics [7–9] to localization [10, 11] and continuous control [12, 13] to name a few.

33rd Conference on Neural Information Processing Systems (NeurIPS 2019), Vancouver, Canada.

The literature on DFO for solving (1) is long and rich. The first approaches were based on deterministic direct search (DDS) and they span half a century of work [14–16]. However, for DDS methods complexity bounds have only been established recently by the work of Vicente and coauthors [17, 18]. In particular, the work of Vicente [17] showed the first complexity results on non-convex f and the results were extended to better complexities when f is convex [18]. However, there have been several variants of DDS, including randomized approaches [19–24]. Only very recently, complexity bounds have also been derived for randomized methods [25–29]. For instance, the work of [25, 29] imposes a decrease condition on whether to accept or reject a step of a set of random directions. Moreover, [30] derived new complexity bounds when the random directions are normally distributed vectors for both smooth and non-smooth f. They proposed both accelerated and non-accelerated zero-order (ZO) methods. Accelerated derivative-free methods in the case of inexact oracle information was proposed in [31]. An extension of [30] for non-Euclidean proximal setup was proposed by Gorbunov et. al. [32] for the smooth stochastic convex optimization with inexact oracle.

More recently and closely related to our work, Bergou et. al. [1] proposed a new randomized direct search method called *Stochastic Three Points* (`STP`). At each iteration k `STP` generates a random search direction s_k according to a certain probability law and compares the objective function at three points: current iterate x_k, a point in the direction of s_k and a point in the direction of $-s_k$ with a certain step size α_k. The method then chooses the best of these three points as the new iterate:

$$x_{k+1} = \arg\min\{f(x_k), f(x_k + \alpha_k s_k), f(x_k - \alpha_k s_k)\}.$$

Momentum. Heavy ball momentum[1] is a special technique introduced by Polyak in 1964 [33] to get faster convergence to the optimum for the first-order methods. In the original paper, Polyak proved that his method converges *locally* with $O\left(\sqrt{L/\mu}\log 1/\varepsilon\right)$ rate for twice continuously differentiable μ-strongly convex and L-smooth functions. Despite the long history of this approach, there is still an open question whether heavy ball method converges to the optimum *globally* with accelerated rate when the objective function is twice continuous differentiable, L-smooth and μ-strongly convex. For this class of functions, only non-accelerated global convergence was proved [34] and for the special case of quadratic strongly convex and L-smooth functions Lessard et. al. [35] recently proved asymptotic accelerated global convergence. However, heavy ball method performs well in practice and, therefore, is widely used. One can find more detailed survey of the literature about heavy ball momentum in [36].

Importance Sampling. Importance sampling has been celebrated and extensively studied in stochastic gradient based methods [37] or in coordinate based methods [38]. Only very recently, [39] proposed, `STP_IS`, the first DFO algorithm with importance sampling. In particular, under coordinate-wise smooth function, they show that sampling coordinate directions, can be generalized to arbitrary directions, with probabilities proportional to the function coordinate smoothness constants, improves the leading constant by the same factor typically gained in gradient based methods.

Contributions. Our contributions can be summarized into three folds.

- **First ZO method with heavy ball momentum.** Motivated by practical effectiveness of first-order momentum heavy ball method, we introduce momentum into `STP` method and propose new DFO algorithm with heavy ball momentum (`SMTP`). We summarized the method in Algorithm 1, with theoretical guarantees for non-convex, convex and strongly convex functions under generic sampling directions $\mathcal{D}$. To the best of our knowledge it is the first analysis of derivative-free method with heavy ball momentum, i.e. we show that the same momentum trick that works for the first order method could be applied for zeroth-order methods as well.
- **First ZO method with both heavy ball momentum and importance sampling.** In order to get more gain from momentum in the case when the sampling directions are coordinate directions and the objective function is coordinate-wise L-smooth (see Assumption 3.1), we consider importance sampling to the above method. In fact, we propose the first zeroth-order momentum method with importance sampling (`SMTP_IS`) summarized in Algorithm 2 with theoretical guarantees for non-convex, convex and strongly convex functions.
- **Practicality.** We conduct extensive experiments on continuous control tasks from the MuJoCo suite [2] following recent success of DFO compared to model-free reinforcement

[1] We will refer to this as momentum.

Algorithm 1 SMTP: Stochastic Momentum Three Points

Require: learning rates $\{\gamma^k\}_{k\geq 0}$, starting point $x^0 \in \mathbb{R}^d$, $\mathcal{D}$ — distribution on $\mathbb{R}^d$, $0 \leq \beta < 1$ — momentum parameter

Set $v^{-1} = 0$ and $z^0 = x^0$
for $k = 0, 1, \ldots$ **do**
  Sample $s^k \sim \mathcal{D}$
  Let $v_+^k = \beta v^{k-1} + s^k$ and $v_-^k = \beta v^{k-1} - s^k$
  Let $x_+^{k+1} = x^k - \gamma^k v_+^k$ and $x_-^{k+1} = x^k - \gamma^k v_-^k$
  Let $z_+^{k+1} = x_+^{k+1} - \frac{\gamma^k \beta}{1-\beta} v_+^k$ and $z_-^{k+1} = x_-^{k+1} - \frac{\gamma^k \beta}{1-\beta} v_-^k$
  Set $z^{k+1} = \arg\min\left\{f(z^k), f(z_+^{k+1}), f(z_-^{k+1})\right\}$
  Set $x^{k+1} = \begin{cases} x_+^{k+1}, & \text{if } z^{k+1} = z_+^{k+1} \\ x_-^{k+1}, & \text{if } z^{k+1} = z_-^{k+1} \\ x^k, & \text{if } z^{k+1} = z^k \end{cases}$ and $v^{k+1} = \begin{cases} v_+^{k+1}, & \text{if } z^{k+1} = z_+^{k+1} \\ v_-^{k+1}, & \text{if } z^{k+1} = z_-^{k+1} \\ v^k, & \text{if } z^{k+1} = z^k \end{cases}$
end for

Assumptions on f	SMTP Compleixty	Theorem	Importance Sampling	SMTP_IS Complexity
None	$2r_0 L\gamma_{\mathcal{D}}/\mu_{\mathcal{D}}^2\varepsilon^2$	2.1	$p_i = L_i/\sum_{i=1}^d L_i$	$2r_0 d\sum_{i=1}^d L_i/\varepsilon^2$
Convex, $R_0 < \infty$	$\ln\left(2r_0/\varepsilon\right) L\gamma_{\mathcal{D}} R_0^2/\mu_{\mathcal{D}}^2\varepsilon$	2.2	$p_i = L_i/\sum_{i=1}^d L_i$	$\ln\left(2r_0/\varepsilon\right) R_0^2 d\sum_{i=1}^d L_i/\varepsilon$
μ-strongly convex	$\ln\left(2r_0/\varepsilon\right) L/\mu\mu_{\mathcal{D}}^2$	2.5	$p_i = L_i/\sum_{i=1}^d L_i$	$\ln\left(2r_0/\varepsilon\right) \sum_{i=1}^d L_i/\mu$

Table 1: Summary of the new derived complexity results of SMTP and SMTP_IS. The complexities for SMTP are under a generic sampling distribution $\mathcal{D}$ satisfying Assumption 2.1 while for SMTP_IS are under an arbitrary discrete sampling from a set of coordinate directions following [39] where we propose an importance sampling that improves the leading constant marked in red. Note that $r_0 = f(x_0) - f(x_*)$ and that all assumptions listed are in addition to Assumption 1.1. Complexity means number of iterations in order to guarantee $\mathbf{E}\|\nabla f(\overline{z}^K)\|_{\mathcal{D}} \leq \varepsilon$ for the non-convex case, $\mathbf{E}\left[f(z^K) - f(x^*)\right] \leq \varepsilon$ for convex and strongly convex cases. $R_0 < \infty$ is the radius in $\|\cdot\|_{\mathcal{D}}^*$-norm of a bounded level set where the exact definition is given in Assumption 2.2. We notice that for STP_IS $\|\cdot\|_{\mathcal{D}} = \|\cdot\|_1$ and $\|\cdot\|_{\mathcal{D}}^* = \|\cdot\|_\infty$ in non-convex and convex cases and $\|\cdot\|_{\mathcal{D}} = \|\cdot\|_2$ in the strongly convex case.

learning [12, 13]. We achieve with SMTP_IS the state-of-the-art results on across all tested environments on the continuous control outperforming DFO [12] and policy gradient methods [40, 41].

2 Stochastic Momentum Three Points (SMTP)

Our analysis of SMTP is based on the following key assumption.

Assumption 2.1. *The probability distribution $\mathcal{D}$ on $\mathbb{R}^d$ satisfies the following properties: (1) the quantity $\gamma_{\mathcal{D}} \overset{def}{=} \mathbf{E}_{s\sim\mathcal{D}}\|s\|_2^2$ is positive and finite and (2) here is a constant $\mu_{\mathcal{D}} > 0$ and norm $\|\cdot\|_{\mathcal{D}}$ on $\mathbb{R}^d$ such that $\mathbf{E}_{s\sim\mathcal{D}}|\langle g, s\rangle| \geq \mu_{\mathcal{D}}\|g\|_{\mathcal{D}}$ for all $g \in \mathbb{R}^d$.*

Some examples of distributions that meet above assumption are described in Lemma 3.4 from [1].

The intuition behind SMTP is very similar to STP. STP is an adaptation of the classical gradient descent method, whereas SMTP is an adaptation of Polyak's heavy ball method with *a slight modification*: following the virtual iterates analysis [42], we introduce new variables $z_\pm^k$ and z^k. This is the key ingredient that allows us to modify heavy ball method into DFO method based on STP paradigm.

By definition of z^{k+1}, we get that the sequence $\{f(z^k)\}_{k\geq 0}$ is monotone: $f(z^{k+1}) \leq f(z^k)$, $\forall k \geq 0$. Our complexity results are based on the key result presented below.

Lemma 2.1. *Assume that f is L-smooth and $\mathcal{D}$ satisfies Assumption 2.1. Then for the iterates of* SMTP *the following inequalities hold:* $f(z^{k+1}) \leq f(z^k) - \frac{\gamma^k}{1-\beta}|\langle\nabla f(z^k), s^k\rangle| + \frac{L(\gamma^k)^2}{2(1-\beta)^2}\|s^k\|_2^2$ *and* $\mathbf{E}_{s^k\sim\mathcal{D}}\left[f(z^{k+1})\right] \leq f(z^k) - \frac{\gamma^k\mu_{\mathcal{D}}}{1-\beta}\|\nabla f(z^k)\|_{\mathcal{D}} + \frac{L(\gamma^k)^2\gamma_{\mathcal{D}}}{2(1-\beta)^2}$.

2.1 Non-Convex Case

In this section, we show our complexity results for Algorithm 1 in the case when f is allowed to be non-convex. In particular, we show that SMTP in Algorithm 1 guarantees complexity bounds with the same order as classical bounds, i.e. $1/\sqrt{K}$ where K is the number of iterations, in the literature.

Theorem 2.1. *Let Assumptions 1.1 and 2.1 be satisfied. Let* SMTP *with* $\gamma^k \equiv \gamma > 0$ *produce points* $\{z^0, z^1, \dots, z^{K-1}\}$ *and* $\overline{z}^K$ *is chosen uniformly at random among them. Then* $\mathbf{E}[\|\nabla f(\overline{z}^K)\|_{\mathcal{D}}] \leq {(1-\beta)(f(x^0)-f(x^*))}/{K\gamma\mu_{\mathcal{D}}} + {L\gamma\gamma_{\mathcal{D}}}/{2\mu_{\mathcal{D}}(1-\beta)}$. *Moreover, if we choose* $\gamma = \gamma_0/\sqrt{K}$ *this inequality reduces to* $\mathbf{E}[\|\nabla f(\overline{z}^K)\|_{\mathcal{D}}] \leq \left({(1-\beta)(f(z^0)-f(x^*))}/{\gamma_0\mu_{\mathcal{D}}} + {L\gamma_0\gamma_{\mathcal{D}}}/{2\mu_{\mathcal{D}}(1-\beta)}\right)/\sqrt{K}$. *Then* $\gamma_0 = \sqrt{{2(1-\beta)^2(f(x^0)-f(x^*))}/{L\gamma_{\mathcal{D}}}}$ *minimizes the right-hand side of the previous inequality and for this choice we have* $\mathbf{E}[\|\nabla f(\overline{z}^K)\|_{\mathcal{D}}] \leq \sqrt{2(f(x^0)-f(x^*))L\gamma_{\mathcal{D}}}/\mu_{\mathcal{D}}\sqrt{K}$.

In other words, the above theorem states that SMTP converges no worse than STP for non-convex problems to the stationary point. However, in practice SMTP significantly outperforms STP. So, the relationship between SMTP and STP is correlated with the known on the literature relationship between Polyak's heavy ball method and gradient descent.

2.2 Convex Case

In this section, we present our complexity results for Algorithm 1 when f is convex. In particular, we show that this method guarantees complexity bounds with the same order as classical bounds, i.e. $1/K$, in the literature. We will need the following additional assumption in the sequel.

Assumption 2.2. *We assume that* f *is convex, has a minimizer* x^* *and has bounded level set at* x^0*:* $R_0 \overset{def}{=} \max\left\{\|x - x^*\|_{\mathcal{D}}^* \mid f(x) \leq f(x^0)\right\} < +\infty$, *where* $\|\xi\|_{\mathcal{D}}^* \overset{def}{=} \max\{\langle \xi, x\rangle \mid \|x\|_{\mathcal{D}} \leq 1\}$ *defines the dual norm to* $\|\cdot\|_{\mathcal{D}}$.

Theorem 2.2 (Constant stepsize)**.** *Let Assumptions 1.1, 2.1 and 2.2 be satisfied. If we set* $\gamma^k \equiv \gamma < {(1-\beta)R_0}/{\mu_{\mathcal{D}}}$, *then for the iterates of* SMTP *method the following inequality holds:* $\mathbf{E}[f(z^k) - f(x^*)] \leq \left(1 - {\gamma\mu_{\mathcal{D}}}/{(1-\beta)R_0}\right)^k \left(f(x^0) - f(x^*)\right) + {L\gamma\gamma_{\mathcal{D}}R_0}/{2(1-\beta)\mu_{\mathcal{D}}}$. *If we choose* $\gamma = {\varepsilon(1-\beta)\mu_{\mathcal{D}}}/{L\gamma_{\mathcal{D}}R_0}$ *for some* $0 < \varepsilon \leq {L\gamma_{\mathcal{D}}R_0^2}/{\mu_{\mathcal{D}}^2}$ *and run* SMTP *for* $k = K$ *iterations where* $K = \ln\left({2(f(x^0)-f(x^*))}/{\varepsilon}\right) {L\gamma_{\mathcal{D}}R_0^2}/{\mu_{\mathcal{D}}^2\varepsilon}$, *then we will get* $\mathbf{E}\left[f(z^K)\right] - f(x^*) \leq \varepsilon$.

In order to get rid of factor $\ln({2(f(x^0)-f(x^*))}/{\varepsilon})$ in the complexity we consider decreasing stepsizes.

Theorem 2.3 (Decreasing stepsizes)**.** *Let Assumptions 1.1, 2.1 and 2.2 be statisified. If* $\gamma^k = {2}/{\alpha k+\theta}$, *where* $\alpha = {\mu_{\mathcal{D}}}/{(1-\beta)R_0}$ *and* $\theta \geq {2}/{\alpha}$, *then for* SMTP *iterates the following inequality holds:* $\mathbf{E}[f(z^k)] - f(x^*) \leq {1}/{\eta k+1} \max\left\{f(x^0) - f(x^*), {2L\gamma_{\mathcal{D}}}/{\alpha\theta(1-\beta)^2}\right\}$, *where* $\eta \overset{def}{=} {\alpha}/{\theta}$. *If* $\gamma^k = {2\alpha}/{\alpha^2 k+2}$ *where* $\alpha = {\mu_{\mathcal{D}}}/{(1-\beta)R_0}$ *and run* SMTP *for* $k = K = \max\left\{(1-\beta)^2(f(x^0) - f(x^*)), L\gamma_{\mathcal{D}}\right\} {2R_0^2}/{\varepsilon\mu_{\mathcal{D}}^2} - {2(1-\beta)^2R_0^2}/{\mu_{\mathcal{D}}^2}$, $\varepsilon > 0$, *we get* $\mathbf{E}\left[f(z^K)\right] - f(x^*) \leq \varepsilon$.

When β is sufficiently close to 1, we will obtain from the theorem above that $K \approx {2R_0^2L\gamma_{\mathcal{D}}}/{\varepsilon\mu_{\mathcal{D}}^2}$.

2.3 Strongly Convex Case

In this section we present our complexity results for Algorithm 1 when f is μ-strongly convex.

Assumption 2.3. *We assume that* f *is* μ*-strongly convex with respect to the norm* $\|\cdot\|_{\mathcal{D}}$*:* $f(y) \geq f(x) + \langle \nabla f(x), y - x\rangle + \frac{\mu}{2}\|y - x\|_{\mathcal{D}}^2$ *for all* $x, y \in \mathbb{R}^d$.

Theorem 2.4 (Solution-dependent stepsizes)**.** *Let Assumptions 1.1, 2.1 and 2.3 be satisfied. If we set* $\gamma^k = \sqrt{2\mu(f(z^k) - f(x^*))}{(1-\beta)\theta_k\mu_{\mathcal{D}}}/{L}$ *for some* $\theta_k \in (0, 2)$ *such that* $\theta = \inf_{k\geq 0}\{2\theta_k - \gamma_{\mathcal{D}}\theta_k^2\} \in (0, {L}/{(\mu_{\mathcal{D}}^2\mu)})$, *then for the iterates of* SMTP, *the following inequality holds:* $\mathbf{E}\left[f(z^k)\right] - f(x^*) \leq \left(1 - {\theta\mu_{\mathcal{D}}^2\mu}/{L}\right)^k \left(f(x^0) - f(x^*)\right)$. *Then, if we run* SMTP *for* $k = K \ln\left({f(x^0)-f(x^*)}/{\varepsilon}\right) {\kappa}/{\theta\mu_{\mathcal{D}}^2}$, $\varepsilon > 0$, *where* $\kappa \overset{def}{=} {L}/{\mu}$ *is the condition number of the objective, we will get* $\mathbf{E}\left[f(z^K)\right] - f(x^*) \leq \varepsilon$.

Note that the previous result uses stepsizes that depends on the optimal solution $f(x^*)$ which is often not known in practice. The next theorem removes this drawback without spoiling the convergence rate. However, we need an additional assumption on the distribution $\mathcal{D}$ and one extra function evaluation.

Algorithm 2 `SMTP_IS`: Stochastic Momentum Three Points with Importance Sampling

Require: stepsize parameters $w_1, \dots, w_n > 0$, probabilities $p_1, \dots, p_n > 0$ summing to 1, starting point $x^0 \in \mathbb{R}^d$, $0 \leq \beta < 1$ — momentum parameter

Set $v^{-1} = 0$ and $z^0 = x^0$
for $k = 0, 1, \dots$ **do**
  Select $i_k = i$ with probability $p_i > 0$
  Choose stepsize γ_i^k proportional to $\frac{1}{w_{i_k}}$
  Let $v_+^k = \beta v^{k-1} + e_{i_k}$ and $v_-^k = \beta v^{k-1} - e_{i_k}$
  Let $x_+^{k+1} = x^k - \gamma_i^k v_+^k$ and $x_-^{k+1} = x^k - \gamma_i^k v_-^k$
  Let $z_+^{k+1} = x_+^{k+1} - \frac{\gamma_i^k \beta}{1-\beta} v_+^k$ and $z_-^{k+1} = x_-^{k+1} - \frac{\gamma_i^k \beta}{1-\beta} v_-^k$
  Set $z^{k+1} = \arg\min \left\{ f(z^k), f(z_+^{k+1}), f(z_-^{k+1}) \right\}$
  Set $x^{k+1} = \begin{cases} x_+^{k+1}, & \text{if } z^{k+1} = z_+^{k+1} \\ x_-^{k+1}, & \text{if } z^{k+1} = z_-^{k+1} \\ x^k, & \text{if } z^{k+1} = z^k \end{cases}$ and $v^{k+1} = \begin{cases} v_+^{k+1}, & \text{if } z^{k+1} = z_+^{k+1} \\ v_-^{k+1}, & \text{if } z^{k+1} = z_-^{k+1} \\ v^k, & \text{if } z^{k+1} = z^k \end{cases}$
end for

Assumption 2.4. *We assume that for all $s \sim \mathcal{D}$ we have $\|s\|_2 = 1$.*

Theorem 2.5 (Solution-free stepsizes)**.** *Let Assumptions 1.1, 2.1, 2.3 and 2.4 be satisfied. If we compute $f(z^k + ts^k)$, set $\gamma^k = {(1-\beta)|f(z^k+ts^k) - f(z^k)|}/{(Lt)}$ for $t > 0$ and assume that $\mathcal{D}$ is such that $\mu_{\mathcal{D}}^2 \leq {L}/{\mu}$, then for the iterates of* `SMTP` *the following inequality holds: $\mathbf{E}\left[f(z^k)\right] - f(x^*) \leq \left(1 - {\mu_{\mathcal{D}}^2 \mu}/{L}\right)^k \left(f(x^0) - f(x^*)\right) + {L^2 t^2}/{(8\mu_{\mathcal{D}}^2 \mu)}$. Moreover, for any $\varepsilon > 0$ if we set t such that $0 < t \leq \sqrt{4\varepsilon\mu_{\mathcal{D}}^2 \mu / L^2}$, and run* `SMTP` *for $k = K$ iterations where $K = \ln\left({2(f(x^0) - f(x^*))}/{\varepsilon}\right) {\kappa}/{\mu_{\mathcal{D}}^2}$, where $\kappa \overset{def}{=} {L}/{\mu}$ is the condition number of f, we will have $\mathbf{E}\left[f(z^K)\right] - f(x^*) \leq \varepsilon$.*

3 Stochastic Momentum Three Points with Importance Sampling (`SMTP_IS`)

In this section we consider another assumption, in a similar spirit to [39], on the objective.

Assumption 3.1 (Coordinate-wise L-smoothness)**.** *We assume that the objective f has coordinate-wise Lipschitz gradient, with Lipschitz constants $L_1, \dots, L_d > 0$, i.e. $f(x + he_i) \leq f(x) + \nabla_i f(x) h + \frac{L_i}{2} h^2$, for all $x \in \mathbb{R}^d, h \in \mathbb{R}$, where $\nabla_i f(x)$ is i-th partial derivative of f at the point x.*

For this kind of problems we modify `SMTP` and present `STMP_IS` method in Algorithm 2. Due to the lack of space, we omit theorems with the complexity results for `SMTP_IS` and state them in Table 1.

4 Experiments

Experimental Setup. We conduct extensive experiments on challenging non-convex problems on the continuous control task from the MuJoCO suit [2]. In particular, we address the problem of model-free control of a dynamical system. Policy gradient methods for model-free reinforcement learning algorithms provide an off-the-shelf model-free approach to learn how to control a dynamical system and are often benchmarked in a simulator. We compare our proposed momentum stochastic three points method `SMTP` and the momentum with importance sampling version `SMTP_IS` against state-of-art DFO based methods as `STP_IS` [39] and ARS [12]. Moreover, we also compare against classical policy gradient methods as TRPO [40] and NG [41]. We conduct experiments on environments with varying difficulty `Swimmer-v1`, `Hopper-v1`, `HalfCheetah-v1`, `Ant-v1`, and `Humanoid-v1`.

Note that due to the stochastic nature of problem where f is stochastic, we use the mean of the function values of $f(x^k), f(x_+^k)$ and $f(x_-^k)$, see Algorithm 1, over K observations. Similar to the work in [39], we use $K = 2$ for `Swimmer-v1`, $K = 4$ for both `Hopper-v1` and `HalfCheetah-v1`, $K = 40$ for `Ant-v1` and `Humanoid-v1`. Similar to [39], these values were chosen based on the validation performance over the grid that is $K \in \{1, 2, 4, 8, 16\}$ for the smaller dimensional problems `Swimmer-v1`, `Hopper-v1`, `HalfCheetah-v1` and $K \in \{20, 40, 80, 120\}$ for larger dimensional problems `Ant-v1`, and `Humanoid-v1`. As for the momentum term, for `SMTP` we set $\beta = 0.5$. For `SMTP_IS`, as the smoothness constants are not available for continuous control, we use the coordinate smoothness constants of a θ parameterized smooth function $\hat{f}_\theta$ (multi-layer perceptron) that estimates f. In particular, consider running any DFO for n steps; with the queried sampled $\{x_i, f(x_i)\}_{i=1}^n$, we

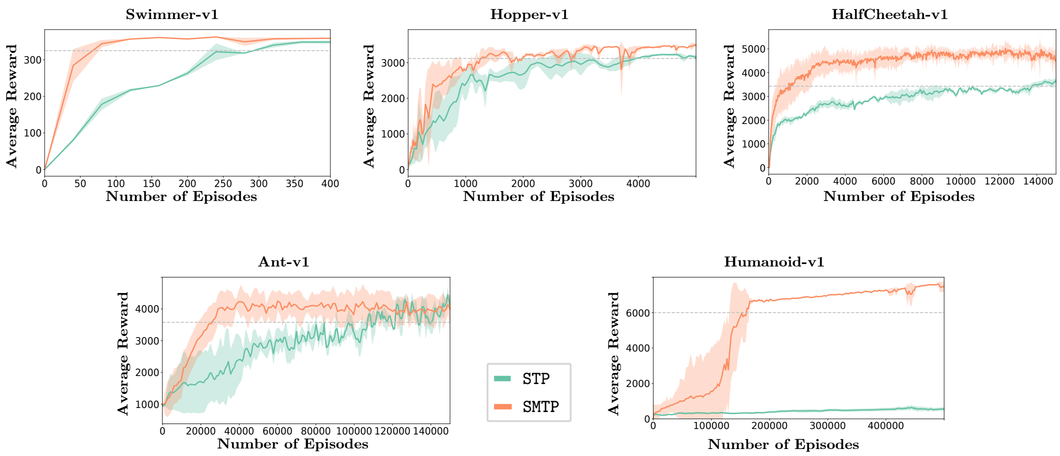

Figure 1: SMTP is far more superior to STP on all 5 different MuJoCo tasks particularly on the high dimensional Humanoid-v1 problem. The horizontal dashed lines are the thresholds used in Table 2 to demonstrate complexity of each method.

Table 2: For each MuJoCo task, we report the average number of episodes required to achieve a predefined reward threshold. Results for our method is averaged over five random seeds, the rest is copied from [12] (N/A means the method failed to reach the threshold. UNK means the results is unknown since they are not reported in the literature.)

	Threshold	STP	STP_{IS}	SMTP	$SMTP_{IS}$	ARS(V1-t)	ARS(V2-t)	NG-lin	TRPO-nn
Swimmer-v1	325	320	110	80	100	100	427	1450	N/A
Hopper-v1	3120	3970	2400	1264	1408	51840	1973	13920	10000
HalfCheetah-v1	3430	13760	4420	1872	1624	8106	1707	11250	4250
Ant-v1	3580	107220	43860	19890	14420	58133	20800	39240	73500
Humanoid-v1	6000	N/A	530200	161230	207160	N/A	142600	130000	UNK

estimate f by solving $\theta_{n+1} = \text{argmin}_\theta \sum_i (f(x_i) - \hat{f}(x_i;\theta))^2$. See [39] for further implementation details as we follow the same experimental procedure. In contrast to STP_IS, our method (SMTP) does not required sampling from directions in the canonical basis; hence, we use directions from standard Normal distribution in each iteration. For SMTP_IS, we follow a similar procedure as [39] and sample from columns of a random matrix B.

Similar to the standard practice, we perform all experiments with 5 different initialization and measure the average reward, in continuous control we are maximizing the reward function f, and best and worst run per iteration. We compare algorithms in terms of reward vs. sample complexity.

Comparison Against STP. Our method improves sample complexity of STP and STP_IS significantly. Especially for high dimensional problems like Ant-v1 and Humanoid-v1, sample efficiency of SMTP is at least as twice as the STP. Moreover, SMTP_IS helps in some experiments by improving over SMTP. However, this is not consistent in all environments. We believe this is largely due to the fact that SMTP_IS can only handle sampling from canonical basis similar to STP_IS.

Comparison Against State-of-The-Art. We compare our method with state-of-the-art DFO and policy gradient algorithms. For the environments, Swimmer-v1, Hopper-v1, HalfCheetah-v1 and Ant-v1, our method outperforms the state-of-the-art results. Whereas for Humanoid-v1, our methods results in a comparable sample complexity.

5 Conclusion

We have proposed, SMTP, the first heavy ball momentum DFO based algorithm with convergence rates for non-convex, convex and strongly convex functions under generic sampling direction. We specialize the sampling to the set of coordinate bases and further improve rates by proposing a momentum and importance sampling version SMPT_IS with new convergence rates for non-convex, convex and strongly convex functions too. We conduct large number of experiments on the task of controlling dynamical systems. We achieve the state-of-the-art performance compared to all DFO based and policy gradient based methods.

References

[1] E. H. Bergou, E. Gorbunov, and P. Richtárik, "Stochastic three points method for unconstrained smooth minimization," *arXiv preprint arXiv:1902.03591*, 2019.

[2] E. Todorov, T. Erez, and Y. Tassa, "Mujoco: A physics engine for model-based control," in *Intelligent Robots and Systems (IROS), 2012 IEEE/RSJ International Conference on*, pp. 5026–5033, IEEE, 2012.

[3] A. R. Conn, K. Scheinberg, and L. N. Vicente, *Introduction to Derivative-Free Optimization*. Philadelphia, PA, USA: SIAM, 2009.

[4] T. G. Kolda, R. M. Lewis, and V. J. Torczon, "Optimization by direct search: New perspectives on some classical and modern methods," *SIAM Review*, vol. 45, pp. 385–482, 2003.

[5] R. Chen, "Stochastic derivative-free optimization of noisy functions," *PhD thesis at Lehigh University.*, 2015.

[6] A. L. Marsden, J. A. Feinstein, and C. A. Taylor, "A computational framework for derivative-free optimization of cardiovascular geometries," *Computer Methods in Applied Mechanics and Engineering*, vol. 197, pp. 1890–1905, 2008.

[7] G. Allaire, *Shape Optimization by the Homogenization Method*. New York, USA: Springer, 2001.

[8] J. Haslinger and R. Mäckinen, *Introduction to Shape Optimization: Theory, Approximation, and Computation*. Philadelphia, PA, USA: SIAM, 2003.

[9] B. Mohammadi and O. Pironneau, *Applied Shape Optimization for Fluids*. Clarendon Press, Oxford, 2001.

[10] A. L. Marsden, M. Wang, J. E. Dennis, and P. Moin, "Optimal aeroacustic shape design using the surrogate management framework," *Optimization and Engineering*, vol. 5, pp. 235–262, 2004.

[11] A. L. Marsden, M. Wang, J. E. Dennis, and P. Moin, "Trailing-edge noise reduction using derivative-free optimization and large-eddy simulation," *Journal of Fluid Mechanics*, vol. 5, pp. 235–262, 2007.

[12] H. Mania, A. Guy, and B. Recht, "Simple random search provides a competitive approach to reinforcement learning," *arXiv preprint arXiv:1803.07055*, 2018.

[13] T. Salimans, J. Ho, X. Chen, S. Sidor, and I. Sutskever, "Evolution strategies as a scalable alternative to reinforcement learning," *arXiv preprint arXiv:1703.03864*, 2017.

[14] R. Hooke and T. Jeeves, "Direct search solution of numerical and statistical problems," *J. Assoc. Comput. Mach*, vol. 8, pp. 212–229, 1961.

[15] Y. W. Su, "Positive basis and a class of direct search techniques," *Scientia Sinica (in Chinese)*, vol. 9, no. S1, pp. 53–67, 1979.

[16] V. Torczon, "On the convergence of pattern search algorithms," *SIAM Journal on optimization*, vol. 7, no. 1, pp. 1–25, 1997.

[17] L. N. Vicente, "Worst case complexity of direct search," *EURO Journal on Computational Optimization*, vol. 1, no. 1-2, pp. 143–153, 2013.

[18] M. Dodangeh and L. N. Vicente, "Worst case complexity of direct search under convexity," *Mathematical Programming*, vol. 155, no. 1-2, pp. 307–332, 2016.

[19] J. Matyas, "Random optimization," *Automation and Remote Control*, vol. 26, pp. 246–253, 1965.

[20] V. G. Karmanov, "Convergence estimates for iterative minimization methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 14, pp. 1–13, 1974.

[21] V. G. Karmanov, "On convergence of a random search method in convex minimization problems," *Theory of Probability and its applications*, vol. 19, pp. 788–794, 1974.

[22] N. Baba, "Convergence of a random optimization method for constrained optimization problems," *Journal of Optimization Theory and Applications*, vol. 33, pp. 1–11, 1981.

[23] C. Dorea, "Expected number of steps of a random optimization method," *Journal of Optimization Theory and Applications*, vol. 39, pp. 165–171, 1983.

[24] M. Sarma, "On the convergence of the Baba and Dorea random optimization methods," *Journal of Optimization Theory and Applications*, vol. 66, pp. 337–343, 1990.

[25] M. A. Diniz-Ehrhardt, J. M. Martinez, and M. Raydan, "A derivative-free nonmonotone line-search technique for unconstrained optimization," *Journal of Optimization Theory and Applications*, vol. 219, pp. 383–397, 2008.

[26] S. U. Stich, C. L. Muller, and B. Gartner, "Optimization of convex functions with random pursuit," *arXiv preprint arXiv:1111.0194*, 2011.

[27] S. Ghadimi and G. Lan, "Stochastic first-and zeroth-order methods for nonconvex stochastic programming," *SIAM Journal on Optimization*, vol. 23, no. 4, pp. 2341–2368, 2013.

[28] S. Ghadimi, G. Lan, and H. Zhang, "Mini-batch stochastic approximation methods for nonconvex stochastic composite optimization," *Mathematical Programming*, vol. 155, no. 1-2, pp. 267–305, 2016.

[29] S. Gratton, C. W. Royer, L. N. Vicente, and Z. Zhang, "Direct search based on probabilistic descent," *SIAM Journal on Optimization*, vol. 25, no. 3, pp. 1515–1541, 2015.

[30] Y. Nesterov and V. Spokoiny, "Random gradient-free minimization of convex functions," *Foundations of Computational Mathematics*, vol. 17, pp. 527–566, 2017.

[31] P. Dvurechensky, A. Gasnikov, and A. Tiurin, "Randomized similar triangles method: A unifying framework for accelerated randomized optimization methods (coordinate descent, directional search, derivative-free method)," *arXiv preprint arXiv:1707.08486*, 2017.

[32] E. Gorbunov, P. Dvurechensky, and A. Gasnikov, "An accelerated method for derivative-free smooth stochastic convex optimization," *arXiv preprint arXiv:1802.09022*, 2018.

[33] B. T. Polyak, "Some methods of speeding up the convergence of iteration methods," *USSR Computational Mathematics and Mathematical Physics*, vol. 4, no. 5, pp. 1–17, 1964.

[34] E. Ghadimi, H. R. Feyzmahdavian, and M. Johansson, "Global convergence of the heavy-ball method for convex optimization," in *2015 European Control Conference (ECC)*, pp. 310–315, IEEE, 2015.

[35] L. Lessard, B. Recht, and A. Packard, "Analysis and design of optimization algorithms via integral quadratic constraints," *SIAM Journal on Optimization*, vol. 26, no. 1, pp. 57–95, 2016.

[36] N. Loizou and P. Richtárik, "Momentum and stochastic momentum for stochastic gradient, newton, proximal point and subspace descent methods," *arXiv preprint arXiv:1712.09677*, 2017.

[37] P. Zhao and T. Zhang, "Stochastic optimization with importance sampling for regularized loss minimization," in *international conference on machine learning*, pp. 1–9, 2015.

[38] P. Richtárik and M. Takáč, "On optimal probabilities in stochastic coordinate descent methods," *Optimization Letters*, vol. 10, no. 6, pp. 1233–1243, 2016.

[39] A. Bibi, E. H. Bergou, O. Sener, B. Ghanem, and P. Richtárik, "Stochastic derivative-free optimization method with importance sampling," *arXiv preprint arXiv:1902.01272*, 2019.

[40] J. Schulman, S. Levine, P. Abbeel, M. Jordan, and P. Moritz, "Trust region policy optimization," in *International Conference on Machine Learning*, pp. 1889–1897, 2015.

[41] A. Rajeswaran, K. Lowrey, E. V. Todorov, and S. M. Kakade, "Towards generalization and simplicity in continuous control," in *Advances in Neural Information Processing Systems*, pp. 6550–6561, 2017.

[42] T. Yang, Q. Lin, and Z. Li, "Unified convergence analysis of stochastic momentum methods for convex and non-convex optimization," *arXiv preprint arXiv:1604.03257*, 2016.