1. Introduction
In this paper, we consider the following minimization problem
\[
\min_{x \in \mathbb{R}^d} f(x),
\]
where \(f : \mathbb{R}^d \to \mathbb{R}\) is L-smooth: \(|\nabla f(x) - \nabla f(y)| \leq L|x - y|\).

2. Stochastic Three Points Method
Stochastic Three Points method [1] is a new method aimed to solve (1). The key properties of STP are its simplicity, generality, and practicality.

### Algorithm 1 (STP).
1. **Parameters:** some distribution \(D\) over \(\mathbb{R}^d\), stepizes \(\gamma^t\) \(t \geq 0\),
2. **Initialization:** Choose \(x_0 \in \mathbb{R}^n\)
3. **for** \(k = 0, 1, 2 \ldots \) **do**
   4. Draw a fresh sample \(s^k\) from \(D\)
   5. \(z^{k+1} = \arg\min\{f(x^k) + \gamma^t s^k, f(x^k - \gamma^t s^k)\}\)
   6. **end for**

3. Key Assumption

#### Assumption 1.
The probability distribution \(D\) on \(\mathbb{R}^d\) satisfies the following properties:
1. The quantity \(\gamma^t \mathbb{E}_{D}(\|s\|_2^2)\) is finite.
2. There is a constant \(\mu_D > 0\) for a norm \(\|\cdot\|_D\) in \(\mathbb{R}^d\) such that for all \(g \in \mathbb{R}^d\)
   \[\mathbb{E}_{D}(\|g(s)\|_D) \geq \mu_D \|g\|_2.\]

4. First Ingredient: Momentum Term
Below we introduce Polyak's heavy ball momentum using special technique inspired by virtual iterates analysis from [4].

### Algorithm 2 (SMTP).
1. **Parameters:** some distribution \(D\) over \(\mathbb{R}^d\), stepizes \(\gamma^t\) \(t \geq 0\),
2. **momentum parameter** \(\beta \in [0, 1)\)
3. **Initialization:** Choose \(x_0 \in \mathbb{R}^n\)
4. **for** \(k = 0, 1, 2 \ldots \) **do**
   5. Draw a fresh sample \(s^k\) from \(D\)
   6. \(v^k = \beta v^{k-1} + s^k\) and \(v^0 = s^0\)
   7. \(x^{k+1} = x^k - \gamma^k v^{k+1}\) and \(x^{k+1} = x^k - \gamma^k s^k\)
   8. \(z^{k+1} = \arg\min\{f(x^{k+1}) + \gamma^k v^{k+1}, f(x^{k+1}) + \gamma^k s^k\}\)
   9. \(v^{k+1} = \gamma v^k + \frac{1}{2} v^{k+1}\)
   10. **end for**

#### Key Lemma.
Assume that \(f\) is \(L\)-smooth and \(D\) satisfies Assumption 1. Then for the iterates of SMTP the following inequalities hold:
   \[f(z^{k+1}) \leq f(z^k) - \frac{\gamma^k}{1 - \beta^2} \|\nabla f(z^k)\|^2 + \frac{L \gamma^k}{2 (1 - \beta^2)} \|z^k\|^2,\]
   and
   \[\mathbb{E}_{D}(f(z^{k+1})) \leq f(z^k) - \frac{\gamma^k}{1 - \beta^2} \|\nabla f(z^k)\|^2 + \frac{L \gamma^k}{2 (1 - \beta^2)} \|z^k\|^2.\]

Using the lemma above one can get convergence guarantees for SMTP in the similar way as it was done in [1].