Theorem 1

For all $x \in Q$

1. Unbiased:

$$\mathbb{E}_e\{\nabla f_\gamma(x,e)\} = \nabla f(x), \gamma \to 0;$$

2.

$$\forall \gamma, \mathbb{E}_e\{\nabla f_\gamma(x,e)\} = \nabla f(x) + \mathbb{E}_e\{e\}(O(\gamma) + \frac{d\Delta}{\gamma})$$

3. Bounded variance:

$$\begin{aligned} &\mathbb{E}_e\{\|\nabla f_\gamma(x,e)\|_q^2\} = \kappa(p,d)(dM_2^2 + \frac{d^2\Delta^2}{\gamma^2}), \\ &1/p + 1/q = 1, \\ &\kappa(p,d) = O(\sqrt{\mathbb{E}_e\|e\|_q^4}) = \begin{cases} O(1), p = 2; \\ O((\ln d)/d), p = 1. \end{cases} \end{aligned}$$

Proof

First two statements

Due to the existence of first derivative:

$$\begin{aligned} &\vec{g} := \mathbb{E}_e\{\nabla f_\gamma(x,e)\} = \\ &d\mathbb{E}_e\{\frac{f(x) + \langle\nabla f(x), \gamma e\rangle + o(\|\gamma e\|_2) - (f(x) - \langle\nabla f(x), \gamma e\rangle + o(\|\gamma e\|_2))}{2\gamma}e\} = \\ &= d\mathbb{E}_e\{\frac{2\langle\nabla f(x), \gamma e\rangle + o(|\gamma|)}{2\gamma}e\} \underset{\gamma\to 0}{=} d\mathbb{E}_e\{\langle\nabla f(x), e\rangle e\} = \nabla f(x) \end{aligned}$$

Let $\delta(x) : |\delta(x)| \leq \Delta$ Oracle's noise, then:

$$\begin{aligned} &\vec{g} := \mathbb{E}_e\{\nabla f_\gamma(x,e)\} = \\ &d\mathbb{E}_e\{\frac{\delta(x+\gamma e) + \langle\nabla f(x), \gamma e\rangle - (\delta(x-\gamma e) - \langle\nabla f(x), \gamma e\rangle) + O(\|\gamma e\|_2^2)}{2\gamma}e\} \leq \\ &\leq d\mathbb{E}_e\{\frac{2\langle\nabla f(x), \gamma e\rangle + 2\Delta + O(\gamma^2)}{2\gamma}e\} = \nabla f(x) + \mathbb{E}_e\{e\}O(d\gamma + \frac{d\Delta}{\gamma}), \end{aligned}$$

More than that, let $\vec{r} = x_0 - x^*, R = \|\vec{r}\|_2$

$$\begin{aligned} &|\langle\vec{g}, \vec{r}\rangle| = |\langle\nabla f(x), \vec{r}\rangle + \langle\mathbb{E}_e\{e\}, \vec{r}\rangle(dO(\gamma) + \frac{d\Delta}{\gamma})| \leq |\langle\nabla f(x), \vec{r}\rangle + \mathbb{E}_e\{\langle e, \vec{r}\rangle\}O(d\gamma + \frac{d\Delta}{\gamma})| \leq \\ &\leq |\langle\nabla f(x), \vec{r}\rangle| + \frac{R}{\sqrt{d}}|O(\gamma + \frac{\Delta}{\gamma})| \end{aligned}$$

Then maximal residual (for $\gamma = \sqrt{d\Delta}$):

$$\varepsilon \approx \frac{R}{\sqrt{d}} \cdot \sqrt{d\Delta} = R\sqrt{\Delta}$$

The third statement

Due to Shamir, 2017 (Lemma 4 and 5) we obtain:

$$\begin{aligned}
&\mathbb{E}_e\{\|\nabla f_\gamma(x,e)\|_q^2\} = \kappa(p,d)(dM_2^2 + \frac{d^2\Delta^2}{\gamma^2}),\\
&1/p + 1/q = 1,\\
&\kappa(p,d) = O(\sqrt{\mathbb{E}_e\|e\|_q^4}) = \begin{cases} O(1), p = 2;\\ O((\ln d)/d), p = 1.\end{cases}
\end{aligned}$$

Conclusion

For reaching ε sub-optimality the noise should be not greater then:

a)

$$R\sqrt{\Delta} \le \varepsilon \implies \Delta \le \frac{\varepsilon^2}{R^2}$$

b)

$$\frac{d^2\Delta^2}{\gamma^2} \le dM^2 \implies \Delta \le \frac{\gamma M}{\sqrt{d}} = [\gamma = \sqrt{d\Delta}] = M$$

$$\Delta \le M^2$$

So $\Delta \le \min\{M^2, \frac{\varepsilon^2}{R^2}\} \approx \frac{\varepsilon^2}{R^2}$

Gradient Descent case

$$x_{k+1} = x_k - a_k \cdot \nabla_\gamma f(x_k, e_k)$$

$$\|x_{k+1} - x_*\|_2^2 \le \|x_k - x_*\|_2^2 - 2a\langle \nabla_\gamma f(x_k, e_k), x_k - x_*\rangle + a^2\|\nabla_\gamma f(x_k, e_k)\|_2^2$$

It is easy to see that x_k does not depends on random vector e_k, so if we take math expectation by e_k with “frozen” x_k:

$$2a\langle \nabla f(x_k) + \vec{\mathbb{1}}^\top e_k \cdot O(\gamma), x_k - x_*\rangle = \|x_k - x_*\|_2^2 - \mathbb{E}_{e_k}\|x_{k+1} - x_*\|_2^2 + a^2(dM_2^2)$$

Now taking $\mathbb{E}_{x_k}$ from both sides and taking $\gamma \to 0$:

$$\mathbb{E}_{x_k}\{2a(f(x_k) - f(x_*)\} \le \mathbb{E}_{x_k}\{\|x_k - x_*\|_2^2\} - \mathbb{E}_{x_k}\{\|x_{k+1} - x_*\|_2^2\} + a^2(dM_2^2)$$

Summing both sides and using Jensen's inequality we obtain:

$$2aN\mathbb{E}_{x_k}\{2a(f(x_k) - f(x_*)\} \le R_2^2 + a^2M_2^2dN$$

Optimal $a = \frac{R_2^2}{M\sqrt{dN}}$, so:

$$\mathbb{E}_{x_k}\{2a(f(x_k) - f(x_*)\} \le \frac{M_2R_2^2\sqrt{d}}{\sqrt{N}}$$

So for reaching ε-suboptimality we need:

$$N \approx \frac{M_2^2R_2^2d}{\varepsilon}$$