Theorem 1

For all $x \in Q$

1. Unbiased:

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

2.

$$orall \gamma, \mathbb{E}_e \{
abla f_\gamma(x,e) \} =
abla f(x) + \mathbb{E}_e \{ e \} (O(\gamma) + rac{d\Delta}{\gamma})$$

3. Bounded variance:

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

Proof

First two statements

Due to the existence of first derivative:

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

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

$$egin{aligned} ec{g} &:= \mathbb{E}_e\{
abla f_{\gamma}(x,e)\} = \ d\mathbb{E}_e\{rac{\delta(x+\gamma e) + \langle
abla f(x), \gamma e
angle - (\delta(x-\gamma e) - \langle
abla f(x), \gamma e
angle) + O(\|\gamma e\|_2^2)}{2\gamma}e\} &\leq \ d\mathbb{E}_e\{rac{2\langle
abla f(x), \gamma e
angle + 2\Delta + O(\gamma^2)}{2\gamma}e\} =
abla f(x) + \mathbb{E}_e\{e\}O(d\gamma + rac{d\Delta}{\gamma}), \end{aligned}$$

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

$$egin{aligned} |\langle ec{g},ec{r}
angle| &= |\langle
abla f(x),ec{r}
angle + \langle \mathbb{E}_e\{e\},ec{r}
angle (dO(\gamma) + rac{d\Delta}{\gamma})| \leq |\langle
abla f(x),ec{r}
angle + \mathbb{E}_e\{\langle e,ec{r}
angle\}O(d\gamma + rac{d\Delta}{\gamma})| \leq \\ &\leq |\langle
abla f(x),ec{r}
angle| + rac{R}{\sqrt{d}}|O(\gamma + rac{\Delta}{\gamma})| \end{aligned}$$

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

$$arepsilon pprox rac{R}{\sqrt{d}} \cdot \sqrt{d\Delta} = R \sqrt{\Delta}$$

The third statement

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

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

Conclusion

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

a)

$$R\sqrt{\Delta} \le arepsilon \implies \Delta \le rac{arepsilon^2}{R^2}$$

b)

$$rac{d^2\Delta^2}{\gamma^2} \leq dM^2 \implies \Delta \leq rac{\gamma M}{\sqrt{d}} = [\gamma = \sqrt{d\Delta}] = M$$
 $\Delta < M^2$

So $\Delta \leq \min\{M^2, rac{arepsilon^2}{R^2}\} pprox rac{arepsilon^2}{R^2}$

Gradient Descent case

$$egin{aligned} x_{k+1} &= x_k - a_k \cdot
abla_{\gamma} f(x_k, e_k) \ & \|x_{k+1} - x_*\|_2^2 \leq \|x_k - x_*\|_2^2 - 2a \langle
abla_{\gamma} f(x_k, e_k), x_k - x_*
angle + a^2 \|
abla_{\gamma} f(x_k, e_k)\|_2^2 \end{aligned}$$

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

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

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

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

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

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

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

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

So for reaching ε -suboptimality we need:

$$Npprox rac{M_2^2R_2^2d}{arepsilon}$$