

A PROOF OF THEOREM 4

Proof. To provide a rough upper bound for inequality (4), one can directly apply the Lipschitz Hessian condition. Particularly, let f_2 be the Taylor polynomial of f expanded at \mathbf{x} up to the quadratic terms, we have

$$|f(\mathbf{y}) - f_2(\mathbf{y})| \leq \frac{1}{6}\rho\|\mathbf{y} - \mathbf{x}\|_2^3. \quad (12)$$

Then by the definition of \hat{H} , we have

$$\begin{aligned} \left\| \mathbb{E}[\hat{H}] - \nabla \nabla f(\mathbf{x}) \right\|_F &\leq \mathbb{E}_{\mathbf{u} \sim S^{d-1}} \left[d(d+2) \left\| \mathbf{u}\mathbf{u}^\top - \frac{I_d}{d+2} \right\|_F \frac{1}{6}\rho r \right] \\ &= \frac{1}{6}d\sqrt{d(d+3)}\rho r. \end{aligned}$$

Although not necessary for our proofs, the above bound can be improved using techniques we applied in the proof of inequality (2).

For inequality (5), we apply the same bound stated in inequality (12) and obtain the follows.

$$\begin{aligned} \text{Tr}(\text{Cov}[\hat{H}]) &\leq \frac{d^2(d+2)^2}{n} \mathbb{E}_{\mathbf{u} \sim S^{d-1}} \left[\text{Tr} \left(\left(\mathbf{u}\mathbf{u}^\top - \frac{I_d}{d+2} \right)^2 \right) \frac{(\frac{1}{3}\rho r^3)^2 + 6}{(4r^2)^2} \right] \\ &= \frac{d^3(d+3)(\frac{1}{18}\rho r^6 + 3)}{8nr^4}. \end{aligned}$$

□

B LOWER BOUNDS: PROOF FOR THE GENERAL CASES

The generalization of the earlier 1D lower bound is obtained by constructing a set of hard-instance functions where the optimization problem over this subset consists of d binary hypothesis estimation problems each identical to a 1D construction. Formally, for any $\mathbf{s} = (s_1, s_2, \dots, s_d) \in \{1, 2\}^d$ and any input $\mathbf{x} = (x_{(1)}, x_{(2)}, \dots, x_{(d)})$, we let

$$f_{\mathbf{s}}(\mathbf{x}) = \sum_{j=1}^d f_{s_j}(x_{(j)}).$$

One can verify that $f_{\mathbf{s}} \in \mathcal{F}(\rho, M, R)$ for all \mathbf{s} for sufficiently large T .

Note that the simple regret for the above function class can be written as the sum of d individual terms $\sum_{j=1}^d (f_{s_j}(x_{(j)}) - \inf_x f_{s_j}(x))$. As proved earlier, the expectation of each term associated with any index j is at least $\Omega\left(\frac{\rho^{\frac{2}{3}}T^{-\frac{2}{3}}}{M}\right)$ even if all entries of \mathbf{s} except s_j is known. Therefore, the total expected regret is lower bounded by $\Omega\left(\frac{d\rho^{\frac{2}{3}}T^{-\frac{2}{3}}}{M}\right)$.