	Best known lower bound	**Best known upper bound**
Stationary, no-corruptions ($\Phi_T = 0, \Lambda_T = 0$)	$\Omega[\sqrt{T}(\sqrt{\text{Trace}(\Sigma)} + \sqrt{\nu_{max}(\Sigma)\ln(1/\delta)})]$ [2]	$\mathcal{O}[\sqrt{T}(\sqrt{\text{Trace}(\Sigma)} + \sqrt{\nu_{max}(\Sigma)\ln(T/\delta)})]$ [3]
Stationary, corruptions ($\Phi_T = 0, \Lambda_T \geq 0$)	$\Omega(\Lambda_T\mathcal{D})$ (Proof sketch below) No previous bound	Thm 5.1: $\widetilde{\mathcal{O}}(T^{\frac{5}{6}}\sigma^2 + T^{\frac{3}{4}}\mathcal{D}\Lambda_T)$. No previous bound
Non-stationary, no-corruptions ($\Phi_T \geq 0, \Lambda_T = 0$).	[1]: $\Omega(T^{2/3}\Phi_T^{1/3})$	Thm 5.1: $\widetilde{\mathcal{O}}(T^{\frac{2}{3}}\Phi_T + T^{\frac{5}{6}}\sigma^2)$ No previous bound. [1] gives in expectation bound only.
Non-stationary, corruptions ($\Phi_T \geq 0, \Lambda_T \geq 0$)	Prop 2.6: $\Omega\left(\Lambda_T\mathcal{D}\right)$. No previous bound	Thm 5.1: $\widetilde{\mathcal{O}}(T^{\frac{2}{3}}\Phi_T + T^{\frac{5}{6}}\sigma^2 + T^{\frac{3}{4}}\Lambda_T\mathcal{D})$. No previous bound

Table 1: Regret bounds for online estimation with heavy-tailed data and strongly convex loss. Ours is the first work to give bounds in many settings as seen above in red. The setting of $\Phi_T = 0, \Lambda_T = 0$ (line 1) is characterized (upto log factors) in [2] and [3]. For the setting of ($\Phi_T = 0, \Lambda_T \geq 0$) (line 2), we give the first dimension free upper and lower bounds. There is a gap however as our lower bound does not characterize dependence on the time horizon T. For the setting of ($\Phi_T \geq 0, \Lambda_T = 0$), [1] gives a lower bound while ours is the first high-probability regret bound for heavy tailed data. For the general case ($\Phi_T \geq 0, \Lambda_T \geq 0$) ours is the only known bounds. Except line 1, there is a gap between known lower and upper bounds as can be seen.

Proof Sketch of lower bound when $\Phi_T = 0, \Lambda_T \geq 0$: Similar to Prop 2.6, consider two scenarios for mean-estimation. In one scenario, the un-corrupted samples are all drawn from a Dirac mass at 0, but the first Λ_T samples are corrupted with all d coordinates set to $\mathcal{D}/\sqrt{d}$. In the other scenario, there are no corruptions and the un-corrupted samples are all from Dirac mass at location with all coordinates $\mathcal{D}/\sqrt{d}$. In both situations, the first Λ_T samples are identical. Thus, no estimator in the first Λ_T samples can distinguish between these two scenarios and will incur regret at-least $(\Lambda_T\mathcal{D})/4$.

References

[1] Omar Besbes, Yonatan Gur, and Assaf Zeevi. Non-stationary stochastic optimization. *Operations research*, 63(5):1227–1244, 2015.

[2] Olivier Catoni. Challenging the empirical mean and empirical variance: a deviation study. In *Annales de l'IHP Probabilités et statistiques*, 2012.

[3] Gábor Lugosi and Shahar Mendelson. Sub-gaussian estimators of the mean of a random vector. *The Annals of Statistics*, 2019.