Impacts of noise to convergence rates

Miaoyan Wang, Feb 5, 2021

Consider the signal plus noise model

$$\mathcal{Y} = \Theta + \mathcal{E},$$

where $\mathcal{E}$ consists of mean-zero, independent (but not necessarily identical) noise entries, and $\Theta \in \mathscr{P}_{\text{sgn}}(r)$ is an α-smooth tensor.

Our earlier result requires bounded observation $\|\mathcal{Y}\|_\infty \leq A$ for some constant $A > 0$. I believe this assumption can be relaxed to the following two assumptions.

Assumption 1 (sub-Gaussian noise)**.**

1. There exists a constant $\alpha > 0$, independent of tensor dimension, such that $\|\Theta\|_\infty \leq \alpha$.
2. The noise entries $\mathcal{E}(\omega)$ are independent and sub-Gaussian, i.e, $\mathbb{P}(|\mathcal{E}(\omega)| \geq L) \leq Ce^{-L^2/\sigma^2}$ for $L > 0$ and $\omega \in [d_1] \times \cdots \times [d_K]$. Here σ^2 is the sub-Gaussian parameter.

Early result states that, under the strong boundedness assumption, there exists a constant $C > 0$ (*Is C linear or quadratic in upper bound $\|\mathcal{Y}\|_\infty$?*) such that

$$\mathbb{P}\left[\text{MAE}(\hat{\Theta}, \Theta) \geq Ct_n\right] \leq \exp(-Cnt_n), \text{ where } t_n \asymp \left(\frac{dr}{n}\right)^{\frac{\alpha}{\alpha+2} \vee \frac{1}{2}}.$$

Conjecture 1. Under the Assumption 1, with very high probability (*polynomial tail or exponential tail?)*

$$\text{MAE}(\hat{\Theta}, \Theta) \leq t_n(\sigma \log d + \alpha).$$

Sketch of proofs. Let $L = L(d, \alpha, \sigma)$ denote the scaling factor we aim to find. Divide the sample space into two cases:

- Case 1: Every tensor entry $|\mathcal{Y}(\omega)| \leq L$.
- Case 2: At least one entry $\mathcal{Y}(\omega)$ such that $|\mathcal{Y}(\omega)| > L$.

We bound the tail event by

$$\begin{aligned}
&\mathbb{P}\left[\text{MAE}(\hat{\Theta}, \Theta) \geq Lt_n\right] \\
&= \mathbb{P}(\text{MAE}(\hat{\Theta}, \Theta) \geq Lt_n | \text{case 1})\mathbb{P}(\text{case 1}) + \mathbb{P}(\text{MAE}(\hat{\Theta}, \Theta) \geq Lt_n | \text{case 2})\mathbb{P}(\text{case 2}) \\
&\leq \exp(-Lnt_n) + 1 - \left[1 - C\exp(-(L-\alpha)^2/\sigma^2)\right]^{Kd} \\
&\lesssim \exp(-Lnt_n) + CK\exp(-(L-\alpha)^2/\sigma^2 + \log d).
\end{aligned}$$

There are two competing considerations when choosing L. We want a small L in the MAE bound whereas a large L in the probability bound. A somewhat balanced L is to set $(L-\alpha)^2/\sigma^2 \approx 4\log d$, i.e. $L \asymp 2\sigma\sqrt{\log d} + \alpha$. In this case, with probability at least $1 - cd^{-c}$, we have

$$\text{MAE}(\hat{\Theta}, \Theta) \leq t_n(\sigma \log d + \alpha).$$

Several details need to fill in:

- Any changes to the algorithm? Perhaps change the range $\pi \in [-1, 1]$ to $\pi \in [-\alpha - 2\sigma\log d, \alpha + 2\sigma\log d]$? Should we also change $H = H(\alpha, \sigma)$?

- Possible to obtain a sharper bound? Intuitively, the bound should be zero when $\sigma = 0$. The current bound has an additional α term. Go through each step of earlier proofs, including sign tensor estimation and aggregation.
- How does the sign estimation bound depend on (α, σ)?
- The $\text{MAE}(\hat{\Theta}, \Theta)$ increases with both noise *and* signal. Does it intuitively make sense?