Smoothness impact

Chanwoo Lee, July 20, 2021

1 When signal is zero

Let us consider $\mathcal{Y} = \Theta + \mathcal{E} \in \mathbb{R}^{d_1\times\cdots\times d_m}$ where $\mathcal{E}$ follows i.i.d. sub-Gaussian noise with $\sigma^2 = 1$ without loss of generality and $\Theta = \mathcal{X} + \mathcal{X}_\perp$ with $\text{rank}(\mathcal{X}) = (\sqrt{d_1},\ldots,\sqrt{d_m})$. For each $k = 1,...,n$, denote

$$X_k = \mathcal{M}_k(\mathcal{X}), \quad X_{k,\perp} = \mathcal{M}_k(\mathcal{X}_\perp), \quad E_k = \mathcal{M}_k(\mathcal{E}), \quad Y_k = \mathcal{M}_k(\mathcal{Y}),$$

and define $Z_k = X_{k,\perp} + E_k$. We consider the high-order spectral method, where we estimate the signal tensor Θ from

$$\begin{aligned}
\tilde{U}_k &= \text{SVD}_{r_k}(Y_k)\\
\hat{U}_k &= \text{SVD}_{r_k}\left(\mathcal{M}_k\left(\mathcal{Y}\times_1 \tilde{U}_1^T \times\cdots\times_{k-1}\tilde{U}_{k-1}^T\times_k \tilde{U}_{k+1}^T\times\cdots\times_m \tilde{U}_m^T\right)\right)\\
\hat{\Theta} &= \mathcal{Y}\times_1 (\hat{U}_1\hat{U}_1^T)\times\cdots\times_m(\hat{U}_m\hat{U}_m^t),
\end{aligned} \tag{1}$$

where r_k can be set arbitrary for all $k\in[m]$.

Theorem 1.1 (Estimation of high-order spectral algorithm)**.** Suppose that $\|X_\perp\|_F \leq \sqrt{\bar{d}}$. Then, with probability at least $1 - C\exp(-c\underline{d})$, $\hat{\Theta}$ defined according to (1) satisfies,

$$\|\hat{\Theta}-\Theta\|_F^2 \lesssim \underbrace{r_* + \bar{r}^2\bar{d} + \bar{r}d_*^{1/2}}_{(*)} + \underbrace{\bar{d}}_{(**)}.$$

Notice that

Remark 1. The condition of $\|X_\perp\|_{\text{sp}} \leq \sqrt{\bar{d}}$ should be changed to $\|X_\perp\|_F \leq \sqrt{\bar{d}}$ in the previous note.

Remark 2. This new theorem incorporates many scenarios with different ranks. Notice that when $r_k = 0$ for all $k\in[m]$, by definition (1), $\hat{\Theta} = 0$, therefore our estimation has

$$\|\hat{\Theta}-\Theta\|_F^2 \leq \|\mathcal{X}_\perp\|_F \leq \bar{d},$$

which is a special case of Theorem 1.1 when $r_k = 0$ for all $k\in[m]$.

Notice that the previous theorem assumes that $r_k = \sqrt{d_k}$ so that $\bar{d}$ term is absorbed into $\bar{r}^2\bar{d}$.

2 When signal is from α-smooth function

For α-smooth functino, we are using the following lemma

Lemma 2.1 (Block approximation)**.** Suppose the true parameter Θ admits α-smoothness, i.e., $f\in\mathcal{H}(\alpha)$. For every integer $r\leq n$, there exists block tensor $\mathcal{X} = \text{Block}_r(\theta)$, satisfying

$$\|\Theta-\mathcal{X}\|_F^2 \lesssim \left(\frac{d^m}{r^{2\alpha}}\right).$$

Notice the rank of $\mathcal{X}$ is at most $(r,r,\ldots,r)$. If we apply this setting to Theorem 1.1, $(*)$ term becomes

$$(*) = r^m + r^2 d + r d^{m/2}$$

under the condition that $\left(\frac{d^m}{r^{2\alpha}}\right) \leq d$ equivalently, $d^{\frac{m-1}{2\alpha}} \leq r$. The second term $(**)$ becomes

$$(**) = \left(\frac{d^m}{r^{2\alpha}}\right).$$

Therefore, α-smooth case, we choose r balancing $(*)$ and $(**)$ under the constraints $d^{\frac{m-1}{2\alpha}} \leq r$. Table 1 shows the combinations giving the optimal convergence rate balancing r between $(*)$ and $(**)$. It is interesting to see that dominating term changes at smoothness $= 1$. I am looking for the general formula of those combinations when $m > 5$.

	Smoothness	Dominating term	Optimal rank	Convergence rate
$m = 2$	$\alpha > 0$	$r^2 d$	$r = d^{1/2(\alpha+1)}$	$d^{(\alpha+2)/(\alpha+1)}$
$m = 3$	$0 < \alpha < 1$	$r^2 d$	$r = d^{3/(2\alpha+1)}$	$d^{(2\alpha+4)/(2\alpha+1)}$
	$\alpha \geq 1$	$r d^{m/4}$	$r = d^{1/(\alpha+1)}$	$d^{(3\alpha+5)/(2\alpha+1)}$
$m = 4$	$0 < \alpha < 1$	r^m	$r = d^{2/(2\alpha+1)}$	$d^{8/(2\alpha+1)}$
	$\alpha > 1$	$r d^{m/4}$	$r = d^{2/(\alpha+2)}$	$d^{(2\alpha+6)/(\alpha+2)}$

Table 1: The best combinations of (Smoothness, tensor mode, rank) to obtain the optimal convergence rate.