Gaussian width and Statistical convergence

Zhuoyan Xu

8/4/2019

1 Gaussian width for tucker tensor

In this section, we show **Gaussian width for a rank** $(R_1, \ldots, R_K)$ **tucker tensor is the Gaussian width for this tensor's tucker core**.

Consider a orthonormal tucker decomposition

$$\mathcal{Y} = \mathcal{C} \times_1 A^1 \ldots \times_K A^K$$

where $\mathcal{Y} \in R^{d_1 \times \cdots \times d_K}$, $\mathcal{C} \in R^{R_1 \times \cdots \times R_K}$.

Consider $\mathcal{E}$ as a random sub-Gaussian tensor that $\forall$ entries $\varepsilon \in \mathcal{E}$, $\varepsilon \in \mathrm{sG}\left(\sigma^2\right)$. We have:

$$\begin{aligned}
\langle \mathcal{Y}, \mathcal{E} \rangle =& \langle \mathcal{C} \times_1 A^1 \ldots \times_K A^K, \mathcal{E} \rangle \\
=& \langle \mathcal{C}, \mathcal{E} \times_1 (A^1)^T \ldots \times_K (A^K)^T \rangle \\
=& \langle \mathcal{C}, \mathcal{B} \rangle
\end{aligned}$$

where

$$\begin{aligned}
\mathcal{B} =& \mathcal{E} \times_1 (A^1)^T \ldots \times_K (A^K)^T \\
\mathcal{B}_{i_1 \cdots i_K} =& \sum_{j_1=1}^{d_1} \cdots \sum_{j_K=1}^{d_K} \mathcal{E}_{i_1 \cdots i_K} A^1_{j_1 i_1} A^2_{j_2 i_2} \cdots A^K_{j_K i_K}
\end{aligned}$$

Since

$$E\left[e^{t \mathcal{E}_{i_1 \cdots i_K}}\right] \leq e^{\sigma^2 t^2 / 2}$$

then we have:

$$E\left[exp\{t \mathcal{E}_{i_1 \cdots i_K} A^1_{j_1 i_1} A^2_{j_2 i_2} \cdots A^K_{j_K i_K}\}\right] \leq exp\{\sigma^2 (A^1_{j_1 i_1})^2 (A^2_{j_2 i_2})^2 \cdots (A^K_{j_K i_K})^2 t^2 / 2\}$$

Thus, we have:

$$
\begin{aligned}
E\left[exp\{t\mathcal{B}_{i_1\cdots i_K}\}\right] \leq & exp\{\frac{\sigma^2 t^2}{2}\sum_{j_1=1}^{d_1}(A^1_{j_1 i_1})^2\cdots\sum_{j_K=1}^{d_K}(A^K_{j_K i_K})^2\} \\
= & e^{\sigma^2 t^2/2}
\end{aligned}
$$

Therefore, $\mathcal{B}$ is also a random sub-Gaussian tensor that $\forall$ entries $\beta \in \mathcal{B}$, $\beta \in \mathrm{sG}\left(\sigma^2\right)$.

Recall

$$
\langle \mathcal{Y}, \mathcal{E}\rangle = \langle \mathcal{C}, \mathcal{B}\rangle
$$

We conclude the Gaussian width for a tucker tensor is the same as the gaussian width for its tucker core.

2 Statistical convergence rate

Recall the note in *Evidence Theory about statistical convergence* in 7/30/2019, consider a K-way tensor, we obtain the convergence rate as

$$
\frac{1}{\sqrt{\prod_k d_k}}\left\|\hat{\Theta}-\Theta_{\text{ true }}\right\|_F = 2C_2\frac{L_\alpha}{\gamma_\alpha}\sqrt{\frac{\prod_{k=1}^{K-1}R_k\sum_{k=1}^K d_k}{\Pi_k d_k}}
$$

We simplify the condition that $R_1=\cdots=R_K=R$ and $d_1=\cdots=d_K=d$, we obtain the rate as R^{K-1}/d^{K-1}. We called it the *initial rate* in our last meeting.

With the property above, we can improve the sharpness of this rate.

Following the main step in *Evidence Theory about statistical convergence* in 7/30/2019, consider the linear term in second-order Taylor's series. We have:

$$
\left|\left\langle S_{\mathcal{Y}}\left(\Theta_{\text{ true }}\right), \Theta-\Theta_{\text{ true }}\right\rangle\right| \leq \max\left|\langle S, \mathcal{C}\rangle\right| \leq \|S\|_\sigma \|\mathcal{C}\|_*
$$

where $\mathcal{S}\in R^{R_1\times\cdots\times R_K}$ is a random tensor whose entries are independently distributed and satisfy

$$
E\left(s_{i_1,\ldots,i_K}\right)=0, E\left(e^{tL_\alpha^{-1}s_{i_1,\ldots,i_K}}\right)\leq e^{t^2/2}
$$

Thus, with probability at least $1 - \exp\left(-C_1 \log K \sum_k R_k\right)$:

$$\|S\|_\sigma \leq C_2 L_\alpha \log K \sqrt{\sum_k R_k} \tag{1}$$

Consider $\mathcal{C}$ is the tucker core tensor of $\Theta - \Theta_{\text{ true }}$.
We still have:

$$\|\mathcal{C}\|_* \leq \sqrt{\prod_{k=1}^{K-1} 2R_k} \, \|\Theta - \Theta_{true}\|_F \tag{2}$$

Combining the results above, we have, with probability at least $1 - \exp\left(-C_1 \log K \sum_k R_k\right)$:

$$\left|\left\langle S_{\mathcal{Y}}\left(\Theta_{\text{ true }}\right), \Theta - \Theta_{\text{ true }}\right\rangle\right| \leq C_2 L_\alpha \sqrt{\prod_{k=1}^{K-1} R_k \sum_{k=1}^{K} R_k} \, \|\Theta - \Theta_{\text{ true }}\|_F$$

Henceforth,

$$\frac{1}{\sqrt{\prod_k d_k}} \left\|\hat{\Theta} - \Theta_{\text{ true }}\right\|_F \leq \frac{2C_2 L_\alpha \sqrt{\prod_{k=1}^{K-1} R_k \sum_{k=1}^{K} R_k}}{\gamma_\alpha \sqrt{\Pi_k d_k}} = 2C_2 \frac{L_\alpha}{\gamma_\alpha} \sqrt{\frac{\prod_{k=1}^{K-1} R_k \sum_{k=1}^{K} R_k}{\Pi_k d_k}}$$

When simplifying the condition, we obtain the rate as R^K/d^K. This rate surpassed all the rate we've got so far, and also attained the rate in the professor's conjecture.