Fundamentals of Mathematical and Computing Sciences:
Applied Mathematical Science

Part III: Low rank matrix estimation (Lecture 4) Statistical property

2014-7-17

Taiji Suzuki (Room W707, post W8-46)
e-mail: suzuki.t.ct@m.titech.ac.jp

Today's topic:

- Convergence rate of the trace norm regularized estimator.

1 Preliminary

Model:

$$y_i = \langle X_i, A^* \rangle + \epsilon_i,$$

where A^* is the true matrix and $\epsilon_i \sim N(0, \sigma^2)$. We assume that the rank of A^* is d^*.
Estimator:

$$\hat{A} \mathop{\arg\min}_{A \in \mathbb{R}^{M \times N}} \frac{1}{n} \|Y - \mathcal{X}(A)\|^2 + \lambda_n \|A\|_{\mathrm{Tr}},$$

where $Y = [y_1, \ldots, y_n]^\top$, and $\mathcal{X}(A) = [\langle X_1, A \rangle, \ldots, \langle X_n, A \rangle]^\top$.

Q: How rapidly does the estimator $\hat{A}$ converge to A^*?
A: Roughly speaking

$$\|\hat{A} - A^*\|_F^2 = O_p\left(\frac{d^*(M+N)\log(MN)}{n}\right).$$

This is much faster than the rate of the standard MLE, $O_p(\frac{MN}{n})$, if $d^* \ll M, N$.

2 Restricted Strong Convexity

Assumption 1 (Restricted Strong Convexity (RSC))**.** $\exists C_1, C_2 > 0$ *such that*

$$\frac{\|\mathcal{X}(A)\|}{\sqrt{n}} \geq C_1 \|A\|_F - C_2 \left(\frac{\sqrt{M} + \sqrt{N}}{\sqrt{n}}\right) \|A\|_{\mathrm{Tr}},$$

for all $A \in \mathbb{R}^{M \times N}$.

Example 2. *If each element of* X_i *is i.i.d.* $N(0,1)$*, then*

$$\frac{\|\mathcal{X}(A)\|}{\sqrt{n}} \geq \frac{1}{4} \|A\|_F - 4 \left(\frac{\sqrt{M} + \sqrt{N}}{\sqrt{n}}\right) \|A\|_{\mathrm{Tr}},$$

for all $A \in \mathbb{R}^{M \times N}$*, with probability at least* $1 - 2\exp(-n/32)$.

This can be shown by using the following propositions.

Definition 3 (Gaussian process)**.** *A set of random variables $\{G_x\}_{x\in\mathcal{X}}$ on a set $\mathcal{X}$ is called Gaussian process if for all finite combination $\{x_1,\dots,x_k\}\subset\mathcal{X}$, the joint distribution of $(G_{x_1},\dots,G_{x_k})$ obeys a multivariate Gaussian distribution.*

Proposition 4 (Gordon-Slepian's Inequality [1])**.** *Consider two centered Gaussian processes $\{G_{u,v}\}_{(u,v)}$ and $\{G'_{u,v}\}_{(u,v)}$ indexed by $(u,v)\in\mathcal{U}\times\mathcal{V}$ ("centered" means $\mathrm{E}[G_{u,v}=0]$ for all u,v). If G and G' satisfy*

$$\begin{aligned}\mathrm{E}[(G_{u,v}-G_{u',v'})^2] &\geq \mathrm{E}[(G'_{u,v}-G'_{u',v'})^2],\\ \mathrm{E}[(G_{u,v}-G_{u,v'})^2] &= \mathrm{E}[(G'_{u,v}-G'_{u,v'})^2],\end{aligned}$$

for all $(u,v),(u',v')\in\mathcal{U}\times\mathcal{V}$. Then

$$\mathrm{E}[\inf_{u\in\mathcal{U}}\sup_{v\in\mathcal{V}}G_{u,v}]\leq\mathrm{E}[\inf_{u\in\mathcal{U}}\sup_{v\in\mathcal{V}}G'_{u,v}].$$

Applying this lemma to $-G_{u,v},-G'_{u,v}$ under an assumption that $\mathcal{V}$ is a singleton, we obtain classical **Slepian's inequality**:

$$\begin{aligned}&\mathrm{E}[(G_u-G_{u'})^2]\geq\mathrm{E}[(G'_u-G'_{u'})^2]\quad(\forall u,u'\in\mathcal{U})\\ \Rightarrow\quad&\mathrm{E}[\sup_u G_u]\geq\mathrm{E}[\sup_u G'_u].\end{aligned}$$

Proposition 5 (Gaussian concentration inequality)**.** *Let $X\in\mathbb{R}^m$ be i.i.d. $N(0,1)$ and $f:\mathbb{R}^m\to\mathbb{R}$ be Lipschitz continuous function with the continuity parameter L, $|f(x)-f(y)|\leq L\|x-y\|$. Then,*

$$P\left(|f(X)-\mathrm{E}[f(X)]|\geq\delta\right)\leq 2\exp\left(-\frac{\delta^2}{L^2}\right)\quad(\forall\delta>0).$$

See Proposition 2.18 of [2] for the proof.

3 Operator norm of sum of matrix Gaussian series

Lemma 6. *Suppose $\{X_i\}_{i=1}^n$ are fixed and satisfy*

$$\max\{\|\sum_{i=1}^n X_iX_i^\top\|_\infty,\|\sum_{i=1}^n X_i^\top X_i\|_\infty\}\leq C_3n(M+N),$$

for some constant $C_3>0$. Then, we have that

$$P\left(\left\|\frac{1}{n}\sum_{i=1}^n\epsilon_iX_i\right\|_\infty\geq\sqrt{C_3\sigma^2\frac{M+N}{n}\log\left(\frac{2(M+N)}{\delta}\right)}\right)\leq\delta\quad(\forall\delta>0).$$

See [3] for the proof. The assumption is satisfied if $\|X_i\|_\infty\leq C(\sqrt{M}+\sqrt{N})$, which is true with high probability if each element of X_i is i.i.d. Gaussian.

3.1 Special case: i.i.d. Gaussian

Lemma 7. *If each element of X_i is i.i.d. $N(0,1)$, then*

$$\mathrm{E}\left[\left\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i X_i\right\|_\infty\right] \le \frac{\sqrt{M}+\sqrt{N}}{\sqrt{n}},$$

and

$$\left\|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i X_i\right\|_\infty \le (1+t)\sigma\frac{\sqrt{M}+\sqrt{N}}{\sqrt{n}}t',$$

with probability $1-2\exp(-t^2(\sqrt{M}+\sqrt{N})^2/2)-\frac{1}{t'}\exp(-t'^2/2)$ for all $t,t'>0$.

4 Convergence rate of trace norm regularized estimator

Combining lemmas shown above, we obtain the following theorem.

Theorem 8. *Suppose $\{X_i\}_{i=1}^n$ are fixed and satisfy*

$$\max\{\|\sum_{i=1}^{n}X_iX_i^\top\|_\infty, \|\sum_{i=1}^{n}X_i^\top X_i\|_\infty\} \le C_3 n(M+N),$$

for some constant $C_3>0$. Assume the number n of samples satisfies

$$n \ge \left[\frac{16C_2\sqrt{d^*}}{C_1}(\sqrt{M}+\sqrt{N})\right]^2.$$

Let the regularization constant $\lambda_n = 2\sqrt{C_3\sigma^2\frac{M+N}{n}\log\left(\frac{2(M+N)}{\delta}\right)}$ for some $\delta>0$. Then under RSC condition (Assumption 1), there exists a universal constant $c>0$ such that

$$\|\hat{A}-A^*\|_F^2 \le c\frac{C_3\sigma^2}{C_1^4}\frac{d^*(M+N)}{n}\log\left(\frac{2(M+N)}{\delta}\right),$$

with probability $1-\delta$.

References

[1] Y. Gordon. Some inequalities for gaussian processes and applications. *Israel Journal of Mathematics*, 50(4):265–289, 1985.

[2] M. Ledoux. *The Concentration of Measure Phenomenon*, volume 89 of *Mathematical Surveys and Monographs*. American Methematical Society, 2001.

[3] J. A. Tropp. User-friendly tail bounds for sums of random matrices. *Foundations of Computational Mathematics*, 12(4):389–434, 2012.