HDP
(16) Symmetrization and Contraction Principle

@ 滕佳烨

前情回顾：Concentration of Quadratic

$$X^\top AX \to \mathbb{E}X^\top AX = \operatorname{tr} A$$

Technique: Decouple from $X^\top AX$ to $X^\top AX'$

Today, we focus on another application of decoupling:

Symmetrization

Symmetrization

Lemma 6.4.2 (Symmetrization). *Let $X_1, \dots, X_N$ be independent, mean zero random vectors in a normed space. Then*

$$\frac{1}{2}\mathbb{E}\Big\|\sum_{i=1}^{N}\varepsilon_i X_i\Big\| \le \mathbb{E}\Big\|\sum_{i=1}^{N}X_i\Big\| \le 2\,\mathbb{E}\Big\|\sum_{i=1}^{N}\varepsilon_i X_i\Big\|.$$

The purpose of this lemma is to let us replace general random variables X_i by the symmetric random variables $\varepsilon_i X_i$.

优势: 比起一般的X，可以考虑一个性质更好的ϵX（具有对称性）

Hint：引入X的独立副本X'，注意 $\epsilon\,(X - X') =^d X - X'$

Similar conclusion follows with $\mathbb{E}X \neq 0$, or on $F(\|\sum X\|)$ when F is increasing convex, or X_i is subGaussian with $\|\|_{\psi_2}$

Symmetrization (Gaussian version)

Lemma 6.4.2 (Symmetrization). *Let $X_1, \ldots, X_N$ be independent, mean zero random vectors in a normed space. Then*

$$\frac{1}{2}\mathbb{E}\left\|\sum_{i=1}^{N}\varepsilon_i X_i\right\| \le \mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\| \le 2\,\mathbb{E}\left\|\sum_{i=1}^{N}\varepsilon_i X_i\right\|.$$

Lemma 6.7.4 (Symmetrization with Gaussians). *Let $X_1, \ldots, X_N$ be independent, mean zero random vectors in a normed space. Let $g_1, \ldots, g_N \sim N(0,1)$ be independent Gaussian random variables, which are also independent of X_i. Then*

$$\frac{c}{\sqrt{\log N}}\mathbb{E}\left\|\sum_{i=1}^{N} g_i X_i\right\| \le \mathbb{E}\left\|\sum_{i=1}^{N} X_i\right\| \le 3\,\mathbb{E}\left\|\sum_{i=1}^{N} g_i X_i\right\|.$$

Contraction Principle

Theorem 6.7.1 (Contraction principle). *Let* $x_1, \ldots, x_N$ *be (deterministic) vectors in some normed space, and let* $a = (a_1, \ldots, a_n) \in \mathbb{R}^n$. *Then*

$$\mathbb{E} \left\| \sum_{i=1}^{N} a_i \varepsilon_i x_i \right\| \le \|a\|_\infty \cdot \mathbb{E} \left\| \sum_{i=1}^{N} \varepsilon_i x_i \right\|.$$

Similar results hold on general independent random vectors X_i, and on random process (Talagrand's contraction principle, will be discussed later.)

Take-away Messages

1. Symmetrization: transfer random variable (X) to its symmetric version (ϵX)
   - Gaussian symmetric version
   - Convex function version (F)
   - Sub-Gaussian X version
2. Contraction Lemma: bound $\mathbb{E}\|\sum a_i \epsilon_i X_i\|$ with $\mathbb{E}\|\sum \epsilon_i X_i\|$ where a_i is constant

Appendix: An application to matrix completion. Note the passing from the operator norm to the F-norm.

$$\mathbb{E}\,\frac{1}{n}\|\hat{X} - X\|_F \le C\sqrt{\frac{rn\log n}{m}}\,\|X\|_\infty,$$

@ 滕佳烨

Thanks!