Probability Inequalities in Sparse Recovery Problems

Vladimir Koltchinskii
School of Mathematics
Georgia Institute of Technology

Let (X, Y) be a random couple in $S \times \mathbb{R}$ with distribution P and let $(X_1, Y_1), \dots, (X_n, Y_n)$ be i.i.d. copies of (X, Y). Given a dictionary consisting of N functions $h_1, \dots, h_N : S \mapsto \mathbb{R}$, denote

$$f_\lambda := \sum_{j=1}^{N} \lambda_j h_j, \ \lambda \in \mathbb{R}^N.$$

Consider the following penalized empirical risk minimization problem (ERM)

$$\hat{\lambda}^\varepsilon := \mathrm{argmin}_{\lambda \in \mathbb{R}^N} \left[n^{-1} \sum_{j=1}^{n} (f_\lambda(X_j) - Y_j)^2 + \varepsilon \|\lambda\|_{\ell_p}^p \right],$$

which is an empirical version of the "true" risk minimization problem

$$\lambda^\varepsilon := \mathrm{argmin}_{\lambda \in \mathbb{R}^N} \left[\mathbb{E}(f_\lambda(X) - Y)^2 + \varepsilon \|\lambda\|_{\ell_p}^p \right]$$

Here $\varepsilon \geq 0$ is a regularization parameter.

Suppose the "true" solution of the problem (λ^ε or λ^0) is "sparse". Does it imply that its empirical solution $\hat{\lambda}^\varepsilon$ is also "sparse" or at least "approximately sparse?" What is the impact of "sparsity" on bounding the errors $\|f_{\hat{\lambda}^\varepsilon} - f_{\lambda^\varepsilon}\|_{L_2(P)}$ and $\|f_{\hat{\lambda}^\varepsilon} - f_{\lambda^0}\|_{L_2(P)}$ of the solution of ERM? There has been been a significant progress in understanding of these problems in the recent years, based on the methods of asymptotic geometric analysis and high-dimensional probability (concentration inequalities, various bounds for Rademacher sums, etc).

The most important case is $p = 1$. In this case, there is an interesting version of the problem due to Candes and Tao (the Dantzig selector). We will discuss this circle of problems for $p = 1$ and also for $p > 1$, but close enough to 1 (so that $p - 1$ is of the order $\frac{1}{\log N}$). Surprisingly enough, in this case one can prove some inequalities describing the sparsity of the empirical solution that seem to be unavailable for $p = 1$.