Adaptivity in high-dimensional statistics

Sara van de Geer
Seminar for Statistics, ETH Zürich

In high-dimensional problems the models are severely over-parametrized, meaning that the complexity of the parameter space $\mathcal{B}$ is much higher than the information contained in the data. An algorithm based on these data is adaptive if its error depends on the particular element $\beta \in \mathcal{B}$. That is, some β's are easier to estimate than others. On way to quantify the complexity of β is in terms of its *sparseness*: its number of non-zero coefficients s. To arrive at adaptivity one may use regularization with a suitable norm on $\mathcal{B}$ such as the ℓ_1-norm, a structured sparsity norm, or the sorted ℓ_1-norm (when $\mathcal{B}$ is a collection of vectors), the nuclear norm (when $\mathcal{B}$ a collection of matrices) or a tensor norm (when $\mathcal{B}$ is a collection of tensors). We consider empirical risk minimization with norm-regularization and present a bound for the excess risk that roughly says that up to log-terms the error of this algorithm behaves like s/n where n is the sample size. Without going into too much detail we will indicate that the theory relies on results from several mathematical fields: convex analysis, concentration of measure, geometry, random matrix theory, approximation theory. As examples, we look at the Lasso, matrix completion from binary observations, robust matrix completion and sparse principal components.

Reference

Sara van de Geer *Estimation and testing under sparsity*, Springer, Saint-Flour Lecture Notes (2016)