Synthesis of optimal sketching estimators

2019-2020 Master Thesis Proposal

Supervisor: Anatoli IOUDITSKI, anatoli.juditsky@univ-grenoble-alpes.fr

Description: We consider the estimation problem as follows:

suppose that $x \in \mathcal{X} \subset \mathbf{R}^n$ is unknown signal to be recovered, and $\mu = A(x)$ is an affine image of x. Our objective is, given a "noisy observation" $\omega_i \sim P_\mu,\ i = 1, ..., N$ where P_μ is a probability distribution, parameterized by μ, build an estimate $\widehat{x}_N$ of the signal. Here $\mathcal{X}$ is a known compact set representing *a priori* information about the signal, $A(\cdot)$ is a given linear mapping and $\mathcal{P} = \{P_\mu,\ \mu \in A(\mathcal{X})\}$ is a given family of distributions. The recovery error is measured, e.g., by the risk

$$\mathrm{Risk}[\widehat{x}|\mathcal{X}] = \sup_{x\in\mathcal{X}} \mathbf{E}_{\omega\sim A(x)} \{\|x - \widehat{x}_N(\omega)\|\} \tag{1}$$

where $\|\cdot\|$ is a given norm on $\mathbf{R}^n$.

Our objective is to study *sketching estimators* of x, e.g., the estimators of the form

$$\widehat{x}(\omega)_N(\omega) \in \underset{x\in\mathcal{X}}{\mathrm{Argmin}}\, \|H(\omega - Ax)\|_\infty \tag{2}$$

where $H(\cdot)$ is a given *contrast linear mapping.*

The idea of construction of the estimator (2) is simple: in the contrast estimator all information about the data is summarized (or "sketched," when using modern statistical slang) in the values $h_k^T\omega$ of (hopefully) few linear functionals of observations (h_k^T being raws of the "contrast matrix" H). Special structure of estimator (2) is of particular interest in the treatment of massive datasets when simple transmission and storage of the data at the treatment site are difficult to implement.

Note that estimator (2) can be seen as an extension of the linear estimator $\widehat{x}_N(\omega) = H(\omega)$, which possesses well studied properties of optimality in a quite general setting, see e.g., [1] and references therein. However, it is well known that in some important cases, linear estimators are suboptimal. For instance, when $\mathcal{X}$ is a unit ball of ℓ_1-norm, and the observation ω is "direct:" $\omega = x + \sigma\xi$ with normal noise $\xi \sim \mathcal{N}(0, I_n)$, the ℓ_2-risk (i.e., when the norm $\|\cdot\|$ in the risk definition (1) is the $\|\cdot\|_2$-norm) of the best linear estimation is, up to an absolute constant, $\frac{\sigma\sqrt{n}}{1+\sigma\sqrt{n}}$. On the other hand, it is also well known that the maximal over $\mathcal{X}$ risk of the simple estimator $\widehat{x}_N(\omega) = \min_{x\in\mathcal{X}} \|x - \omega\|_\infty$ is $O\left(\ln^{1/4}[n/\sigma]\sigma^{1/2}\right)$, which is much less than the best risk of the linear estimation in the range $n^{-1/2} \ll \sigma \ll 1$ (and coincides up to a moderate absolute factor with the *minimax risk* – the maximal over $\mathcal{X}$ risk of the best possible estimator "existing in nature"). It is obvious that the estimator $\widehat{x}_N(\cdot)$ is the simplest *contrast estimator* (2) with identity contrast mapping.

In the hindsight, the contrast estimator can be seen as a further development of the ideas behind Nemirovski's nonlinear estimator [2], which have recently seen a regain of interest [3].

The subject of this thesis is to study efficiently computable via convex programming contrast estimators (2) which outperform linear estimates (are provably near-optimal in the case where linear estimators are), and can also be used (and under favorable circumstances is near-optimal) in a generic problem described above. We will also study applications of this approach to building efficient estimators to the Poisson imaging model (e.g., Positron Emission Tomography) and missing observation model.

The Master thesis will be supervised by Anatoli Iouditski at LJK, and is expected to lead to a PhD thesis on the same topic.

[1] Juditsky, A and Nemirovski, A. (2018). Near-optimality of linear recovery in Gaussian observation scheme under $\|\cdot\|_2^2$-loss. *The Annals of Statistics*, **46**(4), 1603–1629.

[2] Nemirovskii, A. S. (1985). Nonparametric estimation of smooth regression functions. *Izv. Akad. Nauk. SSR Teckhn. Kibernet*, 3:50-60 (Translated as *J. Comput. System Sci.*, 23:1-11, 1986).
[3] Grasmair, M., Li, H., and Munk, A. (2015). Variational multiscale nonparametric regression: smooth functions. *arXiv preprint arXiv:*1512.01068