ԵՐԵՎԱՆԻ ՊԵՏԱԿԱՆ ՀԱՄԱԼՍԱՐԱՆ

Մինասյան Արշակ Գագիկի

Ռոբաստ գնահատում գաուսյան միջինի համար հաշվարկելիության տիրույթում

Ա.01.05 «Հավանականությունների տեսություն և մաթեմատիկական վիճակագրություն» մասնագիտությամբ ֆիզիկամաթեմատիկական գիտությունների թեկնածուի գիտական աստիճանի հայցման ատենախոսության

ՍԵՂՄԱԳԻՐ

Երևան - 2020

YEREVAN STATE UNIVERSITY

Arshak Minasyan

Robust Estimation of Gaussian Mean within the Domain of Computational Tractability

SYNOPSIS

of dissertation for the degree of candidate of physical and mathematical sciences specializing in A.01.05 – ”Probability theory and mathematical statistics”

Yerevan - 2020

Ատենախոսության թեման հաստատվել է Երևանի պետական համալսարանում:

Գիտական ղեկավար՝	ֆիզ.-մաթ. գիտ. դոկտոր Վ.Կ. Օհանյան
Պաշտոնական ընդդիմախոսներ՝	ֆիզ.-մաթ. գիտ. դոկտոր Ռ.Հ. Արամյան
	ֆիզ.-մաթ. գիտ. թեկնածու Ք. Լունիսի
Առաջատար կազմակերպություն՝	ՀՀ ԳԱԱ Ինֆորմատիկայի և ավտոմատացման պրոբլեմների ինստիտուտ

Պաշտպանությունը կայանալու է 2020թ. նոյեմբերի 17-ին, ժ. 15^{00}-ին ԵՊՀ-ում գործող ԲՈԿ-ի 050 մասնագիտական խորհրդի նիստում հետևյալ հասցեով՝ 0025, Երևան, Ալ. Մանուկյան 1:

Ատենախոսությանը կարելի է ծանոթանալ ԵՊՀ-ի գրադարանում:

Սեղմագիրն առաքված է 2020 հոկտեմբերի 6-ին:

Մասնագիտական խորհրդի
գիտական քարտուղար՝ Տ. Ն. Հարությունյան

Dissertation topic was approved at the Yerevan State University.

Supervisor:	Doctor of phys-math sciences V.K. Ohanyan
Official opponents:	Doctor of phys-math sciences R.H. Aramyan
	Candidate of phys-math sciences K. Lounici
Leading organization:	Institute for informatics and automation problems of NAS RA

Defense of the thesis will be held at the meeting of the specialized council 050 of SCC (Supreme Certifying Committee) of Armenia at Yerevan State University on November 17, 2020 at 15^{00} (0025, Yerevan, A. Manoogian str. 1).

You can get acquainted with the thesis in the library of the YSU.

Synopsis was sent on October 6, 2020.

Scientific secretary of specialized council, T.N. Harutyunyan

General characteristics of the work

Relevance of the theme.

Robustness-to-outliers is a fundamental problem in statistics, which aims at designing statistical procedures that remain stable in presence of outliers. A significant breakthrough in the field of robust statistics was done by Peter J. Huber with seminal papers ([10, 11]). History and experience of collecting datasets shows that the outliers were mistakes in data collection, being it incorrect input by a human or machine or including observations from a different population than that of interest. Having robust-to-outliers methods is becoming more and more critical, since automatically collected datasets are often heterogeneous. Recent advances in data acquisition and computational power provoked a revival of interest in robust estimation and learning, with a focus on finite sample results and computationally tractable procedures. This was in contrast with more traditional studies analyzing asymptotic properties of statistical methods.

Loss functions that lead to robust M-estimators [15] are the most popular ones across the field of statistics and machine learning, since they share a number of useful and desirable properties for proposed estimators. It is well known that the minimization of the ℓ_1 loss function leads to the median. In a more general way, Lipschitz loss function minimization leads to more stable (to outliers) methods. Examples of such loss function are Huber's loss function [10], quantile losses, above mentioned ℓ_1 loss and many others.

To the best of our knowledge, the form $\sqrt{p/n} + \varepsilon$ of the minimax risk in the Gaussian mean estimation problem has been first obtained by [2]. They proved that this rate holds with high probability for the Tukey median, which is known to be computationally intractable in the high-dimensional setting. The first nearly-rate-optimal and computationally tractable estimators have been proposed in [12] and [6]. The methods analyzed in these papers are different, but they share the same idea: If for a subsample of points the empirical covariance matrix is sufficiently close to the theoretical one, then the arithmetic mean of this subsample is a good estimator of the theoretical mean.

Further improvements in running times—up to obtaining a linear in np computational complexity in the case of a constant ε—are presented in [3]. Some lower bounds suggesting that the log-factor in the term $\varepsilon\sqrt{\log(1/\varepsilon)}$ cannot be removed from the rate of computationally tractable estimators are established in [9]. In a slightly weaker model of corruption, [7] propose an iterative filtering algorithm that achieves the optimal rate ε without the extra factor $\sqrt{\log(1/\varepsilon)}$. On a related note [4] shows that in a weaker contamination model termed as parametric contamination, the carefully trimmed mean can achieve a better rate than that of the coordinatewise/geometric median.

An overview of the recent advances on robust estimation with a focus on computational aspects can be found in [8]. All these results are proved to hold on an event with a prescribed probability, see [1] for a relation between results in expectation and those with high probability, as well as for the definitions of various types of contamination.

An appealing feature of the risk bounds that hold with high probability is that they allow us to apply Lepski's method [13] for obtaining an adaptive estimator with respect to ε. The obtained adaptive estimator enjoys all the five properties enumerated above except the asymptotic efficiency, since the adaptation results in a inflation of the risk bound by a factor 3, see [4] for more details.

The aim of the thesis:

1. Establish an information theoretic bound for the ℓ_2 loss of coordinate-wise median $\hat{\boldsymbol{\mu}} = \mathtt{Med}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$ of observations $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ both in probability and in expectation.
2. Construct an estimator $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ that is translation and orthogonal transformation invariant.
3. Establish the breakdown point of the estimator $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$.
4. Prove that $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ is nearly minimax rate optimal with the worst-case risk of order
$$\|\boldsymbol{\Sigma}\|_{\mathsf{op}}^{1/2}\big(\sqrt{\mathbf{r}_{\boldsymbol{\Sigma}}/n} + \varepsilon\sqrt{\log(1/\varepsilon)}\big),$$
with $\mathbf{r}_{\boldsymbol{\Sigma}} = \frac{\mathsf{Tr}(\boldsymbol{\Sigma})}{\|\boldsymbol{\Sigma}\|_{\mathsf{op}}}$ being the effective rank of matrix $\boldsymbol{\Sigma}$.
5. Prove that estimator $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ is asymptotically efficient.
6. Construct the robust estimators in the setting of excess-risk setup such that
$$\overline{\lim_{n\to\infty}}\, \mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}, n, p_n, \varepsilon_n) = 0.$$
7. Modify the estimator $\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}$ to obtain the minimax rate for excess-risk loss
$$\mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}, \varepsilon) \le \sqrt{\varepsilon\frac{p}{n}} + C\varepsilon^2\sqrt{p}.$$
8. Illustrate numerical robustness of Algorithms 1 and 2.
9. Illustrate numerical robustness of Projected Iterative Reweighted Least Squares (Pr-IRLS) optimization method and prove convergence of the algorithm.

The methods of investigation.

In this thesis we apply methods and techniques obtained on the basis of high-dimensional statistics, gaussian processes, probabilistic inequalities, minimax theory and related topics.

Scientific innovation.

All results are new and are published in local and international conferences and journals.

Practical and theoretical value.

The results of the work both have theoretical and practical character. The theoretical results are devoted the minimax theory yielding nearly minimax optimal estimators. Presented various estimators are shown (numerically) to be robust to the existence of outliers.

Approbation of the results.

The presented results were presented in the scientific seminar at Yerevan State University. Some parts of obtained results were presented in local and international conferences.

Publications.

The main results of this thesis have been published in 3 scientific articles in journals, 2 abstracts in conferences and 1 arXiv paper. The list of the articles is given at the end of the Synopsis.

The structure and the volume of the thesis.

The thesis consists of introduction, 3 chapters of main results followed by conclusion and discussion, a list of references and 5 appendices. The number of references is 104. The volume of the thesis is 129 pages. The thesis contains 15 figures and 1 table.

The main results of the thesis

Chapter 1.

The aim of the first chapter is to introduce to the problem of robust estimation both from formal and from intuitive point of views. We present the two main robust estimators: *coordinate-wise median* and *trimmed/truncated mean*. Theorems concerning their minimax rates were presented.

Then, the challenges in high-dimensions of robust estimation were briefly discussed and the following overview of the results available in literature was given.

Dimensionality	Error guarantee	Efficiency
In low dimensions		
Coordinate-wise median (Ch. 1)	$\Theta(\varepsilon)$	**Yes**
Truncated mean (Ch. 1)	$\Theta(\varepsilon\sqrt{\log 1/\varepsilon})$	**Yes**
In dimension $p \gg 1$		
Coordinate-wise median (Ch. 2)	$\Theta(\varepsilon\sqrt{p})$	**Yes**
Truncated mean	$\Theta(\varepsilon\sqrt{p})$	**Yes**
Tournament (Ch. 6 [5])	$\Theta(\varepsilon)$	**No**
Geometric median ([14], Ch. 2)	$\Theta(\varepsilon\sqrt{p})$	**Yes**
Tukey's median ([16])	$\Theta(\varepsilon)$	**No**
Iterative Soft Thresholding ([4])	$\Theta(\varepsilon p^{1/4})$	**Yes**
[6]	$\mathcal{O}(\varepsilon\sqrt{\log 1/\varepsilon})$	**Yes**
Iterative reweighted mean (Ch. 3)	$\mathcal{O}(\varepsilon\sqrt{\log 1/\varepsilon})$	**Yes**
Excess-risk setting		
Group Hard Thresholding (Ch. 3)	$\mathcal{O}\left(\sqrt{\varepsilon\frac{p}{n}}\right)$	**Yes**

The main aim of this thesis is to construct robust estimators with optimal minimax rates, low computational cost and high breakdown point. In the end of the Chapter we also discuss the directions of robust estimation beyond mean estimation.

Chapter 2.

Chapter 2 introduces 3 contamination models that will be used further in the thesis and the robust mean estimators under these contamination models in high-dimensions.

§1. Contamination models.

Huber's contamination model. Let $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n \overset{\text{i.i.d}}{\sim} (1-\varepsilon)\boldsymbol{P}_{\boldsymbol{\mu}} + \varepsilon \boldsymbol{Q}$, where $\boldsymbol{P}_{\boldsymbol{\mu}}$ is the target distribution and $\boldsymbol{Q}$ is an arbitrary distribution of outliers. Notice that under this model assumption we assume that the outliers are all coming from the same distribution. In other words, this assumption means that the presence of outliers is caused by a single reason.

An equivalent formulation is the following: suppose that there exist $Z_1, \dots, Z_n \overset{\text{i.i.d.}}{\sim} \mathcal{B}\text{ern}(\varepsilon)$ such that $\{(\boldsymbol{X}_i, Z_i)\}_{i=1}^n$ are i.i.d. with

$$\mathbb{P}(\boldsymbol{X}_i \in \mathcal{A} \mid Z_i = 0) = \mathbb{P}_{\boldsymbol{\mu}}(\mathcal{A}), \quad \mathbb{P}(\boldsymbol{X}_i \in \mathcal{A} \mid Z_i = 1) = \boldsymbol{Q}(\mathcal{A}).$$

Then, the number of outliers o is defined as follows: $o \overset{\text{def}}{=} \sum_{i=1}^n Z_i$. Huber's contamination model can be summarizes as follows:

$$\mathcal{M}_{\text{HC}}(\varepsilon) = \left\{ \left[(1-\varepsilon)\boldsymbol{P}_{\boldsymbol{\mu}} + \varepsilon \boldsymbol{Q}\right]^{\otimes n} \ : \ \boldsymbol{\mu} \in \mathcal{M}, \boldsymbol{Q} \in \mathcal{P} \right\}, \tag{1}$$

where $\mathcal{P}$ is an arbitrary family of probability distributions.

Parameter contamination model. Fix $o \in \{1, \dots, n\}$ and let $\boldsymbol{X}_i \overset{\text{i.i.d.}}{\sim} \boldsymbol{P}_{\boldsymbol{\mu}_i}$ so that for some fixed set $\mathcal{O} \subset \{1, \dots, n\}$ with $|\mathcal{O}| \le o$ we have $\boldsymbol{\mu}_i = \boldsymbol{\mu}$ for all $i \in \mathcal{O}^c$. Then, we can write

$$\mathcal{M}_{\text{PC}}(o) = \left\{ \boldsymbol{P}_{\boldsymbol{\mu}_1} \otimes \dots \otimes \boldsymbol{P}_{\boldsymbol{\mu}_n} \ : \ \boldsymbol{\mu}_i \in \mathcal{M}, \ \exists \mathcal{O} \subset [n] \text{ s.t. } |\mathcal{O}| \le o \text{ and } \boldsymbol{\mu}_i \ne \boldsymbol{\mu}_j, \ \ \forall i, j \in \mathcal{O} \right\}. \tag{2}$$

In other words, the observations $\boldsymbol{X}_1, \dots, \boldsymbol{X}_n$ follow the following rule:

$$\boldsymbol{X}_i = \boldsymbol{\mu} + \boldsymbol{\theta}_i + \boldsymbol{\xi}_i, \quad \forall i \in [n], \tag{3}$$

where $\boldsymbol{\theta}_i \equiv 0$ for all $i \in \mathcal{O}^c$ and $\boldsymbol{\xi}_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(0, \boldsymbol{\Sigma})$. Notice that under this contamination model assumptions the outliers are still Gaussian itself with the same covariance structure as inliers. The only difference is that they might be shifted in any (one or more) of p directions.

Adversarial contamination model. Fix $o \in \{1, \dots, n\}$. Let all but o vectors from sample $\{\boldsymbol{X}_i\}_{i=1}^n$ are from the distribution $\boldsymbol{P}_{\boldsymbol{\mu}}$ and are independent. Hence $\boldsymbol{X}_I \overset{\text{i.i.d.}}{\sim} \mathcal{P}_{\boldsymbol{\mu}}$ for a set of inliers I with $|I| \ge n - o$. The rest of observation may be arbitrarily selected vectors. Hence,

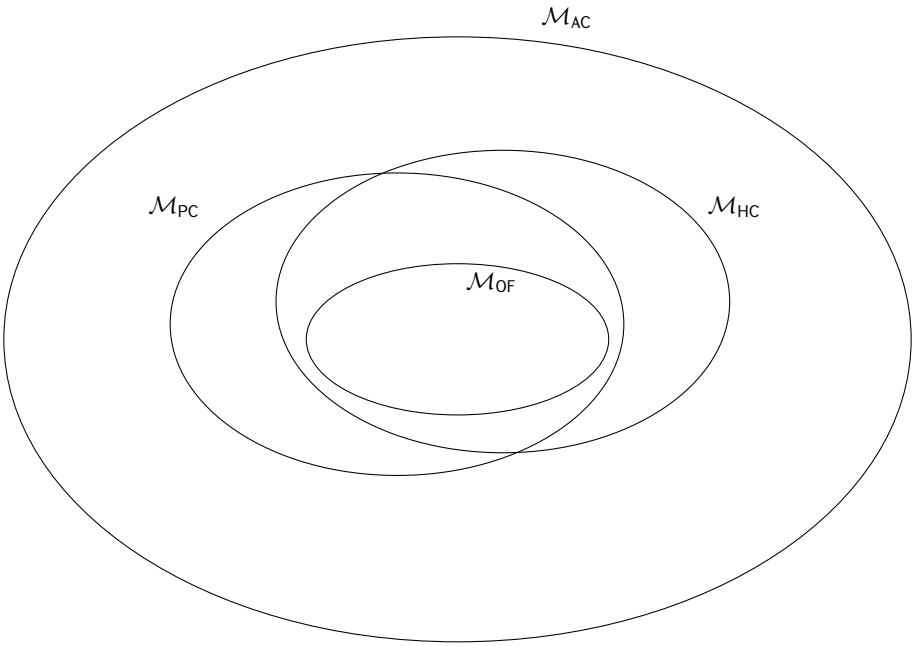
$$\mathcal{M}_{\mathsf{AC}}(o) = \Big\{ \sigma(\boldsymbol{P}_{\boldsymbol{\mu}}^{\bigotimes n-o} \otimes \boldsymbol{P}_1 \otimes \cdots \otimes \boldsymbol{P}_o) \ : \ \boldsymbol{\mu} \in \mathcal{M}, \sigma \in S_n$$

$$\text{permutation group and } \boldsymbol{P}_1, \ldots, \boldsymbol{P}_o \text{ are arbitrary} \Big\}.$$

Definition. We say that the distribution P_n of data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n$ is Gaussian with adversarial contamination, denoted by $\mathsf{P}_n \in \mathsf{GAC}(\boldsymbol{\mu}^*, \boldsymbol{\Sigma}, \varepsilon)$ with $\varepsilon \in (0, 1/2)$ and $\boldsymbol{\Sigma} \succeq 0$, if there is a set of n independent and identically distributed random vectors $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ drawn from $\mathcal{N}_p(\boldsymbol{\mu}^*, \boldsymbol{\Sigma})$ satisfying

$$\big|\{i : \boldsymbol{X}_i \neq \boldsymbol{Y}_i\}\big| \leq \varepsilon n. \tag{4}$$

The chart below indicates how these contamination models are connected between each other:



§2. Robust mean estimation in high-dimensions.

Coordinate-wise median. The variational formulation of coordinate-wise median is the following

$$\hat{\boldsymbol{\mu}}_n^{\mathsf{CM}} \in \arg\min_{\boldsymbol{\mu} \in \mathbb{R}^p} \sum_{i=1}^n \|\boldsymbol{Y}_i - \boldsymbol{\mu}\|_1. \tag{5}$$

Theorem 1. *For any* $\mathsf{a} < 1/2$ *there exist two constants* c_1 *and* c_2 *depending only on* a *such that if* $\delta \in [2pe^{-n/c_1}, 1]$, *then*

$$\inf_{\varepsilon \le \mathsf{a}} \inf_{\mathsf{P}_n \in \mathsf{GAC}(\boldsymbol{\Sigma}, \varepsilon)} \mathbb{P}\left(\|\hat{\boldsymbol{\mu}}_n^{CM} - \boldsymbol{\mu}^*\|_2 \le c_2 \left(\varepsilon \sqrt{\mathsf{Tr}(\boldsymbol{\Sigma})} + \sqrt{\frac{\mathsf{Tr}(\boldsymbol{\Sigma}) \log(2p/\delta)}{2n}} \right) \right) \ge 1 - \delta.$$

Moreover,

$$\sup_{\mathsf{P}_n \in \mathsf{GAC}(\boldsymbol{\Sigma}, \varepsilon)} \mathbb{E}\big[\|\hat{\boldsymbol{\mu}}_n^{CM} - \boldsymbol{\mu}^*\|_2^4\big]^{1/4} \le c_2 \sqrt{\mathsf{Tr}(\boldsymbol{\Sigma})} \Big(\varepsilon + \frac{1}{\sqrt{n}} \Big) + \frac{4\sqrt{\mathsf{Tr}(\boldsymbol{\Sigma})}}{1 - 2\mathsf{a}} \, e^{-n/(8c_1)}$$

holds for any $\varepsilon \le \mathsf{a}$.

Tukey's median. For any $\eta \in \mathbb{R}^p$ and distribution $\mathbb{P}$ on $\mathbb{R}^p$, the Tukey's depth of η with respect to $\mathbb{P}$ is defined as follows

$$\mathcal{D}(\eta, \mathbb{P}) = \inf_{u \in \mathcal{S}^{p-1}} \mathbb{P}(u^\top \boldsymbol{X} \le u^\top \eta), \text{ where } \quad \boldsymbol{X} \sim \mathbb{P}. \tag{6}$$

The empirical counterpart of (6) for i.i.d. observations $\{\boldsymbol{X}_i\}_{i=1}^n$ is defined as follows:

$$\mathcal{D}(\eta, \{\boldsymbol{X}_i\}_{i=1}^n) = \min_{u \in \mathcal{S}^{p-1}} \frac{1}{n} \sum_{i=1}^n \mathbb{1}\big\{u^\top \boldsymbol{X}_i \le u^\top \eta\big\},$$

where $\mathbb{P}_n = \frac{1}{n} \sum_{i=1}^n \delta_{\boldsymbol{X}_i}$ being the empirical distribution. Then, Tukey's median is defined to be the deepest point with respect to given observations $\{\boldsymbol{X}_i\}_{i=1}^n$, hence

$$\widehat{\boldsymbol{\mu}}^{\mathsf{TM}} = \arg\max_{\eta \in \mathbb{R}^p} \mathcal{D}(\eta, \{\boldsymbol{X}_i\}_{i=1}^n). \tag{7}$$

In case of multiple maxima in (7) we take any vector that has the deepest Tukey depth $\mathcal{D}$.

Theorem 2 (Theorem 2.1 from [2]). *Assume* $\varepsilon < 1/5$ *and there exists universal constants* $C, C_1 > 0$ *such that for any* $\delta \in (0, 0.5)$ *satisfying* $C_1 \left(p/n + \log \delta^{-1}/n\right) < 1$, *Tukey's median* $\widehat{\boldsymbol{\mu}}^{TM}$ *has the following convergence rate*

$$\|\widehat{\boldsymbol{\mu}}^{TM} - \boldsymbol{\mu}\|_2^2 \le C \left(\frac{p}{n} \vee \varepsilon^2 + \frac{\log \delta^{-1}}{n} \right)$$

with probability of at least $1 - 2\delta$ *uniformly over* $\boldsymbol{\mu}$ *and* $\boldsymbol{Q}$.

Chapter 3.

§1. All inclusive robust estimator. In this part we construct an estimator that has many good properties for robust estimation. Below we list these properties.

1. The estimator is computationally tractable.
2. The estimator is translation and orthogonal transformation invariant.
3. The breakdown point ε_n^* and the nearly-minimax-rate-breakdown point $\varepsilon_{\mathsf{r}}^*$ of the estimator satisfy, respectively $\varepsilon_n^* = 0.5$ and $\varepsilon_{\mathsf{r}}^* \ge (5 - \sqrt{5})/10 \approx 0.28$.

4. The estimator is nearly minimax rate optimal, in the sense that its worst-case risk is of order $\|\mathbf{\Sigma}\|_{\mathsf{op}}^{1/2}\big(\sqrt{\mathbf{r}_{\mathbf{\Sigma}}/n}+\varepsilon\sqrt{\log(1/\varepsilon)}\big)$.

5. In the setting $\varepsilon=\varepsilon_n\to 0$ so that $\varepsilon^2\log(1/\varepsilon)=o(\mathbf{r}_{\mathbf{\Sigma}}/n)$ when $n\to\infty$, the estimator is asymptotically efficient.

The main theorem is formulated as follows

Theorem 3. *There is a universal constant $C>0$ such that for any $n,p\geq 1$ and for every $\varepsilon<(5-\sqrt{5})/10$, we have*

$$\sup_{\boldsymbol{\mu}^*\in\mathbb{R}^p}\sup_{\mathsf{P}_n\in\mathsf{GAC}(\boldsymbol{\mu}^*,\mathbf{\Sigma},\varepsilon)}\mathbb{E}^{1/2}[\|\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}-\boldsymbol{\mu}^*\|_2^2]\leq\frac{C\|\mathbf{\Sigma}\|_{\mathsf{op}}^{1/2}}{1-2\varepsilon-\sqrt{\varepsilon(1-\varepsilon)}}\big(\sqrt{\mathbf{r}_{\mathbf{\Sigma}}/n}+\varepsilon\sqrt{\log(1/\varepsilon)}\big).$$

If, in addition, $p\geq 2$ and $n\geq p\vee 10$, then

$$\sup_{\boldsymbol{\mu}^*\in\mathbb{R}^p}\sup_{\mathsf{P}_n\in\mathsf{GAC}(\boldsymbol{\mu}^*,\mathbf{\Sigma},\varepsilon)}\mathbb{E}^{1/2}[\|\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}-\boldsymbol{\mu}^*\|_2^2]\leq\frac{10\|\mathbf{\Sigma}\|_{\mathsf{op}}^{1/2}}{1-2\varepsilon-\sqrt{\varepsilon(1-\varepsilon)}}\big(\sqrt{p/n}+\varepsilon\sqrt{\log(1/\varepsilon)}\big).$$

§2. Excess risk setup.

Let $\hat{\boldsymbol{\mu}}^i=\mathtt{Med}(\boldsymbol{Y}_1,\ldots,\boldsymbol{Y}_{i-1},\boldsymbol{Y}_{i+1},\ldots,\boldsymbol{Y}_n)\equiv\mathtt{Med}(\mathcal{Y}_{-i})$ be the sample median of the sample $\mathcal{Y}$ excluding the observation Y_i for each $i\in\{1,\ldots,n\}$. The estimator $\hat{\boldsymbol{\mu}}^{\mathsf{GHT}}$ is then defined by the following simple procedure. Put

$$\hat{\theta}_i=\mathsf{HT}_\lambda(\boldsymbol{Y}_i-\hat{\boldsymbol{\mu}}^i)=(\boldsymbol{Y}_i-\hat{\boldsymbol{\mu}}^i)\mathbb{1}(\|\boldsymbol{Y}_i-\hat{\boldsymbol{\mu}}^i\|_2>\lambda)\tag{8}$$

and take

$$\hat{\boldsymbol{\mu}}^{\mathsf{GHT}}=\frac{1}{n}\sum_{i=1}^n(\boldsymbol{Y}_i-\hat{\theta}_i)=L_n(\mathbf{Y}-\hat{\mathbf{\Theta}})\tag{9}$$

as the final estimator.

Even in the more weaken setup when $\hat{\boldsymbol{\mu}}^i\equiv\hat{\boldsymbol{\mu}}=\mathtt{Med}(Y_1,Y_2,\ldots,Y_n)$ the following theorem takes place for the excess-risk loss

$$\mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}},n,p_n,\varepsilon_n)=\sup_{\boldsymbol{\mu}\in\mathbb{R}^p;\|\mathbf{\Theta}\|_{0,2}\leq\varepsilon n}R[\hat{\boldsymbol{\mu}},\boldsymbol{\mu};\mathbf{\Theta}]-\sqrt{\frac{p}{n}}$$

Theorem 4. *For $\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}$ and $\lambda^2=p+8\sqrt{p\log\varepsilon^{-1}}+16\log\varepsilon^{-1}$ we have*

$$\varlimsup_{n\to\infty}\mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}},n,p_n,\varepsilon_n)=0$$

provided that $\varepsilon_n p_n^{1/4}\log^{1/2}\varepsilon_n^{-1}=o(1)$ and $p_n=O(n)$ as $n\to\infty$.

For the minimax rate the estimator (9) enjoys the following bound:

Theorem 5. *Let $\varepsilon \in (0, \mathsf{a})$ with $\mathsf{a} < 1/2$ and $\mathcal{Y}$ be the sample generated according to parameter contamination model. For the estimator $\hat{\boldsymbol{\mu}}^{\mathsf{GHT}}$ and*

$$\lambda = \sqrt{2\log(2/\varepsilon^8)} + \sqrt{\mathsf{Tr}(\boldsymbol{\Sigma})} + c_2^2\sqrt{\mathsf{Tr}(\boldsymbol{\Sigma})}\left(\varepsilon^2 + \frac{\log(4p/\varepsilon^8)}{2(n-1)}\right),$$

where c_2 is a constant dependent only on a, then we have the following bound for the excess risk

$$\mathcal{E}(\hat{\boldsymbol{\mu}}^{\mathsf{GHT}}, \varepsilon) \leq \sqrt{\varepsilon\frac{\mathsf{Tr}(\boldsymbol{\Sigma})}{n}} + C\varepsilon \cdot \left[\varepsilon\sqrt{\mathsf{Tr}(\boldsymbol{\Sigma})} + \sqrt{\log(1/\varepsilon)} + \left(\|\boldsymbol{\Sigma}\|_F^{1/2} \vee [\log(1/\varepsilon)\mathsf{Tr}(\boldsymbol{\Sigma})]^{1/4}\right)\right]$$

for some universal constant C.

Chapter 4.

Chapter 4 addresses the problem of robust estimation from the optimizational point of view.

§1. Iteratively reweighted mean estimator.

Algorithm 1 Iteratively reweighted mean estimator

Input: data $\boldsymbol{X}_1, \ldots, \boldsymbol{X}_n \in \mathbb{R}^p$, contamination rate ε and $\boldsymbol{\Sigma}$
Output: parameter estimate $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$
Initialize: compute $\boldsymbol{\mu}^0$ as a minimizer of $\sum_{i=1}^n \|\boldsymbol{X}_i - \boldsymbol{\mu}\|_2$
Set $K = 0 \vee \left\lceil \frac{\log(4\mathbf{r}_{\boldsymbol{\Sigma}}) - 2\log(\varepsilon(1-2\varepsilon))}{2\log(1-2\varepsilon) - \log\varepsilon - \log(1-\varepsilon)} \right\rceil$.
For $k = 1 : K$
  For $i = 1 : n$
    Set $\mathsf{M}_i = (\boldsymbol{X}_i - \boldsymbol{\mu}^{k-1})(\boldsymbol{X}_i - \boldsymbol{\mu}^{k-1})^\top$
  EndFor
  Compute current weights:

$$\boldsymbol{w} \in \underset{(n-n\varepsilon)\|\boldsymbol{w}\|_\infty \leq 1}{\arg\min} \lambda_{\max}\left(\sum_{i=1}^n w_i \mathsf{M}_i - \boldsymbol{\Sigma}\right) \vee 0.$$

  Update the estimator: $\hat{\boldsymbol{\mu}}^k = \sum_{i=1}^n w_i \boldsymbol{X}_i$.
EndFor
Return $\hat{\boldsymbol{\mu}}^K$.

§2. Minimizing Huber function. From the intuitive point of view the algorithm from [4] approximates $\widehat{\boldsymbol{\mu}}^\star$ defined as follows

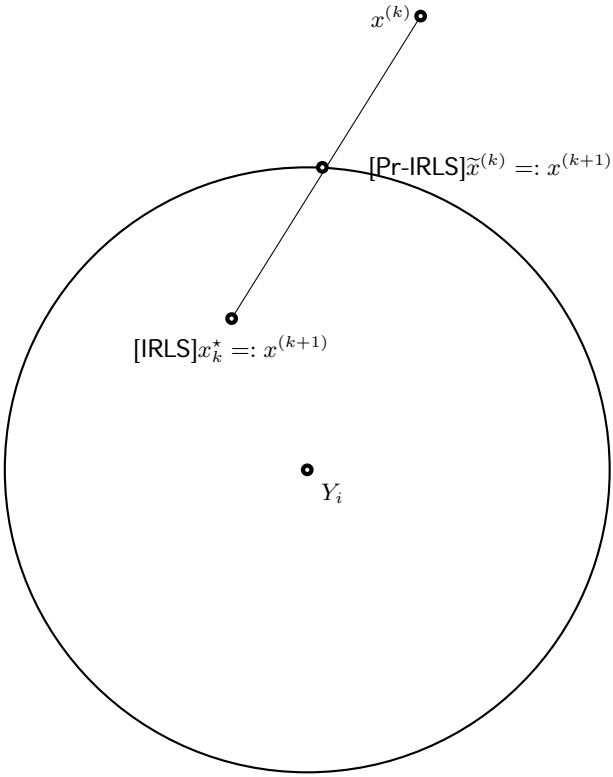
$$\widehat{\boldsymbol{\mu}}^\star \in \arg\min_{\boldsymbol{\mu} \in \mathbf{R}^p} \sum_{i=1}^n \rho_\delta\left(\frac{(\|Y_i - \boldsymbol{\mu}\|_2^2 - \sigma^2 p)_+^{1/2}}{\sigma}\right) \tag{10}$$

for appropriately chosen tuning parameter $\delta > 0$. The objective function from the optimization problem (10) is unfortunately not convex with respect to $\boldsymbol{\mu}$.

§3. Projected IRLS. The algorithm of IRLS was first introduced in [17]. A modification of this algorithm was studied and an illustration of one Pr-IRLS algorithm step from $x^{(k)}$ to $x^{(k+1)}$ given that $x^{(k)}$ is outside of the ball around Y_i and its IRLS projection is inside.

Algorithm 2 IRLS for mean estimation

Given: data $Y_1, \ldots, Y_n \in \mathbf{R}^p$ and δ.
Initialize: $\boldsymbol{\mu}^{(0)} = \frac{1}{n}\sum_{i=1}^n Y_i$.
$k \leftarrow 0$
while stopping criterion is **False**, **do**
  $k \leftarrow k+1$
  $t_k = [(\|\sigma^{-1}(Y_i - \boldsymbol{\mu}^{(k-1)})\|_2^2 - p)_+^{1/2}$ for $i = 1, \ldots, n]$
  $\mathcal{O} = \{i : t_k^i > \delta\}$.
  Update: $\boldsymbol{\mu}^{(k)} = \frac{\sum_{i\in\mathcal{O}} Y_i \cdot \delta/t_k^{(i)} + \sum_{i\in\mathcal{O}^c} Y_i/\sigma}{\sum_{i\in\mathcal{O}} \delta/t_k^{(i)} + |\mathcal{O}^c|/\sigma}$.
  Check stopping criterion.
end while
$N \leftarrow k$ and $\widehat{\boldsymbol{\mu}} \leftarrow \boldsymbol{\mu}^{(N)}$
Return: $\widehat{\boldsymbol{\mu}}$.



Theorem 6. *For all but countable set of initial values $x^{(0)}$ and for all $i \in [N]$ if at each iteration $k \geq 1$ $x^{(k)} \neq Y_i$ then the above defined sequence $\{x^{(k)}\}_{k\geq 1}$ converges to $x^\star$.*

Chapter 5.

Chapter 5 concludes the main results of the thesis and contains discussion on the relevance of obtained results with possible further research directions.

List of author's publications

1. Minasyan, A. Alternating least squares in generalized linear models. *Journal of Contemporary Mathematical Analysis (Armenian Academy of Sciences)*, 54(5):302–312, 2019.
2. Minasyan, A. Excess-risk consistency of group-hard thresholding estimator in robust estimation of gaussian mean. *Journal of Contemporary Mathematical Analysis (Armenian Academy of Sciences)*, 55(2):208–212, 2020.
3. Minasyan, A. On the convergence and robustness of mean-vector estimation of inliers distribution. In *Proceedings of 21st European Young Statistiticians Meeting,* 62–67, 2019.
4. Dalalyan, A. and Minasyan, A. All-in-one robust estimator of the gaussian mean. *arXiv preprint arXiv:2002.01432,* 2020.
5. Minasyan, A. Robust Estimation through Huber Function. *The 12th International Conference on Numerical Optimization and Numerical Linear Algebra*, p. 29, 2019.
6. Minasyan, A. Assessing excess risk in robust estimation of gaussian mean with parameter contamination. *International Conference Dedicated to the 100th Anniversary of Yerevan State University*, p. 37, 2019.

References

[1] Amir-Hossein Bateni and Arnak S. Dalalyan. Confidence regions and minimax rates in outlier-robust estimation on the probability simplex. *Electron. J. Statist.*, 14(2):2653–2677, 2020.

[2] Mengjie Chen, Chao Gao, and Zhao Ren. Robust covariance and scatter matrix estimation under huber's contamination model. *Ann. Statist.*, 46(5):1932–1960, 10 2018.

[3] Yu Cheng, Ilias Diakonikolas, and Rong Ge. High-dimensional robust mean estimation in nearly-linear time. *Proceedings of the Thirtieth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2755–2771, 2019.

[4] Olivier Collier and Arnak S. Dalalyan. Multidimensional linear functional estimation in sparse gaussian models and robust estimation of the mean. *Electron. J. Statist.*, 13(2):2830–2864, 2019.

[5] Luc Devroye and Gábor Lugosi. *Combinatorial methods in density estimation*. Springer Science & Business Media, 2012.

[6] Ilias Diakonikolas, Gautam Kamath, Daniel Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robust estimators in high-dimensions without the computational intractability. *SIAM Journal on Computing*, 48(2):742–864, 2019.

[7] Ilias Diakonikolas, Gautam Kamath, Daniel M Kane, Jerry Li, Ankur Moitra, and Alistair Stewart. Robustly learning a gaussian: Getting optimal error, efficiently. *Proceedings of the Twenty-Ninth Annual ACM-SIAM Symposium on Discrete Algorithms*, pages 2683–2702, 2018.

[8] Ilias Diakonikolas and Daniel M Kane. Recent advances in algorithmic high-dimensional robust statistics. *arXiv preprint arXiv:1911.05911*, 2019.

[9] Ilias Diakonikolas, Daniel M. Kane, and Alistair Stewart. Statistical query lower bounds for robust estimation of high-dimensional gaussians and gaussian mixtures. *2017 IEEE 58th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 73–84, 2017.

[10] Peter J. Huber. Robust estimation of a location parameter. *Ann. Math. Statist.*, 35(1):73–101, 03 1964.

[11] Peter J. Huber. A robust version of the probability ratio test. *Ann. Math. Statist.*, 36(6):1753–1758, 12 1965.

[12] K. A. Lai, A. B. Rao, and S. Vempala. Agnostic estimation of mean and covariance. *2016 IEEE 57th Annual Symposium on Foundations of Computer Science (FOCS)*, pages 665–674, 2016.

[13] OV Lepskii. Asymptotically minimax adaptive estimation. i: Upper bounds. optimally adaptive estimates. *Theory of Probability & Its Applications*, 36(4):682–697, 1992.

[14] Stanislav Minsker et al. Geometric median and robust estimation in banach spaces. *Bernoulli*, 21(4):2308–2335, 2015.

[15] Leonard A Stefanski and Dennis D Boos. The calculus of m-estimation. *The American Statistician*, 56(1):29–38, 2002.

[16] John W Tukey. Mathematics and the picturing of data. In *Proceedings of the International Congress of Mathematicians, Vancouver, 1975*, volume 2, pages 523–531, 1975.

[17] Endre Weiszfeld and Frank Plastria. On the point for which the sum of the distances to n given points is minimum. *Annals of Operations Research*, 167(1):7–41, 2009.

ԱՄՓՈՓՈՒՄ

Արշակ Գագիկի Մինասյան

Ռոբաստ գնահատում գաուսյան միջինի համար հաշվարկելիության տիրույթում

Ատենախոսությունում ստացվել են հետևյալ արդյունքները.

1. Ստացվել են $\boldsymbol{Y}_1, \ldots, \boldsymbol{Y}_n$ տվյալների կոորդինատային կիսորդի $\hat{\boldsymbol{\mu}} = \mathtt{Med}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$ ℓ_2 կորստի ֆունկցիայի համար վերին սահմաններ մեծ հավանականությամբ և մաթեմատիկական սպասումով:

2. Կառուցված է այնպիսի գնահատական $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$, որը ինվարիանտ է շեղման և օրթոգոնալ տրանսֆորմացիաների նկատմամբ:

3. Հաշվված է $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ գնահատականի ձախողման սահմանը:

4. Ապացուցված է, որ $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ գնահատականի մինիմաքս սխալանքը օպտիմալ է (լոգարիթմիկ ֆակտորի ճշտությամբ), այսինքն ռիսկը վատագույն դեպքում ունի հետևյալ կարգը

$$\|\boldsymbol{\Sigma}\|_{\mathsf{op}}^{1/2}\big(\sqrt{\mathbf{r}_{\boldsymbol{\Sigma}}/n} + \varepsilon\sqrt{\log(1/\varepsilon)}\big),$$

   որտեղ $\mathbf{r}_{\boldsymbol{\Sigma}} = \frac{\mathsf{Tr}(\boldsymbol{\Sigma})}{\|\boldsymbol{\Sigma}\|_{\mathsf{op}}}$-ը $\boldsymbol{\Sigma}$ մատրիցի էֆեկտիվ ռանկն է:

5. Ապացուցված է, որ $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ գնահատականը ասիմպտոտիկ էֆեկտիվ է:

6. Դիտարկված է ռոբաստ գնահատումը հավելյալ ռիսկի կորստի ֆունկցիայի դեպքում, և կառուցված է $\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}$ գնահատական, որի համար ճիշտ է հետևյալ արդյունքը

$$\overline{\lim_{n\to\infty}}\, \mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}, n, p_n, \varepsilon_n) = 0.$$

7. Ձևափոխվել է $\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}$ գնահատականը, այնպես որ ստացվել է հետևյալ մինիմաքս սխալանքը հավելյալ ռիսկի կորստի ֆունկցիայի դեպքում

$$\mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}, \varepsilon) \le \sqrt{\varepsilon\frac{p}{n}} + C\varepsilon^2\sqrt{p}.$$

8. Հաշվարկելիորեն ցույց է տրված, որ ալգորիթմ 1 և 2-ը օժտված են վերոնշյալ ռոբաստ հատկություններով:

9. Հաշվարկելիորեն ցույց է տրված, որ պրոյեկտած իտերատիվ վերակշռված փոքրագույն քառակուսիների մեթոդը նույնպես օժտված է ռոբաստ հատկություններով և ապացուցված է ալգորիթմի զուգամիտությունը:

РЕЗЮМЕ

Минасян Аршак Гагикович

Робастное оценивание гауссовского среднего в области вычислительной эффективности

В диссертации получены следующие результаты:

1. Получены верхние оценки для ℓ_2 функции потерь покоординатной медианы $\hat{\boldsymbol{\mu}} = \mathtt{Med}(\boldsymbol{Y}_1, \boldsymbol{Y}_2, \ldots, \boldsymbol{Y}_n)$ с большой вероятностью и по математическому ожиданию.

2. Построена оценка $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$, которая инвариантна по отношению к сдвигам и ортогональным трансформациям.

3. Посчитана точка невозврата для оценки $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$.

4. Доказано, что оценка $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ обладает почти оптимальным минимаксным порядком, в худшем случае функция потерь имеет порядок

$$\|\boldsymbol{\Sigma}\|_{\mathsf{op}}^{1/2}\big(\sqrt{\mathbf{r}_{\boldsymbol{\Sigma}}/n} + \varepsilon\sqrt{\log(1/\varepsilon)}\big),$$

   где $\mathbf{r}_{\boldsymbol{\Sigma}} = \frac{\mathsf{Tr}(\boldsymbol{\Sigma})}{\|\boldsymbol{\Sigma}\|_{\mathsf{op}}}$ эффективный ранк матрицы $\boldsymbol{\Sigma}$.

5. Доказано, что оценка $\hat{\boldsymbol{\mu}}_n^{\mathsf{IR}}$ асимптотически эффективна.

6. Изучены робастные оценки в случае избыточного риска и построена оценка $\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}$ такая, что

$$\overline{\lim_{n\to\infty}}\, \mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}, n, p_n, \varepsilon_n) = 0.$$

7. Модифицирована оценка $\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}$ и получен ксный порядок в случае избыточного риска

$$\mathcal{E}(\hat{\boldsymbol{\mu}}_{\mathsf{GHT}}, \varepsilon) \le \sqrt{\varepsilon\frac{p}{n}} + C\varepsilon^2\sqrt{p}.$$

8. Иллюстрирована робастность Алгоритмов 1 и 2.

9. Иллюстрирована робастность проектированного итеративного взвешенного метода наименьших квадратов и доказана теорема о сходимости метода.