National Research University Higher School of Economics

as a manuscript

Bulgakov Stanislav Aleksandrovich

Stochastic recovery of square-integrable functions

DISSERTATION SUMMARY

for the purpose of obtaining academic degree

Doctor of Philosophy in Applied Mathematics

Academic supervisor:

Doctor of physical and mathematical Sciences, Professor

Khametov Vladimir Minirovich

Moscow 2024

Statement of the stochastic recovery problem. This dissertation is devoted to the theory of stochastic recovery. The stochastic recovery is usually understood as follows:

i) there is an unknown scalar function with values observed with random errors;

ii) the problem is to construct the best estimator of this function in terms of some optimality criterion.

The dissertation considers the stochastic recovery problem under following conditions:

i) the unknown scalar function is defined on the finite-dimensional compact set and is square integrable in the sense of the Lebesgue measure;

ii) the errors are the family of random scalar functions defined on the same finite-dimensional compact set as the estimated function;

iii) the observations are the family of scalar functions where each element is the additive mixture of this unknown function with error function, and it is possible to make an observation at every point of aforementioned compact set;

iv) the optimality criterion is the minimum standard deviation of the estimator for the unknown function.

The degree of the problem's development. Such estimation procedure is known to be an infinite-dimensional one, i.e. nonparametric. The theory of nonparametric estimation considers problems of two types.

The first type of problem is to construct the estimator of an unknown distribution density from observations for corresponding random sequence of functions, delivering an extremum to performance criterion. There are exist three approaches to solve this problem.

The first approach is based on Kolmogorov's theory of n-widths and the Glivenko-Cantelli theorem. Corresponding results can be found in the following papers Vapnik et al. [1], Vapnik [2], Ibragimov [3]. It should be noted that in the work by Vapnik et al. [1] algorithms and programs implementing the recovery of various dependencies are provided.

The second approach is to construct the maximum likelihood kernel estimators of smooth unknown functions. There are quite a lot of works devoted to its description and justification, here are the main ones: Parzen [4], Rosenblatt [5], Murthy [6], Watson [7], Konakov [8], Nadaraya [9] etc. They give an explicit form of these estimators, describe their statistical properties and find the rate of convergence to an unknown function.

The third approach is based on using projection estimators in the maximum likelihood method. Such approach turns original nonparametric estimation problem to infinite-dimensional linear estimation problem. The performance is obtained in terms of the minimum standard deviation. Works devoted to this approach of constructing estimators for smooth unknown functions are: Chentsov [10], Tsybakov [11], Juditsky and Nemirovski [12]. They give a description of almost optimal finite-dimensional recovery procedures and their properties.

Now describe the second type of recovery problem of an unknown function from some class which is observed with random errors at points from an unknown function's domain. Based on these observations, it is necessary to construct nonparametric estimator of this function optimal in terms of the minimum standard deviation. Works devoted to solving such problems are: Stratonovich [13], Ibragimov and Has'minskii [14], Golubev [15], Darkhovskii [16], Nemirovski at el. [17]. They give a description of optimal and almost optimal recovery as well as statistical and some other properties of these algorithms.

Goals and objectives of the research. The goal of the work is to develop and justify stochastic recovery procedures. Its objectives are:

i) to establish the existence of such procedures and form of the optimal stochastic recovery as well as their performance;

ii) to establish conditions under which there exist the ε-optimal stochastic recovery and its form.

Relevance. Stochastic recovery problem is well known to belong to (N, p)-hard nonparametric estimation. Therefore developing estimation procedures with polynomial complexity is a topical problem. For this reason, the dissertation pays great attention to creation of finite-dimensional linear recovery procedures such as ε-optimal and equivalent to them by the order of magnitude.

Methodology of research. The results of the dissertation are largely based on mathematical methods of:

i) probability theory;
ii) random sequence statistics;
iii) stochastic analysis;
iv) Fourier series theory;
v) optimal stochastic systems theory.

Principal results to be defended.

1) The observation scheme allowing to construct the optimal projection estimator for an unknown function is proposed and conditions under which such estimators are unbiased and consistent are established (chapter 1).
2) The dependency of the optimal projection estimator's standard deviation from the number of orthogonal functions and number of observations is established. The recurrence-based relations satisfied by this standard deviation are found (chapter 1).
3) The definition of the Chentsov projection estimator (CPE) is given, and criterion (necessary and sufficient condition) under which exist the minimum number of orthogonal functions $N^0(m)$ such what projection estimator for the unknown function is optimal in terms of the minimum standard deviation for any fixed number of observations m. The algorithm of finding $N^0(m)$ is established (chapter 2).
4) The following CPE's standard deviation properties are established:
   i) order of magnitude $V_m(N^0(m))$ (see formula (2.9) of theorem 2.2.1) (chapter 2);
   ii) independency of CPE's standard deviation upper bound from an unknown function $f(x) \in L_2(\mathrm{K}, \Lambda)$ (chapter 2).

5) The definition of ε-optimal projection estimator is given, its existence criterion and sufficient conditions for its ε-optimality are proved (chapter 3).

Scientific novelty. All results to be defended are new; i.e. they have no analogues in both Russian and foreign scientific papers.

General conclusions. The dissertation proposes a model for observations of an unknown deterministic scalar square integrable function defined on the finite-dimensional compact set such that its values are measured with independent Gaussian errors at every point of its domain.

Conditions under which there exist an optimal and ε-optimal (in terms of the minimum standard deviation) projection estimators as well as Chentsov's estimators are proved for such model of observations. Conditions of unbiasedness and consistency for such estimators are found.

Contribution of the author. All the problems were stated by professor Khametov V.M. All results to be defended were proved by the author.

List of scientific papers that reflect the main results of the dissertation. The main scientific results of the dissertation research are presented in the following publications.

[1] Bulgakov S. A., Gorshkova V. M., Khametov V. M. Stochastic recovery of square-integrable functions // *Herald of the Bauman Moscow State Technical University, Series Natural Sciences.* — 2020. — no. 6. — P. 4–22. — [in Russian].

[2] Bulgakov S. A., Khametov V. M. Modelling optimal and ε-optimal recovery algorithms of square-integrable function from observations with Gaussian errors // *Nanostructures. Mathematical physics and modelling.* — 2020. — Vol. 20, no. 1. — P. 57–69. — [in Russian].

[3] Bulgakov S. A., Khametov V. M. Optimal Recovery of a Square Integrable Function from Its Observations with Gaussian Errors // *Automation and Remote Control.* — 2023. — Vol. 84, no. 4. — P. 412–433.

[4] Bulgakov S. A. ε-optimal recovery algorithm for square-integrable functions with finite-dimensional compact set domain // Modern problems in mathematics and its applications: proceedings of International (49-th National) Youth School-Conference. — Yekaterinburg : N.N. Krasovsky Institute of Mathematics and Mechanics of the UB RAS. — 2018. — P. 61–61. — [in Russian].

[5] Bulgakov S. A. About some approaches to the image restoration // Modern problems of mathematical modelling, data image processing and parallel computing. — Rostov-on-Don. — 2017. — P. 36–44. — [in Russian].

[6] Bulgakov S. A., Khametov V. M. Recovery of square-integrable function from observations with Gaussian errors // *Large-scale Systems Control*. — 2015. — Vol. 54. — P. 45–65. — [in Russian].

[7] Bulgakov S. A. Optimal stochastic recovery of square-integrable function from observations with Gaussian errors // Scientific-technical conference of students, graduate students and young professionals of E.V. Armensky. Conference papers. — Moscow : MIEM NRU HSE. — 2015. — [in Russian].

The work has been tested at the following international and national conferences.

[1] Modern problems in mathematics and its applications International (49-th National) Youth School-Conference. — Yekaterinburg. — 2018. — [in Russian].

[2] Modern problems of mathematical modelling, data image processing and parallel computing. — Rostov-on-Don. — 2017. — [in Russian].

[3] Research seminar of the Doctoral School of Computer Science. — Moscow : Faculty of Computer Science NRU HSE. — 2015. — [in Russian].

[4] Scientific-technical conference of students, graduate students and young professionals of E.V. Armensky. — Moscow. — 2015. — [in Russian].

References

[1] Algoritmy i programmy vosstanovleniya zavisimostey / ed. by Vapnik V. N. — Moscow : Nauka Publ., 1984. — [in Russian].

[2] Vapnik V. N. Estimation of Dependences Based on Empirical Data. — New York : Springer, 2010.

[3] Ibragimov I. A. Estimation of Multivariate Regression // *Theory of Probability & Its Applications*. — 2004. — Vol. 48, no. 2. — P. 256–272.

[4] Parzen E. On Estimation of a Probability Density Function and Mode // *Ann. Math. Statist.* — 1962. — Vol. 33, no. 3. — P. 1065–1076.

[5] Rosenblatt M. Curve Estimates // *Ann. Math. Statist.* — 1971. — Vol. 42, no. 6. — P. 1815–1842.

[6] Murthy V. K. Nonparametric estimation of multivariate densities with applications // *Multivariate Analysis*. — 1966. — P. 43–56.

[7] Watson G. S. Density Estimation by Orthogonal Series // *Ann. Math. Statist.* — 1969. — Vol. 40, no. 4. — P. 1496–1498.

[8] Konakov V. D. Non-Parametric Estimation of Density Functions // *Theory of Probability & Its Applications*. — 1973. — Vol. 17, no. 2. — P. 361–362.

[9] Nadaraya E. A. Nonparametric Estimation of Probability Densities and Regression Curves. — Dordrecht : Springer, 1988.

[10] Chentsov N. N. Statistical Decision Rules and Optimal Inference. — American Mathematical Society, 1982.

[11] Tsybakov A. B. Introduction to Nonparametric Estimation. — New York : Springer, 2009. — 214 p.

[12] Juditsky A. B., Nemirovski A. S. Signal Recovery by Stochastic Optimization // *Automation and Remote Control*. — 2019. — no. 10. — P. 1878–1893.

[13] Stratonovich R. L. The Efficiency of Mathematical Statistics Methods in the Design of Algorithms to Recover an Unknown Function // *Izv. Akad. Nauk SSSR. Tekh. Kibern.* — 1969. — no. 1. — P. 32–46.

[14] Ibragimov I. A., Has'minskii R. Z. Statistical Estimation: Asymptotic Theory. — New York : Springer, 1981.

[15] Golubev G. K. On Adaptive Estimation of Linear Functionals from Observations against White Noise // *Problems of Information Transmission*. — 2020. — Vol. 56, no. 2. — P. 185–200.

[16] Darkhovskii B. S. On a Stochastic Renewal Problem // *Theory of Probability & Its Applications*. — 1999. — Vol. 43, no. 2. — P. 282–288.

[17] Nemirovskij A. S., Polyak B. T., Tsybakov A. B. Signal Processing by the Nonparametric Maximum-Likelihood Method // *Problems of Information Transmission*. — 1984. — Vol. 20, no. 3. — P. 177–192.