Федеральное государственное автономное образовательное учреждение
высшего образования «Московский физико-технический институт
(национальный исследовательский университет)»
Физтех-школа прикладной математики и информатики
Кафедра математических основ управления

На правах рукописи

Алкуса Мохаммад

Алкуса Мохаммад

Численные методы решения негладких задач выпуклой оптимизации с функциональными ограничениями

Специальность 01.01.07 — Вычислительная математика

АВТОРЕФЕРАТ
диссертации на соискание ученой степени
кандидата физико-математических наук

Научный руководитель
д.ф.-м.н. А. В. Гасников

Москва — 2020

Работа выполнена на кафедре математических основ управления Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» Физтех-школа прикладной математики и информатики.

Научный руководитель: Доктор физико-математических наук,
Гасников Александр Владимирович

Ведущая организация: Федеральное государственное автономное образовательное учреждение высшего образования «Крымский федеральный университет имени В. И. Вернадского»

Защита состоится «20» мая 2020 г. в 14:00 на заседании диссертационно- го совета ФПМИ. 01.01.07.002 по адресу: 141701, Московская область, г. Долгопрудный, Институтский переулок, д. 9, МФТИ.

С диссертацией можно ознакомиться в библиотеке и на сайте Федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт (национальный исследовательский университет)» `https://mipt.ru/education/post-graduate/soiskateli-fiziko-matematicheskie-nauki.php`

Работа представлена «15» февраля 2020 г. в Аттестационную комиссию федерального государственного автономного образовательного учреждения высшего образования «Московский физико-технический институт» (национальный исследовательский университет) для рассмотрения советом по защите диссертаций на соискание учёной степени кандидата наук в соответствии с п. 3.1 ст. 4 Федерального закона «О науке и государственной научно-технической политике».

Federal State Autonomous Educational Institution of
Higher Education «Moscow Institute of Physics and Technology
(National Research University)»
Phystech School of Applied Mathematics and Informatics
Department of Control and Applied Mathematics

Manuscript

Mohammad Alkousa

Mohammad S. Alkousa

Numerical Methods for Non-Smooth Convex Optimization Problems with Functional Constraints

Specialty 01.01.07 — Computational mathematics

Synopsis
dissertation for the degree of
candidate of physical and mathematical sciences

Scientific supervisor
Sc.D. math. A. V. Gasnikov

Moscow — 2020

The work has been performed at the Department of Control and Applied Mathematics Federal State Autonomous Educational Institution of Higher Education «Moscow Institute of Physics and Technology (National Research University)» Phystech School of Applied Mathematics and Informatics. Department of Control and Applied Mathematics

Scientific supervisor: Doctor of Physical and Mathematical Sciences,
Alexander V. Gasnikov

Lead organization: Federal State Autonomous Educational Institution of Higher Education «V. I. Vernadsky Crimean Federal University»

The defense of the dissertation will be held on May, 20, 2020 at 14:00, at the meeting of the dissertation council FPMI. 01.01.07.002, at 141701, Moscow region, Dolgoprudniy, Institutskiy per., 9, MIPT.

The dissertation is stored in the library and published on the site of Moscow Institute of Physics and Technology (national research university). `https://mipt.ru/education/post-graduate/soiskateli-fiziko-matematicheskie-nauki.php`

The work was submitted on February, 15, 2020 to the Attestation Commission of the Moscow Institute of Physics and Technology (National Research University) for consideration by the council for the defense of dissertations for the degree of candidate of science, doctor of science in accordance with paragraph 3.1 Art. 4 of the Federal Law «On Science and State scientific and technical policy».

General description of the subject of work

The dissertation is devoted to the development of algorithms for non-smooth convex optimization problems with several convex non-smooth functional constraints.

Optimization problems arise naturally in many different fields, but unfortunately for a majority of optimization problems, there is no hope to find a solution analytically (i.e. find an explicit-form to an optimal solution). Therefore in order to solve an optimization problem, we have to use numerical methods. There are different classifications and types of these methods, one of them is first-order methods, which go back to 1847 with the work of Cauchy on the steepest descent method. With the increase in the amount of applications that can be modeled as large- or even huge-scale optimization problems (some of such applications arising in: machine learning, deep learning, data science, control, signal processing, statistics, and so on), first-order methods, which require low iteration cost as well as low memory storage, have received much interest over the past few decades in order to solve the convex optimization problems, in the smooth and non-smooth cases[1]. Especially, the optimization of non-smooth functions with functional constraints attracts widespread interest, in large-scale optimization and its applications[2] [3]. On continuous optimization with functional constraints, there is a long history of studies. In this area, there are some monographs[4] [5], many recent works on first-order methods for convex optimization problems with convex functional constraints for the deterministic setting[6] [7] [8] and for the stochastic setting[9] [10] [11]. However, the parallel development for problems with non-convex objective functions and also with non-convex constraints, especially for theoretically provable algorithms,

[1] A. Beck: First-Order Methods in Optimization. Society for Industrial and Applied Mathematics, 2017.

[2] A. Ben-Tal, A. Nemirovski: Robust Truss Topology Design via semidefinite programming. SIAM Journal on Optimization,**7**(4), pp. 991–1016, 1997.

[3] S. Shpirko, Y. Nesterov: Primal-dual subgradient methods for huge-scale linear conic problem. SIAM Journal on Optimization, **24**(3), pp. 1444–1457, 2014.

[4] D. P. Bertsekas: Constrained optimization and Lagrange multiplier methods. Academic press, 2014.

[5] J. Nocedal, S. J. Wright: Numerical Optimization. Springer, New York, 2006.

[6] O. Fercoq, A. Alacaoglu, I. Necoara, V. Cevher: Almost surely constrained convex optimization. Proceedings of the 36th International Conference on Machine Learning, PMLR 97, pp. 1910–1919, 2019.

[7] Q. Lin, R. Ma, Y. Xu: Inexact Proximal-Point Penalty Methods for Non-Convex Optimization with Non-Convex Constraints. 2019. `https://arxiv.org/pdf/1908.11518.pdf`

[8] Y. Xu: Iteration complexity of inexact augmented lagrangian methods for constrained convex programming. Mathematical Programming, Series A, pp. 1–46, 2019.

[9] K. Basu, P. Nandy: Optimal Convergence for Stochastic Optimization with Multiple Expectation Constraints. 2019. `https://arxiv.org/pdf/1906.03401.pdf`

[10] G. Lan, Z. Zhou: Algorithms for stochastic optimization with functional or expectation constraints. 2019. `https://arxiv.org/pdf/1604.03887.pdf`

[11] Y. Xu: Primal-dual stochastic gradient method for convex programs with many functional constraints. 2019. `https://arxiv.org/pdf/1802.02724.pdf`

remains limited[12]. There are various first-order methods, for solving the convex optimization problems in the case of non-smooth objective function. Among them, Mirror Descent method, which was originated in the works of Nemirovski and Yudin more 30 years ago[13] [14], and was later analyzed in 2003[15]. This method is considered as the non-Euclidean extension of subgradient methods, and used in many applications[16] [17] [18]. The standard subgradient methods employ the Euclidean distance function with a suitable step-size in the projection step. Mirror Descent extends the standard projected subgradient methods by employing a nonlinear distance function with an optimal step-size in the nonlinear projection step[19]. Mirror Descent method not only generalizes the standard subgradient descent method, but also achieves a better convergence rate and it is applicable to optimization problems in Banach spaces where subgradient descent is not[20].Also, in some works[21] [22], it was proposed an extension of the Mirror Descent method for constrained problems. Usually, the step-size and stopping rule for Mirror Descent method require to know the Lipschitz constant of the objective function and constraints, if any. Adaptive step-sizes, which do not require this information, are considered for unconstrained problems[23], and for constrained problems[24]. Recently some adaptive optimal Mirror Descent methods were proposed for convex optimization problems with non-smooth functional constraints, in deterministic and stochastic settings[25]. In order to solve

[12]see "Q. Lin, R. Ma, Y. Xu: Inexact Proximal-Point Penalty Methods for Non-Convex Optimization with Non-Convex Constraints. 2019. `https://arxiv.org/pdf/1908.11518.pdf`" and references therein.

[13]A. Nemirovskii: Efficient methods for large-scale convex optimization problems. Ekonomika i Matematicheskie Metody, 1979. (in Russian)

[14]A. Nemirovsky, D. Yudin: Problem Complexity and Method Efficiency in Optimization. J. Wiley & Sons, New York 1983.

[15]A. Beck, M. Teboulle: Mirror descent and nonlinear projected subgradient methods for convex optimization. Oper. Res. Lett., **31**(3), pp. 167–175, 2003.

[16]A. V. Nazin, B. M. Miller: Mirror Descent Algorithm for Homogeneous Finite Controlled Markov Chains with Unknown Mean Losses. Proceedings of the 18th World Congress The International Federation of Automatic Control Milano (Italy) August 28 - September 2, 2011.

[17]A. Nazin, S. Anulova, A. Tremba: Application of the Mirror Descent Method to Minimize Average Losses Coming by a Poisson Flow. European Control Conference (ECC) June 24-27, 2014.

[18]A. Tremba, A. Nazin: Extension of a saddle point mirror descent algorithm with application to robust PageRank. 52nd IEEE Conference on Decision and Control December 10-13, 2013.

[19]D. V. N. Luong, P. Parpas, D. Rueckert, B. Rustem: A Weighted Mirror Descent Algorithm for Nonsmooth Convex Optimization Problem. J Optim Theory Appl **170**(3), pp. 900–915, 2016.

[20]T. T. Doan, S. Bose, D. H. Nguyen, C. L. Beck: Convergence of the Iterates in Mirror Descent Methods. IEEE Control Systems Letters, **3**(1), pp. 114–119, 2019.

[21]A. Beck, A. Ben-Tal, N. Guttmann-Beck, L. Tetruashvili: The comirror algorithm for solving nonsmooth constrained convex problems. Operations Research Letters, **38**(6), pp. 493–498, 2010.

[22]A. Nemirovsky, D. Yudin: Problem Complexity and Method Efficiency in Optimization. J. Wiley & Sons, New York 1983.

[23]A. Ben-Tal, A. Nemirovski: Lectures on Modern Convex Optimization. Society for Industrial and Applied Mathematics, Philadelphia 2001.

[24]A. Beck, A. Ben-Tal, N. Guttmann-Beck, L. Tetruashvili: The comirror algorithm for solving nonsmooth constrained convex problems. Operations Research Letters, **38**(6), pp. 493–498, 2010.

[25]A. Bayandina, P. Dvurechensky, A. Gasnikov, F. Stonyakin, A. Titov: Mirror descent and convex optimization

the optimization problem in the stochastic setting, the Mirror Descent method also has been widely used[26] [27] [28].

Also in recent years, online convex optimization (OCO) has become a leading online learning framework, via its powerful modeling capability for a lot of problems from diverse domains. OCO plays a key role in solving the problems where statistical information is being updated[29] [30]. There are a lot of examples of such problems, concerning Internet network, consumer data sets or financial market, and in machine learning applications such as adaptive routing in networks, online display advertising[31], dictionary learning, classification, and regression[32]. In recent years, methods for solving online optimization problems have been actively developed, in both deterministic and stochastic settings[33] [34] [35]. Also, some adaptive methods were considered for OCO problem with constraints, but with only a standard Euclidean prox-structure[36]. Some algorithms were proposed for OCO with stochastic constraints, where the objective function varies arbitrarily but the functionals constraints are varying independently and identically distributed (i.i.d.) over time[37].

For the display to be complete, it is useful to take into account the saddle point problems, which are closely related to the optimization problems. Where in the general case we can consider the saddle point problem as a non-smooth convex optimization problem with a certain structure. Moreover, saddle point problems

problems with non-smooth inequality constraints. In: Large-Scale and Distributed Optimization, Springer, Cham pp. 181–213, 2018.

[26]G. Lan, A. Nemirovski, A. Shapiro: Validation analysis of mirror descent stochastic approximation method. Math. Program., **134**(2), pp. 425–458, 2012.

[27]A. V. Nazin, A. S. Nemirovsky, A. B. Tsybakov, A. B. Juditsky: Algorithms of Robust Stochastic Optimization Based on Mirror Descent Method. Automation and Remote Control, **80**(9), pp. 1607–1627, 2019.

[28]A. V. Nazin: Algorithms of Inertial Mirror Descent in Convex Problems of Stochastic Optimization. Automation and Remote Control, **79**(1), pp. 78–88, 2018.

[29]E. Hazan, S. Kale: Beyond the regret minimization barrier: Optimal algorithms for stochastic strongly-convex optimization. JMLR. **15**, pp. 2489–2512, 2014.

[30]E. Hazan: Introduction to online convex optimization. Foundations and Trends in Optimization, **2**(3–4), pp. 157–325, 2015.

[31]B. Awerbuch, R. Kleinberg: Online linear optimization and adaptive routing. Journal of Computer and System Sciences. **74**(1), pp. 97–114, 2008.

[32]See "J. Yuan, A. Lamperski: Online convex optimization for cumulative constraints. Published in NIPS, pp. 6140–6149, 2018." and references therein.

[33]S. Bubeck, R. Eldan: Multi-scale exploration of convex functions and bandit convex optimization. JMLR: Workshop and Conference Proceedings **49**, pp. 1–7, 2016.

[34]A. V. Gasnikov, A. A. Lagunovskaya, L. E. Morozova: On the relationship between simulation logit dynamics in the population game theory and a mirror descent method in the online optimization using the example of the shortest path problem. Proceedings of MIPT, **7**(4), pp. 104–113, 2015. (in Russian)

[35]A. V. Gasnikov, A. A. Lagunovskaya, I. N. Usmanova, F. A. Fedorenko, E. A. Krymova: Stochastic online optimization. Single-point and multi-point non-linear multi-armed bandits. Convex and strongly-convex case. Automation and Remote Control, **78**(2), pp. 224–234, 2017.

[36]R. Jenatton, J. Huang, C. Archambeau: Adaptive Algorithms for Online Convex Optimization with Long-term Constraints. Proceedings of The 33rd International Conference on Machine Learning, PMLR 48, pp. 402–411, 2016.

[37]Y. Hao, M. J. Neely, W. Xiaohan: Online Convex Optimization with Stochastic Constraints. Published in NIPS, pp. 1427–1437, 2017.

appear when applying the Lagrange multiplier method to the convex optimization problems with functional constraints (linear types of equalities and nonlinear convex types of inequalities). Recently, many researchers actively working in the subject of accelerated methods for saddle point problems, based on their structure. Over the past 15 years, the Nesterov's acceleration scheme[38] has been successfully transferred to smooth constrained optimization problems and to problems with structure (in particular, the so-called composite problems). Also, the acceleration scheme was successfully transferred to gradient-free methods, subgradient methods, directional descents, coordinate descents and methods using higher derivatives. It was also possible to accelerate randomized methods (for example, variance reduction methods in minimizing the sum of functions) and methods for solving smooth stochastic optimization problems. The success of the transfer, mentioned above, is understood to mean the achievement of well-known lower bounds of estimates by using the corresponding accelerated methods (with accuracy to a numerical factor). Very recently[39], it was generalized our results[40] and resolved a longstanding open question pertaining to the design of near-optimal first-order algorithms for smooth and μ_x–strongly-convex μ_y–strongly-concave minimax problems. They proposed a near-optimal algorithm, achieves a gradient complexity $\tilde{O}\left(1/\sqrt{\mu_x\mu_y}\right)$, which matches the lower complexity bound[41] up to logarithmic factors.

Despite the noted achievements, as mentioned previously, there remains a number of problems that are quite important for practice, in which it is not yet completely clear how to accelerate the available methods.

The main theme of the thesis *focuses on the non-smooth convex optimization problems with convex functional constraints and on their connection with the convex-concave composite saddle point problems.*

The goals of the thesis.

1. Development adaptive Mirror Descent algorithms, in order to solve convex optimization problems with functional constraints, in both deterministic and stochastic settings, with different levels of smoothness for the convex or strongly

[38] Y. Nesterov: A method of solving a convex programming problem with convergence rate $O(1/k^2)$. Soviet Mathematics Doklady, **27**(2), pp. 372–376, 1983.

[39] T. Lin, C. Jin, M. I. Jordan: Near-Optimal Algorithms for Minimax Optimization. 2020. `https://arxiv.org/pdf/2002.02417.pdf`

[40] M. S. Alkousa, D. M. Dvinskikh, F. S. Stonyakin, A. V. Gasnikov, D. Kovalev: Accelerated Methods for Saddle Point Problems. Accepted paper to the print in the journal *Computational Mathematics and Mathematical Physics*. `https://arxiv.org/pdf/1906.03620.pdf`

[41] J. Zhang, M. Hong, S. Zhang: On lower iteration complexity bounds for the saddle point problems. 2019. `https://arxiv.org/pdf/1912.07481.pdf`

convex objective function, such as Lipschitz-continuous, the gradient or the Hessian of the objective function is Lipschitz-continuous.

2. Development some Mirror Descent algorithms for online optimization problems with functional constraints, in both deterministic and stochastic settings.

3. Development accelerated methods for a more general class of the convex-concave composite saddle point problems and obtain the optimal estimates of the convergence rate for accelerated methods for the considered class of problems.

The tasks of the thesis.

1. Developing modifications of adaptive Mirror Descent algorithms to solve the optimization problem of a convex function with convex functional constraints, in deterministic and stochastic settings.

2. Developing some algorithms for online optimization problems with functional constraints, in both deterministic and stochastic settings.

3. Developing adaptive algorithms for solving strongly convex optimization problems with one functional constraint.

4. Studying a more general class of the convex-concave composite saddle point problems and obtain the estimates of the convergence rate for accelerated methods for non-bilinear convex-concave smooth composite saddle point problems.

Scientific novelty.

It was proposed a new modification of an adaptive deterministic and stochastic Mirror Descent algorithms, in order to solve the convex optimization problems with non-smooth functional constraints, in the case when the objective function is Lipschitz-continuous. The proposed modification allows saving the running time of the algorithms due to the consideration of not all functional constraints on non-productive steps. Specific estimates of the convergence rate for some adaptive and partial adaptive Mirror Descent algorithms, in the case, when the objective function is Lipschitz-continuous and when its Hessian is Lipschitz-continuous, were obtained. Also, adaptive and non-adaptive algorithms, for stochastic online optimization problems with functional constraints were proposed. In order to solve the non-smooth, strongly convex optimization problems with one functional constraint, it was proposed an adaptive algorithm, its stopping criterion can speed up the work of the

algorithm compared to the other optimal algorithm for some examples of non-smooth strongly convex problems. Also, it was studied a more general class of the convex-concave saddle point problems and obtained the estimates of the convergence rate for accelerated methods for non-bilinear convex-concave smooth composite saddle point problems. Furthermore, numerical experiments were carried out for some examples, to show the advantages of the proposed algorithms, in deterministic and stochastic settings of the considered optimization problem, and the advantages of the using the technique of restarts, in order to solve the optimization problems with functional constraints in the case when the objective function and the functional constraints are strongly convex.

The theoretical and practical value of the work in the thesis.

The proposed algorithms in the thesis, for convex optimization and online optimization problems with functional constraints, in both deterministic and stochastic settings, are adaptive. The adaptive adjustment is with respect to the Lipschitz constant of the objective function itself or of its gradient or Hessian, as well as the Lipschitz constant of the functional constraints. This adaptivity of the proposed algorithms is very important in practice and many applications, such as in Machine Learning scenarios, large-scale optimization problems, and their applications. Also, the proposed modified algorithms are applicable to the objective functions of various levels of smoothness: the Lipschitz condition holds either for the objective function itself or for its gradient or Hessian (while the function itself can fail to satisfy the Lipschitz condition) and they allow saving the running time of the algorithms. In the proposed algorithms, we consider an arbitrary proximal structure, which allows us to solve the optimization problems in the case of non-Euclidean distance. Also, in order to solve the classical problem of minimizing a strongly convex function with one non-smooth functional constraint, in the proposed algorithm there is an adaptive adjustment with respect to the strong convexity parameter, where the strong convexity of the functional constraint is not required, and there is also no need to know the value of the strong convexity parameter of the objective function, which is not available in some optimal algorithms, such as Mirror Descent. The studied approach of accelerated methods for saddle point problems was for a more general class of the convex-concave composite saddle point problems. Such problems arise, for example, in image processing and in solving various inverse problems. The obtained results can be generalized to the case of more general settings, such as a stochastic setting. Also, instead of the Euclidean norm, one could consider more general norms and proximal structures; finally, one could try to consider more than

two terms in the structure of the objective function.

Statements to be defended.

The statements to be defended in the thesis can be enumerated as follows:

1. A new modification of adaptive deterministic and stochastic Mirror Descent algorithms, in order to solve the convex optimization problems with non-smooth functional constraints, in the case when the objective function is Lipschitz-continuous.

2. Specific estimates of the convergence rate for some adaptive and partial adaptive Mirror Descent algorithms, in the case, when the objective function is Lipschitz-continuous and when its Hessian is Lipschitz-continuous.

3. Adaptive and non-adaptive algorithms, for the stochastic online optimization problems with functional constraints.

4. An adaptive algorithm, in order to solve non-smooth, strongly convex optimization problems with one functional constraint and comparison with an optimal Mirror Descent algorithm.

5. Studying a more general class of the convex-concave saddle point problems and obtain the estimates of the convergence rate for accelerated methods for non-bilinear convex-concave smooth composite saddle point problems.

Presentations and validation of research results.

The results of the thesis were reported and discussed at the following scientific conferences and seminars:

1. 61st Scientific Conference MIPT. Moscow, 24.11.2018;

2. Conference on graphs, networks, and their applications (Workshop Network Optimization). Moscow, MIPT, 16.05.2019;

3. 18th International Conference on Mathematical Optimization Theory and Operation Research (MOTOR-2019), Ekaterinburg, Russia, July 8–12, 2019;

4. International conference "Equation of Convolution Type in Science and Technology"ECTST-2019, Simferopol, Russian Federation, September 25–28, 2019;

5. 62nd Scientific Conference MIPT. Moscow, 23.11.2019;

6. Fifth international conference Quasilinear Equations, Inverse Problems and Their Applications. Moscow, MIPT, 02.12.2019;

7. Scientific seminar in the laboratory of numerical methods of applied structural optimization (under the guidance of Professor Yu. G. Evtushenko). Moscow, MITP, 29.05.2019 and 12.02.2020;

8. Scientific seminar of the Department of Algebra and Functional Analysis, Faculty of Mathematics and Computer Science, Taurida Academy, V. I. Vernadsky Crimea Federal University (under the guidance of Professor I. V. Orlov). Simferopol, Russia, 22.09.2019.

Publications

The results in the thesis are represented in five published papers **(1)**–**(5)** and in two accepted papers to print, **(6)** and **(7)**.

(1) F. S. Stonyakin, M. S. Alkousa, A. N. Stepanov, M. A. Barinov: Adaptive mirror descent algorithms in convex programming problems with Lipschitz constraints. Trudy Instituta Matematiki i Mekhaniki URO RAN, **24**(2), pp. 266–279, 2018. http://journal.imm.uran.ru/node/287 *"Web of Science"*

(2) A. A. Titov, F. S. Stonyakin, A. V. Gasnikov, M. S. Alkousa: Mirror Descent and Constrained Online Optimization Problems// Optimization and Applications. 9th International Conference OPTIMA-2018. Communications in Computer and Information Science, **974**, pp. 64–78, 2019. https://link.springer.com/chapter/10.1007/978-3-030-10934-9_5 *"Scopus"*

(3) F. S. Stonyakin, M. S. Alkousa, A. N. Stepanov, A. A. Titov: Adaptive Mirror Descent Algorithms for Convex and Strongly Convex Optimization Problems with Functional Constraints. Journal of Applied and Industrial Mathematics, **13**(3), pp. 557–574, 2019. https://link.springer.com/article/10.1134/S1990478919030165 *"Scopus"*

(4) M. S. Alkousa: On Some Stochastic Mirror Descent Methods for Constrained Online Optimization Problems. Computer Research and Modeling, **11**(2), pp. 205–217, 2019. DOI: 10.20537/2076-7633-2019-11-2-205-217. http://crm-en.ics.org.ru/journal/article/2775/ *"Scopus"*

(5) F. S. Stonyakin, M. S. Alkousa, A. A. Titov, V. V. Piskunova: On Some Methods for Strongly Convex Optimization Problems with One Functional Constraint. In book: Mathematical Optimization Theory and Operations Research, 18th International Conference, MOTOR 2019, LNCS 11548, pp. 82–96, 2019. https://doi.org/10.1007/978-3-030-22629-9_7 *"Scopus"*

(6) M. S. Alkousa: On Modification of an Adaptive Stochastic Mirror Descent Algorithm for Convex Optimization Problems with Functional Constraints. Accepted to the print as a chapter in the forthcoming book: *Communications in Mathematical Computations and Applications, Springer.* https://arxiv.org/pdf/1904.09513.pdf

(7) M. S. Alkousa, D. M. Dvinskikh, F. S. Stonyakin, A. V. Gasnikov, D. Kovalev: Accelerated Methods for Saddle Point Problems. Accepted paper to the print in the journal *Computational Mathematics and Mathematical Physics.* https://arxiv.org/pdf/1906.03620.pdf

Personal contribution

The main contributions of the author in the thesis can be summarized as follows: The author in **(1) and (6)** proposed a new modification of adaptive deterministic and stochastic Mirror Descent algorithms, in order to solve the convex optimization problems with non-smooth convex functional constraints, in the case when the objective function is Lipschitz-continuous. An optimal estimate of a non-adaptive Mirror Descent algorithm, proposed in **(2)**, obtained for the deterministic setting of online optimization problems with functional constraints. In **(3)**, obtained specific estimates of the convergence rate of some adaptive and partial adaptive Mirror Descent algorithms, when the objective function is Lipschitz-continuous and when its Hessian is Lipschitz-continuous. In **(4)** proposed adaptive and non-adaptive algorithms, for the stochastic online optimization problems with functional constraints. In **(5)** proposed an adaptive algorithm, in order to solve the non-smooth strongly convex optimization problems with one functional constraint and comparison with an optimal Mirror Descent algorithm. Also, in **(7)**, the author studied a more general class of the convex-concave composite saddle point problems and obtained the estimates of the convergence rate for accelerated methods for non-bilinear convex-concave smooth composite saddle point problems.

The implementation of algorithms, as well as the preparation of numerical experiments in works **(1)**—**(6)** were performed by the author. The problem statement

in the works **(1)**, **(2)**, **(3)**, **(5)** and **(7)** was carried out by F. S. Stonyakin and A. V. Gasnikov.

The structure and amount of the thesis.

The thesis consists of an abstract, introduction, four chapters and list of 131 references. The full volume of the dissertation is 149 pages, including 16 figures and 15 tables.

The content of the work

The introduction justifies the relevance of the research conducted within the framework of the thesis, provides an overview of the scientific literature on the problem under study and an overview of each chapter in the thesis, formulates the goals, the tasks of the thesis, the scientific novelty, the theoretical and practical value of the presented thesis.

The first chapter briefly describes some fundamental concepts and tools in convex analysis and convex optimization, which will be useful in the remaining chapters of the thesis. **The first section 1.1** is devoted to the convex analysis tools, such as convex sets, differentiable and non-differentiable convex functions, and Lipschitz continuity. **The second section 1.2** is devoted to some basics and fundamental properties of the convex optimization problems and numerical methods in order to solve them. The focus was on the first-order methods and more attention on the basics of the Mirror Descent method.

Let $(\mathbf{E}, \|\cdot\|)$ be a normed finite-dimensional vector space, with an arbitrary norm $\|\cdot\|$, and $\mathbf{E}^*$ be the conjugate space of $\mathbf{E}$ with the standard norm $\|\cdot\|_* = \max_x\{\langle y, x\rangle, \|x\| \leq 1\}$, where $\langle y, x\rangle$ is the value of the continuous linear y at $x \in \mathbf{E}$. Let $d : Q \to \mathbb{R}$, where $Q \subset \mathbf{E}$ is a closed convex set, be a distance generating function (also called *prox-function*), which is continuously differentiable and 1-strongly convex with respect to the norm $\|\cdot\|$, i.e.

$$d(y) \geq d(x) + \langle \nabla d(x), y - x\rangle + \frac{1}{2}\|y - x\|^2 \quad \forall\, x, y \in Q.$$

For all $x, y \in Q$ we consider the corresponding *Bregman divergence*,

$$V_x(y) = d(y) - d(x) - \langle \nabla d(x), y - x\rangle.$$

For all $x \in Q$ and $p \in \mathbf{E}^*$, *the proximal mapping operator* (Mirror Descent step) is defined as follows

$$\mathrm{Mirr}_x(p) = \arg\min_{u \in Q} \big\{\langle p, u\rangle + V_x(u)\big\}.$$

We make the simplicity assumption, which means that $\mathrm{Mirr}_x(p)$ is easily computable.

The second chapter was under the title "*Mirror Descent Algorithms for Deterministic and Stochastic Constrained Optimization Problems*". In this chapter, it was considered the optimization problem of a convex function with several convex non-smooth functional constraints (see problem (2), below) and proposed Mirror

Descent algorithms in order to solve such problem, in deterministic or stochastic (randomized) setting, and in different situations: smooth or non-smooth, convex or strongly convex objective function and constraints. It was demonstrated that the proposed algorithms are applicable to the objective functions of various levels of smoothness: the Lipschitz condition holds either for the objective function itself or for its gradient or Hessian (while the function itself can fail to satisfy the Lipschitz condition).

Consider a set of convex functions f and $g_i : Q \to \mathbb{R}$, $i \in [m] \stackrel{\text{def}}{=} \{1, 2, \ldots, m\}$. Assume that all functionals g_i are Lipschitz-continuous with some constant $M_g > 0$, i.e.

$$|g_i(x) - g_i(y)| \leq M_g \|x - y\| \quad \forall\, x, y \in Q \ \text{ and } \ \forall i \in [m]. \tag{1}$$

In the thesis, it was focused on the most general constrained convex optimization problem

$$\min \{f(x) : \quad x \in Q \ \text{ and } \ g_i(x) \leq 0 \ \text{ for all } \ i \in [m]\}. \tag{2}$$

It is clear that instead of a set of functionals $\{g_i(\cdot)\}_{i=1}^m$ we can see one functional constraint $g : Q \to \mathbb{R}$, such that $g(x) = \max_{i \in [m]}\{g_i(x)\}$, and the function g also will be Lipschitz-continuous, with constant $M_g > 0$. Therefore, by this setting, the problem (2) will be equivalent to the following constrained minimization problem

$$f(x) \to \min_{x \in Q,\, g(x) \leq 0}. \tag{3}$$

The first section 2.1 is devoted to a new modification of an adaptive Mirror Descent algorithm (see Algorithm 1, below) which was proposed for the deterministic setting of the problem (2)[42]. The proposed modification allows saving the running time of the algorithm, due to the consideration of not all functional constraints on non-productive steps (i.e. skipping some of the functional constraints). Note that we obtain the non-productive steps when we have non-feasible points.

In order to mention the proposed modified algorithm (see Algorithm 2, below), suppose we have a constant $\Theta_0 > 0$, such that $d(x^*) \leq \Theta_0^2$, where x^* is a solution of (2). Note that if there is a set, $X_* \subset Q$, of optimal points for the problem (2), we may assume that $\min\limits_{x^* \in X_*} d(x^*) \leq \Theta_0^2$. We say that a point $\hat{x} \in Q$ is an ε–solution of (2) if

$$f(\hat{x}) - f(x^*) \leq \varepsilon \ \text{ and } \ g(\hat{x}) \leq \varepsilon. \tag{4}$$

In thesis it was proved the following result, which gives the complexity estimate for the proposed Algorithm 2 in the case when the objective function f is Lipschitz-continuous.

[42] See the article referenced in footnote [25].

Algorithm 1. *Adaptive Mirror Descent: objective function is Lipschitz-continuous.*

Algorithm 2. *The modification of Algorithm 1.*

	Algorithm 1	**Algorithm 2**
Require	$\varepsilon > 0,\ \Theta_0$.	$\varepsilon > 0,\ \Theta_0$.
1:	$x^0 = \arg\min\limits_{x \in Q} d(x)$	$x^0 = \arg\min\limits_{x \in Q} d(x)$
2:	$I =: \emptyset$	$I =: \emptyset$
3:	$N \leftarrow 0$	$N \leftarrow 0$
4:	**repeat**	**repeat**
5:	**if** $g(x^N) \leq \varepsilon$ **then**	**if** $g(x^N) \leq \varepsilon$ **then**
6:	$h_N = \frac{\varepsilon}{\|\nabla f(x^N)\|_*^2}$,	$h_N = \frac{\varepsilon}{\|\nabla f(x^N)\|_*^2}$,
7:	$x^{N+1} = \mathrm{Mirr}_{x^N}\left(h_N \nabla f(x^N)\right)$,	$x^{N+1} = \mathrm{Mirr}_{x^N}\left(h_N \nabla f(x^N)\right)$,
8:	$N \to I$	$N \to I$
9:	**else** // $\left(g(x^N) > \varepsilon\right)$	**else** // $(g_{j(N)}(x^N) > \varepsilon)$ *for some* $j(N) \in [m]$
10:	$h_N = \frac{\varepsilon}{\|\nabla g(x^N)\|_*^2}$,	$h_N = \frac{\varepsilon}{\left\|\nabla g_{j(N)}(x^N)\right\|_*^2}$,
11:	$x^{N+1} = \mathrm{Mirr}_{x^N}\left(h_N \nabla g(x^N)\right)$,	$x^{N+1} = Mirr_{x^N}\left(h_N \nabla g_{j(N)}(x^N)\right)$,
12:	**end if**	**end if**
13:	$N \leftarrow N + 1$	$N \leftarrow N + 1$
14:	**until** $\sum\limits_{j=0}^{N-1} \frac{1}{M_j^2} \geq \frac{2\Theta_0^2}{\varepsilon^2}$.	**until** $\sum\limits_{j=0}^{N-1} \frac{1}{M_j^2} \geq \frac{2\Theta_0^2}{\varepsilon^2}$.
Ensure	$\bar{x}^N := \sum\limits_{k \in I} x^k h_k \big/ \sum\limits_{k \in I} h_k$.	$\bar{x}^N := \sum\limits_{k \in I} x^k h_k \big/ \sum\limits_{k \in I} h_k$.

Theorem 1. *let $\varepsilon > 0$ be a fixed positive number and the stopping criterion of Algorithm 2 holds. Then $\bar{x}^N$ is an ε–solution to the problem (2) in the sense of (4), i.e*

$$f(\bar{x}^N) - f(x^*) \leq \varepsilon \ \ and \ \ g(\bar{x}^N) \leq \varepsilon, \tag{5}$$

and the Algorithm 2 stops after no more than

$$N = \left\lceil \frac{2\max\{M_f^2, M_g^2\}\Theta_0^2}{\varepsilon^2} \right\rceil \tag{6}$$

iterations.

For the case when the objective function is with Lipschitz gradient, two algorithms (Algorithm 3 and its modification Algorithm 4) were proposed by F. S. Stonyakin[43].

It was proved that, Algorithm 4 (as well as Algorithm 3) stops after no more than

$$N = \left\lceil \frac{2\max\{1, M_g^2\}\Theta_0^2}{\varepsilon^2} \right\rceil \tag{7}$$

[43]See "F. S. Stonyakin, M. S. Alkousa, A. N. Stepanov, M. A. Barinov: Adaptive mirror descent algorithms in convex programming problems with Lipschitz constraints. Trudy Instituta Matematiki i Mekhaniki URO RAN, **24**(2), pp. 266–279, 2018. `http://journal.imm.uran.ru/node/287`"

Algorithm 3. *Adaptive Mirror Descent: objective function with Lipschitz gradient.*

Algorithm 4. *The modification of Algorithm 3.*

	Algorithm 3	**Algorithm 4**
Require	$\varepsilon > 0,\ \Theta_0$.	$\varepsilon > 0,\ \Theta_0$.
1:	$x^0 = \arg\min\limits_{x \in Q} d(x)$	$x^0 = \arg\min\limits_{x \in Q} d(x)$
2:	$I =: \emptyset$	$I =: \emptyset$
3:	$N \leftarrow 0$	$N \leftarrow 0$
4:	**repeat**	**repeat**
5:	**if** $g(x^N) \leq \varepsilon$ **then**	**if** $g(x^N) \leq \varepsilon$ **then**
6:	$h_N = \frac{\varepsilon}{\|\nabla f(x^N)\|_*^2}$,	$h_N = \frac{\varepsilon}{\|\nabla f(x^N)\|_*}$,
7:	$x^{N+1} = \mathrm{Mirr}_{x^N}\left(h_N \nabla f(x^N)\right)$,	$x^{N+1} = \mathrm{Mirr}_{x^N}\left(h_N \nabla f(x^N)\right)$,
8:	$N \to I$	$N \to I$
9:	**else** // $(g(x^N) > \varepsilon)$	**else** // $(g_{j(N)}(x^N) > \varepsilon)$ *for some* $j(N) \in [m]$
10:	$h_N = \frac{\varepsilon}{\|\nabla g(x^N)\|_*^2}$,	$h_N = \frac{\varepsilon}{\left\|\nabla g_{j(N)}(x^N)\right\|_*^2}$,
11:	$x^{N+1} = Mirr_{x^N}\left(h_N \nabla g(x^N)\right)$,	$x^{N+1} = Mirr_{x^N}\left(h_N \nabla g_{j(N)}(x^N)\right)$,
12:	**end if**	**end if**
13:	$N \leftarrow N + 1$	$N \leftarrow N + 1$
14:	**until** $\Theta_0^2 \leq \frac{\varepsilon^2}{2}\left(|I| + \sum\limits_{k \notin I} \frac{1}{\|\nabla g(x^k)\|_*^2}\right)$.	**until** $\Theta_0^2 \leq \frac{\varepsilon^2}{2}\left(|I| + \sum\limits_{k \notin I} \frac{1}{\|\nabla g_{j(N)}(x^k)\|_*^2}\right)$.
Ensure	$\bar{x}^N := \arg\min_{x^k,\, k \in I} f(x^k)$.	$\bar{x}^N := \arg\min_{x^k,\, k \in I} f(x^k)$.

iterations.

Also, in order to solve the problem (2), when the objective functions with Lipschitz gradient, in thesis, it was mentioned to the partial adaptive Mirror Descent algorithm[44] (see Algorithm 5). The difference between Algorithms 3 and 5, is that in Algorithm 5 the step-sizes and the stopping criterion are non-adaptive, they require the constant M_g. Note that the number of iterations for Algorithm 5 is fixed and it is equaling

$$N = \left\lceil \frac{2M_g^2 \Theta_0^2}{\varepsilon^2} \right\rceil. \tag{8}$$

Also in section 2.1, it was obtained specific estimates of the convergence rate for Algorithms 4 and 5, which justify their optimality from the point of view of the theory of lower bound of estimates. Moreover, we consider various classes (various level of smoothness) of objective functions: Lipschitz-continuous functions and functions with Lipschitz Hessian[45].

In the case, when **the objective function is Lipschitz-continuous**, for

[44]This algorithm was proposed by F. S. Stonyakin in "F. S. Stonyakin, A. A. Titov: One Mirror Descent Algorithm for Convex Constrained Optimization Problems with Non-Standard Growth Properties. In Proceedings of the School-Seminar on Optimization Problems and their Applications (OPTA-SCL 2018) Omsk, Russia, July 8-14, 2018. CEUR Workshop Proceedings, **2098**, pp. 372–384, 2018."

[45]The case when the functions with Lipschitz gradient was studied by F. S. Stonyakin, see the article referenced in footnote [44].

Algorithm 5. *Partial adaptive Mirror Descent version of Algorithm 3.*

Algorithm 6. *Modification of an adaptive stochastic Mirror Descent: objective function is Lipschitz-continuous.*

	Algorithm 5	**Algorithm 6**
Require	$\varepsilon > 0,\ M_g > 0,\ \Theta_0$.	$\varepsilon > 0,\ \Theta_0$.
1:	$x^0 = \arg\min\limits_{x \in Q} d(x)$	$x^0 = \arg\min\limits_{x \in Q} d(x)$
2:	$I =: \emptyset$	$I =: \emptyset$
3:	$N \leftarrow 0$	$N \leftarrow 0$
4:	**repeat**	**repeat**
5:	**if** $g(x^N) \leq \varepsilon$ **then**	**if** $g(x^N) \leq \varepsilon$ **then**
6:	$h_N = \frac{\varepsilon}{M_g \cdot \|\nabla f(x^N)\|_*}$,	$M_N := \left\|\nabla f(x^N, \xi^N)\right\|_*$,
7:	$x^{N+1} = Mirr_{x^N}\left(h_N \nabla f(x^N)\right)$,	$h_N = \Theta_0 \left(\sum\limits_{t=0}^{N} M_t^2\right)^{-1/2}$,
8:	$N \to I$	$x^{N+1} := Mirr_{x^N}\left(h_N \nabla f(x^N, \xi^N)\right)$,
9:	**else** // $g(x^N) > \varepsilon$	$N \to I$
10:	$h_N = \frac{\varepsilon}{M_g^2}$,	**else** // $g_{j(N)}(x^N) > \varepsilon$ for some $j(N) \in [m]$
11:	$x^{N+1} = Mirr_{x^N}\left(h_N \nabla g(x^N)\right)$,	$M_N := \left\|\nabla g_{j(N)}(x^N, \zeta^N)\right\|_*$,
12:	**end if**	$h_N = \Theta_0 \left(\sum\limits_{t=0}^{N} M_t^2\right)^{-1/2}$;
13:	$N \leftarrow N + 1$	$x^{N+1} := Mirr_{x^N}\left(h_N \nabla g_{j(N)}(x^N, \zeta^N)\right)$,
14:	**until** $N \geq \left\lceil \frac{2M_g^2 \Theta_0^2}{\varepsilon^2} \right\rceil$.	**end if**
15:	—	$N \leftarrow N + 1$
16:	—	**until** $N \geq \frac{2\Theta_0}{\varepsilon} \left(\sum\limits_{t=0}^{N-1} M_t^2\right)^{1/2}$.
Ensure	$\bar{x}^N := \arg\min_{x^k,\, k \in I} f(x^k)$.	$\bar{x}^N := \frac{1}{N_I} \sum_{k \in I} x^k$.

Algorithms 4 and 5 we have the following corollary

Corollary 1. *Let $f : Q \to \mathbb{R}$ satisfies the Lipschitz condition on Q, with constant $M_f > 0$. Then*

- *after*

$$N = \left\lceil \frac{2 \max\{1, M_g^2\} \Theta_0^2}{\varepsilon^2} \right\rceil$$

steps of the work of Algorithm 4 (as well as Algorithm 3), the following estimate holds:

$$\min_{k \in [N]} f(x^k) - f(x^*) \leq M_f\, \varepsilon; \tag{9}$$

- *after*

$$N = \left\lceil \frac{2M_g^2\Theta_0^2}{\varepsilon^2} \right\rceil$$

steps of the work of Algorithm 5, the following estimate holds:

$$\min_{k\in[N]} f(x^k) - f(x^*) \leq \frac{M_f}{M_g}\varepsilon. \tag{10}$$

Remark 1. *In thesis it was clarified when the partial adaptive Algorithm 5 may be more advantageous and more effective than adaptive Algorithm 3. But here, in order to the brief: note that for these algorithms there are different stopping criteria. In addition, from* (7) *and* (8)*, we can see that, when $M_g < 1$, then the Algorithm 5 works faster than Algorithm 3 and the contrary when $M_g > 1$. Also, we can see that the difference between the estimates* (9) *and* (10)*, is only by the Lipschitz constant M_g. Therefore, after achieving the stopping criteria of Algorithms 3 and 5, we get a solution to the considered problem, with different estimates of the quality, dependently on the value of M_g.*

Now, suppose that, the objective function $f : Q \to \mathbb{R}$ is twice differentiable at each $x \in Q$ and its Hessian is Lipschitz-continuous with constant $L_H > 0$, i.e.

$$\left\|\nabla^2 f(x) - \nabla^2 f(y)\right\| \leq L_H\|x - y\| \quad \forall\, x, y \in Q, \tag{11}$$

where the norm $\|\cdot\|$ in (11), denotes the standard Euclidean norm; when applied to matrices it denotes the l_2-operator norm. In the case, when the **Hessian of the function f is Lipschitz-continuous**, we have the following result

Corollary 2. *Let $f : Q \to \mathbb{R}$ ba a twice differentiable function in Q and have the Lipschitz Hessian, i.e.* (11) *holds. Then*

- *after*

$$N = \left\lceil \frac{2\max\{1, M_g^2\}\Theta_0^2}{\varepsilon^2} \right\rceil$$

steps of the work of Algorithm 4 (as well as Algorithm 3), we have the following estimate

$$\min_{k\in I} f(x^k) - f(x^*) \leq \varepsilon \cdot \|\nabla f(x^*)\|_* + \frac{\varepsilon^2}{2}\left\|\nabla^2 f(x^*)\right\| + \frac{L_H}{6}\varepsilon^3; \tag{12}$$

- *after*

$$N = \left\lceil \frac{2M_g^2\Theta_0^2}{\varepsilon^2} \right\rceil$$

steps of the work of Algorithm 5, we have the following estimate

$$\min_{k\in I} f(x^k) - f(x^*) \leq \frac{\varepsilon}{M_g} \cdot \|\nabla f(x^*)\|_* + \frac{1}{2} \cdot \left\|\nabla^2 f(x^*)\right\| \cdot \frac{\varepsilon^2}{M_g^2} + \frac{L_H}{6}\frac{\varepsilon^3}{M_g^3}. \quad (13)$$

Remark 2. *We can formulate the analogue of the previous estimates for the class of non-smooth functions, every one has the form* $f(x) = \max\limits_{i\in[r]} f_i(x)$*, where* f_i *is twice differentiable at each* $x \in Q$ *and*

$$\left\|\nabla^2 f_i(x) - \nabla^2 f_i(y)\right\| \leq L_i\|x - y\| \quad \forall\, x, y \in Q,$$

where $L_i > 0,\ \forall\ i \in [r]$*. Then we have the same two items in Corollary 2, with* $L_H = \max\limits_{i\in[r]} L_i$*.*

Remark 3. *At the end of the first section of chapter 2, in order to show the advantages of the proposed modified Algorithm 2, some numerical tests were carried out, and compared Algorithms 1–4 for some examples. From the performed experiments, we saw that, the proposed modified Algorithm can significantly reduce both the number of iterations necessary to achieve the desired quality of the solution and the running time of the algorithms, for the considered problem* (2)*, with different forms of the functional constraints. Additionally, although one of Algorithms 3 and 5 works faster than another (dependently on the value of* M_g*, see Remark 1), but maybe the fastest algorithm give a worse quality of the solution. In order to show this, also some numerical experiments were carried out.*

In **section 2.2**, it was considered the problem (2) with assumption of strong convexity of f and g with the same parameter $\mu > 0$. By using the technique of restart another algorithm, some adaptive (restart Algorithms 1 and 3) and partially adaptive (restart Algorithm 5) optimal algorithms were mentioned for the problem under consideration. Some numerical experiments were carried out, in order to show that the technique of restarts justifies a higher convergence rate of the Mirror Descent algorithms for strongly convex constrained optimization problems.

Section 2.3 is devoted to a new modification of a recently proposed adaptive stochastic Mirror Descent algorithm[46], for the stochastic setting of the problem (2), in the case when the objective function is Lipschitz-continuous (the modified algorithm is listed as Algorithm 6). This means that we can still use the value of the objective function and functional constraints, but instead of their (sub)gradient, we use their stochastic (sub)gradient. For the stochastic setup of the problem (2), we introduce

[46]See Algorithm 4 in the article referenced in footnote [25].

the following assumptions. Given a point $x \in Q$, we can calculate the stochastic (sub)gradients $\nabla f(x,\xi)$ and $\nabla g(x,\zeta)$, where ξ and ζ are random vectors. These stochastic (sub)gradients satisfy

$$\mathbb{E}[\nabla f(x,\xi)] = \nabla f(x) \in \partial f(x) \quad \text{and} \quad \mathbb{E}[\nabla g(x,\zeta)] = \nabla g(x) \in \partial g(x), \tag{14}$$

where $\mathbb{E}$ denote to the expectation, and

$$\|\nabla f(x,\xi)\|_* \leq M_f \quad \text{and} \quad \|\nabla g(x,\zeta)\|_* \leq M_g, \quad a.s. \text{ in } \xi, \zeta. \tag{15}$$

We say that a (random) point $\hat{x} \in Q$ is an expected ε–solution to the problem (2), in stochastic setup, if

$$\mathbb{E}[f(\hat{x})] - f(x^*) \leq \varepsilon \ \text{ and } \ g(\hat{x}) \leq \varepsilon. \tag{16}$$

For the modified Algorithm 6, in thesis it was proved the following theorem, which shows that this algorithm is optimal

Theorem 2. *Let equalities* (14) *and inequalities* (15) *hold. Assume that a known constant* $\Theta_0 > 0$ *is such that* $\sup\limits_{x,y \in Q} V_x(y) \leq \Theta_0^2$. *Then Algorithm 6 stops after no more than*

$$N = \left\lceil \frac{4 \max\{M_f^2, M_g^2\} \Theta_0^2}{\varepsilon^2} \right\rceil \tag{17}$$

iterations and $\bar{x}^N$ *is an expected* ε*–solution in the sense of* (16).

In order to show the advantages of the proposed modified Algorithm 6, some numerical tests were carried out for some examples with different values of the accuracy and the dimension of the problem. From the performed experiments we can see that the proposed Algorithm 6 can significantly reduce both the number of iterations necessary to achieve the desired quality of the solution and the running time of the algorithm, for the considered problem (2) in the stochastic setting.

The third chapter was under the title *"Mirror Descent and Constrained Online Optimization Problems"*. In problems of online convex optimization (OCO), it is required to minimize the sum (or the arithmetic mean) of several convex functionals f_i ($i \in [N]$) given on some closed convex set $Q \subset \mathbb{R}^n$, with several convex non-smooth functional constraints, i.e. the following type of problems

$$\frac{1}{N} \sum_{i=1}^{N} f_i(x) \to \min_{x \in Q,\, g(x) \leq 0}. \tag{18}$$

We assume that f_i (for all $i \in [N]$) and g are Lipschitz-continuous functionals, with constant $M > 0$.

This chapter is devoted to some deterministic and stochastic Mirror Descent algorithms for the problem(18). In **section 3.1**, it was mentioned to some algorithms, for the deterministic setting[47] of (18), with an arbitrary proximal structure. The first one of these algorithms is a non-adaptive, while the second is adaptive with one modification. In the thesis, it was proved a theorem, which shows the optimality of the proposed non-adaptive algorithm.

Section 2.3, is devoted to a stochastic setting of the problem (18). This means that we can still use the value of the function g, but instead of (sub)gradient of f_i (for all $i \in [N]$) and g, we use their stochastic (sub)gradients $\nabla f_i(x, \xi), \nabla g(x, \zeta)$, where ξ, ζ are random vectors. It was proposed two algorithms, non-adaptive (Algorithm 7) and adaptive (Algorithm 8), for an arbitrary prox-structure. Each one of these Algorithms, produces N productive steps and in each step, the (sub)gradient of exactly one functional of the objectives is calculated. Denote the number of non-productive steps by N_J. As a result, we get a sequence $\{x^k\}_{k \in I}$ (on productive steps), which can be considered as a solution to the problem (18) with accuracy δ.

Algorithm 7. *Non-adaptive stochastic online Mirror Descent algorithm.*

Algorithm 8. *Adaptive stochastic online Mirror Descent algorithm.*

	Algorithm 7	**Algorithm 8**
Require	$\varepsilon,\ N,\ \Theta_0,\ Q,\ d(\cdot),\ x^0$.	$\varepsilon,\ N,\ \Theta_0,\ Q,\ d(\cdot),\ x^0$.
1:	$i := 1,\ k := 0,$	$i := 1,\ k := 0,$
2:	**repeat**	**repeat**
3:	**if** $g(x^k) \leq \varepsilon$ **then**	**if** $g(x^k) \leq \varepsilon$ **then**
4:	$h = \frac{\varepsilon}{M^2},$	$M_k := \left\|\nabla f_i(x^k, \xi^k)\right\|_*,$
5:	$x^{k+1} := Mirr_{x^k}\left(h \nabla f_i(x^k, \xi^k)\right),$	$h_k = \Theta_0 \left(\sum\limits_{t=0}^{k} M_t^2\right)^{-1/2},$
6:	$i := i + 1,$	$x^{k+1} := Mirr_{x^k}\left(h_k \nabla f_i(x^k, \xi^k)\right),$
7:	$k := k + 1,$	$i := i + 1,$
8:	**else**	$k := k + 1,$
9:	$h = \frac{\varepsilon}{M^2},$	**else**
10:	$x^{k+1} := Mirr_{x^k}\left(h \nabla g(x^k, \zeta^k)\right),$	$M_k := \left\|\nabla g(x^k, \zeta^k)\right\|_*,$
11:	$k := k + 1,$	$h_k = \Theta_0 \left(\sum\limits_{t=0}^{k} M_t^2\right)^{-1/2},$
12:	**end if**	$x^{k+1} := Mirr_{x^k}\left(h_k \nabla g(x^k, \zeta^k)\right),$
13:	**until** $i = N + 1$	$k := k + 1,$
14:	Guaranteed accuracy:	**end if**
15:	$\delta := \frac{\varepsilon}{2} + \frac{M^2 \Theta_0^2}{\varepsilon N} - \frac{\varepsilon N_J}{2N}.$	**until** $i = N + 1$
16:		Guaranteed accuracy:
		$\delta := \frac{2\Theta_0}{N} \left(\sum\limits_{i=0}^{N+N_J-1} M_i^2\right)^{1/2} - \varepsilon \cdot \frac{N_J}{N}.$

[47]These algorithms proposed by F. S. Stonyakin in "A. A. Titov, F. S. Stonyakin, A. V. Gasnikov, M. S. Alkousa: Mirror Descent and Constrained Online Optimization Problems// Optimization and Applications. 9th International Conference OPTIMA-2018. Communications in Computer and Information Science, **974**, pp. 64–78, 2019."

In order to Algorithms 7 and 8, it was proved the following result, which shows that the proposed algorithms are optimal[48].

Theorem 3. *Suppose Algorithm 7 (or Algorithm 8) works exactly N productive steps. After the stopping of the Algorithm 7 (or Algorithm 8), the following inequality holds*

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} f_i(x^k)\right] - \min_{x\in Q}\frac{1}{N}\sum_{i=1}^{N} f_i(x) \leq \delta.$$

For the case $\delta \leq \varepsilon = \frac{C}{\sqrt{N}}$, for some $C > 0$, and

$$\mathbb{E}\left[\frac{1}{N}\sum_{i=1}^{N} f_i(x^k)\right] - \min_{x\in Q}\frac{1}{N}\sum_{i=1}^{N} f_i(x) \geq 0,$$

there will be no more than $O(N)$ non-productive steps.

Remark 4. *At the end of each section in chapter 3, in order to compare the proposed algorithms, some numerical experiments were carried out. From all performed experiments, it was shown that the adaptive variant of algorithms, in both settings, deterministic and stochastic, works better than non-adaptive algorithms, with respect to the number of iterations, the running time of algorithms and the guaranteed accuracy δ, where the number of the non-productive steps and the value of δ obtained by adaptive algorithms are very small compared to the non-adaptive algorithms, in both settings of the proposed problem.*

The forth chapter was under the title "*Accelerated Methods for Saddle Point Problems*". In **section 4.1**, it was considered the following more general class of the saddle point problems

$$\min_{x\in Q_x}\max_{y\in Q_y}\{S(x,y) := r(x) + F(x,y) - h(y)\}. \tag{19}$$

where, $Q_x \subseteq \mathbb{R}^m, Q_y \subseteq \mathbb{R}^n$ are non-empty, convex and compact sets. $r : Q_x \to \mathbb{R}$ and $h : Q_y \to \mathbb{R}$ are μ_x–strongly convex and μ_y–strongly concave functions, respectively. The functional $F : Q_x \times Q_y \to \mathbb{R}$ is convex by x, concave by y and given in some neighborhood of the set $Q_x \times Q_y$. Moreover, we consider F to be sufficiently smooth on $Q_x \times Q_y$, i.e. for any $x, x' \in Q_x$ and $y, y' \in Q_y$, the following inequalities are satisfied

$$\|\nabla_x F(x,y) - \nabla_x F(x',y)\|_2 \leq L_{xx}\|x - x'\|_2, \quad \|\nabla_x F(x,y) - \nabla_x F(x,y')\|_2 \leq L_{xy}\|y - y'\|_2,$$

[48]For the concept of the optimality, see "E. Hazan, S. Kale: Beyond the regret minimization barrier: Optimal algorithms for stochastic strongly-convex optimization. JMLR. **15**, pp. 2489–2512, 2014."

$$\|\nabla_y F(x,y) - \nabla_y F(x',y)\|_2 \le L_{xy}\|x - x'\|_2, \quad \|\nabla_y F(x,y) - \nabla_y F(x,y')\|_2 \le L_{yy}\|y - y'\|_2,$$

where $L_{xx}, L_{xy}, L_{yy} \ge 0$.

Note that, the problem (19) is equivalent to the following optimization problem

$$f(x) := r(x) + \underbrace{\max_{y \in Q_y}\{F(x,y) - h(y)\}}_{g(x) := F(x,y^*(x)) - h(y^*(x))} \to \min_{x \in Q_x}, \tag{20}$$

where $y^*(x) = \arg\max_{y \in Q_y}\{F(x,y) - h(y)\}$.

From the point of view of accelerated methods, the class of problems (19) has been sufficiently studied in some details, mainly in the case when $F(x,y)$ is bilinear[49], i.e. $F(x,y) = \langle Ax, y\rangle$ for some linear operator A. In this case $L_{xx} = L_{yy} = 0$, $L_{xy} = L_{yx} = \sqrt{\lambda_{\max}(A^T A)}$.

In thesis, it was proposed to reduce the considered saddle problem (19) to a combination of auxiliary smooth strongly convex optimization problems of separately for each group of variables. Therefore, the application of acceleration Nesterov's gradient method lead to improve the estimates of convergence. We got an analogue of the Lan's results and showed that the best-known bounds for the bilinear convex-concave smooth composite saddle point problems keep true for the non-bilinear problems.

Definition 1. *We say that the function $r : Q_x \to \mathbb{R}$ is proximal-friendly, if the problem of the form*

$$\min_{x \in Q_x}\left\{\langle c_1, x\rangle + r(x) + c_2\|x\|_2^2\right\}, \tag{21}$$

where $c_1 \in Q_x$ and $c_2 > 0$, can be solved explicitly.

Similarly, $h : Q_y \to \mathbb{R}$ is proximal-friendly function, if the problem of the form

$$\min_{y \in Q_y}\left\{\langle c_3, y\rangle + h(y) + c_4\|y\|_2^2\right\}, \tag{22}$$

where $c_3 \in Q_y$ and $c_4 > 0$, can be solved explicitly.

In order to review the best-known results[50] [51], regarding the complexity of solving the problem (19), we distinguish several cases, according to some assumptions on the functions r and h. We mention to these best-known results in Table 1, below.

By the special case, in Table 1, we mean that $F(x,y) = \langle Ax, y\rangle$ and by the general case when the function $F(x,y)$ is non-bilinear.

[49] See "G. Lan: Lectures on Optimization Methods for Machine Learning. H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology, Atlanta, GA. 2019."

[50] L. T. K. Hien, R. Zhao, W. B. Haskell: An Inexact Primal-Dual Smoothing Framework for Large-Scale Non-Bilinear Saddle Point Problems. 2019. https://arxiv.org/pdf/1711.03669v3.pdf

[51] See the work referenced in footnote [49]

Таблица 1: The best-known results regarding the complexity of solving the problem (19).

State (1): both functions r and h are proximal-friendly.		
special case	$\tilde{O}\left(\frac{L_{xy}}{\sqrt{\mu_x\mu_y}}\right)$	calculations of (21), $\nabla_x F(x,y)$ and (22), $\nabla_y F(x,y)$.
general case	$\tilde{O}\left(\frac{\max\{L_{xx},L_{xy},L_{yy}\}}{\min\{\mu_x,\mu_y\}}\right)$	calculations of (21), $\nabla_x F(x,y)$ and (22), $\nabla_y F(x,y)$.
State (2): the function r is L_x–smooth and it is not proximal-friendly.		
special case	$\tilde{O}\left(\sqrt{\frac{L_x}{\mu_x}}\right)$	calculations of $\nabla r(x)$
	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\right)$	calculations of $\nabla_x F(x,y)$
	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\sqrt{\max\left\{\frac{L_{yy}}{\mu_y},1\right\}}\right)$	calculations of (22), $\nabla_y F(x,y)$
State (3): the function h is L_y–smooth and it is not proximal-friendly.		
special case	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\sqrt{\frac{L_y}{\mu_y}}\right)$	calculations of $\nabla h(y)$
	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\right)$	calculations of (21), $\nabla_x F(x,y)$
	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\sqrt{\max\left\{\frac{L_{yy}}{\mu_y},1\right\}}\right)$	calculations of $\nabla_y F(x,y)$
State (4): r is L_x–smooth, h is L_y–smooth and both they are not proximal-friendly.		
special case	$\tilde{O}\left(\sqrt{\frac{L_x}{\mu_x}}\right)$	calculations of $\nabla r(x)$
	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\right)$	calculations of $\nabla_x F(x,y)$
	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\sqrt{\frac{L_y}{\mu_y}}\right)$	calculations of $\nabla h(y)$
	$\tilde{O}\left(\sqrt{\frac{L_{xx}}{\mu_x}+\frac{L_{xy}^2}{\mu_x\mu_y}}\sqrt{\max\left\{\frac{L_{yy}}{\mu_y},1\right\}}\right)$	calculations of $\nabla_y F(x,y)$

From this table, we can say that an ε–solution of the problem (19) can be achieved in $\tilde{O}()$ calculations, which placed in the second column, of placed items in the third column.

In order to the accelerated methods for problem (19), the main result in the thesis was the justification of the results for states (2), (3) and (4) in Table 1, when $F(x,y) = \langle Ax, y\rangle$, for the general case, in which the function $F(x,y)$ is non-bilinear.

Another result was the refinement of the results in the case when $F(x,y) = \langle Ax, y\rangle$, for some linear operator A, $Q_y = \mathbb{R}^m$, h is L_y–smooth and $\frac{\lambda_{\min}\left(A^TA\right)}{L_y} \gg \mu_x$. In this case, in all formulas, we can replace μ_x with $\frac{\lambda_{\min}\left(A^TA\right)}{L_y}$.

In the case of non-smooth problems, the application of accelerated gradient method does not improve the estimates of complexity, it gives the same estimate as the non-accelerated method, which is $O\left(1/\varepsilon^2\right)$. Therefore, we try to highlight some other classes of problems. In **section 4.2**, for the non-smooth case of saddle point problems, as a special case, we consider the classical optimization problem of minimizing a strongly convex function with one non-smooth functional constraint

$$f(x) \to \min_{x\in Q,\, g(x)\le 0}. \tag{23}$$

where $f : Q \to \mathbb{R}$ is a Lipschitz-continuous with constant $M_f > 0$ and μ_f–strongly convex function, the functional constraint g is Lipschitz-continuous with constant $M_g > 0$. In the case of several strongly convex non-smooth constraints, we consider one max-type constraint which is also strongly convex. Note that the dual problem to the problem (23) has the following form

$$\varphi(\lambda) = f(x(\lambda)) + \lambda g(x(\lambda)) \to \max_{\lambda \geq 0}, \tag{24}$$

where

$$x(\lambda) = \arg\min_{x \in Q} \{f(x) + \lambda g(x)\}. \tag{25}$$

In order to solve the problem (23), an analogue of an adaptive Algorithm 9, which was proposed by F. S. Stonyakin[52], was proposed in thesis (see Algorithm 10). The difference between Algorithms 9 and 10, is represented only on the stooping criterion.

The approach in these algorithms is based on the transition to a strongly convex dual problem, in this case, the dual function depends on one dual variable, which allows us to use the dichotomy method and solving an auxiliary one-dimensional problem at each iteration.

Algorithm 9. *In Require, the interval* $[\lambda^0_{min}, \lambda^0_{max}]$ *is the initial localization interval of the dual variable.*

Algorithm 10. *Also, as in Algorithm 9,* δ *is an accuracy for the auxiliary problem, see item 4 in each algorithm.*

	Algorithm 9	**Algorithm 10**
Require	$[\lambda^0_{min}, \lambda^0_{max}]$; δ; ε.	$[\lambda^0_{min}, \lambda^0_{max}]$; δ; ε.
1:	$N := 0$	$N := 0$
2:	**repeat**	**repeat**
3:	$\lambda^N := \frac{\lambda^N_{min}+\lambda^N_{max}}{2}$,	$\lambda^N := \frac{\lambda^N_{min}+\lambda^N_{max}}{2}$,
4:	$x_\delta(\lambda^N) = \arg\min^{\delta}_{x \in Q}\{f(x) + \lambda^N g(x)\}$,	$x_\delta(\lambda^N) = \arg\min^{\delta}_{x \in Q}\{f(x) + \lambda^N g(x)\}$,
5:	$\varphi'(\lambda^N) = g(x_\delta(\lambda^N))$,	$\varphi'(\lambda^N) = g(x_\delta(\lambda^N))$,
6:	**if** $\varphi'(\lambda^N) < 0$ **then** $\lambda^{N+1}_{max} := \frac{\lambda^N_{min}+\lambda^N_{max}}{2}$,	**if** $\varphi'(\lambda^N) < 0$ **then** $\lambda^{N+1}_{max} := \frac{\lambda^N_{min}+\lambda^N_{max}}{2}$,
7:	**if** $\varphi'(\lambda^N) > 0$ **then** $\lambda^{N+1}_{min} := \frac{\lambda^N_{min}+\lambda^N_{max}}{2}$,	**if** $\varphi'(\lambda^N) > 0$ **then** $\lambda^{N+1}_{min} := \frac{\lambda^N_{min}+\lambda^N_{max}}{2}$,
8:	$N := N + 1$;	$N := N + 1$;
9:	**until** $\lambda^N \lvert g(x_\delta(\lambda^N))\rvert \leq \varepsilon$.	**until** $\lvert g(x_\delta(\lambda^N))\rvert \leq \varepsilon$.
Ensure	λ^N, with $\lambda^N \lvert g(x_\delta(\lambda^N))\rvert \leq \varepsilon$; $x_\delta(\lambda^N)$.	λ^N, with $\lvert g(x_\delta(\lambda^N))\rvert \leq \varepsilon$; $x_\delta(\lambda^N)$.

In order to estimate the accuracy of the solution of the considered problem by Algorithm 10, the following lemma was concluded in thesis

[52]See "F. S. Stonyakin, M. S. Alkousa, A. A. Titov, V. V. Piskunova: On Some Methods for Strongly Convex Optimization Problems with One Functional Constraint. In book: Mathematical Optimization Theory and Operations Research, 18th International Conference, MOTOR 2019, LNCS 11548, pp. 82–96, 2019."

Lemma 1. *Suppose the stopping criterion of Algorithm 10 holds for $\lambda = \lambda^N$. Then the following inequalities hold*

$$f(x_\delta(\lambda)) - f(x^*) \leq \lambda\varepsilon + \delta \quad and \quad g(x_\delta(\lambda)) \leq \varepsilon. \tag{26}$$

where $x^* = x(\lambda^*)$. For the case $\delta = \varepsilon$ we get

$$f(x_\delta(\lambda)) - f(x^*) \leq (\lambda + 1)\varepsilon \quad and \quad g(x_\delta(\lambda)) \leq \varepsilon.$$

Remark 5. *Lemma 1, is the analogue of a lemma, proposed by F. S. Stonyakin, for Algorithm 9, where in this lemma, instead of the inequalities* (26)*, we have the following estimates*

$$f(x_\delta(\lambda)) - f(x^*) \leq \varepsilon + \delta \quad and \quad g(x_\delta(\lambda)) \leq \frac{\varepsilon}{\lambda}, \tag{27}$$

Remark 6. *By analyzing* (26) *and* (27) *we can see that, Algorithm 9 guarantees the desired accuracy of the solution with respect to the objective function, but, possibly, unsatisfactory accuracy of the solution with respect to the constraint, as the estimate is huge in case λ is small. Algorithm 10 provides the desired accuracy of the solution with respect to the constraint and, possibly, unsatisfactory accuracy of the solution with respect to the objective function in a case λ is huge. So one of the Algorithms 9 and 10 surely guarantees the desired accuracy with respect to both objective function and constraint.*

To estimate the convergence rate of the proposed Algorithm 10, in the case when the objective function is Lipschitz-continuous, it was obtained that, the Algorithm 10 stops after no more than

$$O\left(\log_2 \frac{M_g^2 \lambda_{\max}}{\varepsilon \mu_f}\right)$$

iterations. Where, $\lambda_{max} = \frac{1}{\gamma}\Big(f(\overline{x}) - \min\limits_{x \in Q} f(x)\Big)$, and $\overline{x} \in Q$ is an arbitrary point, such that $g(\overline{x}) = -\gamma < 0$. Also, in thesis it was proved that the general complexity of algorithm 10 will be

$$O\left(\frac{1}{\delta^2} log_2 \frac{1}{\varepsilon}\right). \tag{28}$$

Remark 7. *If $\delta = O(\varepsilon)$ then the general complexity* (28) *of Algorithm 10 will be $O\left(\frac{1}{\varepsilon^2} log_2 \frac{1}{\varepsilon}\right)$, which is generally not optimal. However, due to the adaptivity of the stopping criteria of Algorithms 9 and 10, these algorithms can work faster than optimal algorithms, such as Mirror Descent algorithms. Note that we require, in*

Algorithms 9 and 10, the strong convexity only of the objective function f. In this case, the functional g, in general, may not be strongly convex. Also, there is no need to know the value of the strong convexity parameter of the objective function, which is not available in some optimal methods, such as Mirror Descent, wherein these methods require the strong convexity and knowing the value of the strong convexity parameter of both objective function and functional constraints.

Remark 8. *At the end of the last section 4.2, some numerical experiments were carried out, in order to the comparison between Algorithms 9 and 10. Also, in order to compare these algorithms with an optimal adaptive Mirror Descent algorithm. It was compared the running time of Algorithms and the quality of a solution, produced by these algorithms, with respect to the objective function and the functional constraints. It was shown that in some examples, the Algorithm 9 can work faster than the proposed Algorithm 10, but although Algorithm 10 works slower than Algorithm 9, it gives better quality of a solution with respect to the objective function f. The contrary was shown in other examples. Also, It was shown, in some examples, that Algorithms 9 and 10 work better than the adaptive Mirror Descent Algorithm, which is based on the technique of restart Algorithm 1.*

Author's publications on the dissertation topic

1. F. S. Stonyakin, M. S. Alkousa, A. N. Stepanov, M. A. Barinov: Adaptive mirror descent algorithms in convex programming problems with Lipschitz constraints. Trudy Instituta Matematiki i Mekhaniki URO RAN, **24**(2), pp. 266–279, 2018. http://journal.imm.uran.ru/node/287

2. A. A. Titov, F. S. Stonyakin, A. V. Gasnikov, M. S. Alkousa: Mirror Descent and Constrained Online Optimization Problems// Optimization and Applications. 9th International Conference OPTIMA-2018. Revised Selected Papers. Communications in Computer and Information Science, **974**, pp. 64–78, 2019. https://link.springer.com/chapter/10.1007/978-3-030-10934-9_5

3. F. S. Stonyakin, M. S. Alkousa, A. N. Stepanov, A. A. Titov: Adaptive Mirror Descent Algorithms for Convex and Strongly Convex Optimization Problems with Functional Constraints. Journal of Applied and Industrial Mathematics, **13**(3), pp. 557–574, 2019. https://link.springer.com/article/10.1134/S1990478919030165

4. M. S. Alkousa: On Some Stochastic Mirror Descent Methods for Constrained Online Optimization Problems. Computer Research and Modeling, **11**(2), pp. 205–217, 2019. DOI: 10.20537/2076-7633-2019-11-2-205-217. http://crm-en.ics.org.ru/journal/article/2775/

5. F. S. Stonyakin, M. S. Alkousa, A. A. Titov, V. V. Piskunova: On Some Methods for Strongly Convex Optimization Problems with One Functional Constraint. In book: Mathematical Optimization Theory and Operations Research, Proceedings Springer Nature Switzerland AG 2019. M. Khachay et al. (Eds.): MOTOR 2019, LNCS 11548, pp. 82–96, 2019. https://doi.org/10.1007/978-3-030-22629-9_7

6. M. S. Alkousa: On Modification of an Adaptive Stochastic Mirror Descent Algorithm for Convex Optimization Problems with Functional Constraints. Accepted to the print as a chapter in the forthcoming book: *Communications in Mathematical Computations and Applications, Springer.* https://arxiv.org/pdf/1904.09513.pdf

7. M. S. Alkousa, D. M. Dvinskikh, F. S. Stonyakin, A. V. Gasnikov, D. Kovalev: Accelerated Methods for Saddle Point Problems. Accepted paper to the print in the journal *Computational Mathematics and Mathematical Physics.* https://arxiv.org/pdf/1906.03620.pdf