COMPLEXITY AND SIMPLICITY OF OPTIMIZATION PROBLEMS

Yurii Nesterov

REPRINT | 3312

UCLouvain

LIDAM

CORE | Louvain Institute of Data Analysis and Modeling in economics and statistics

CORE

Voie du Roman Pays 34, L1.03.01

B-1348 Louvain-la-Neuve

Tel (32 10) 47 43 04

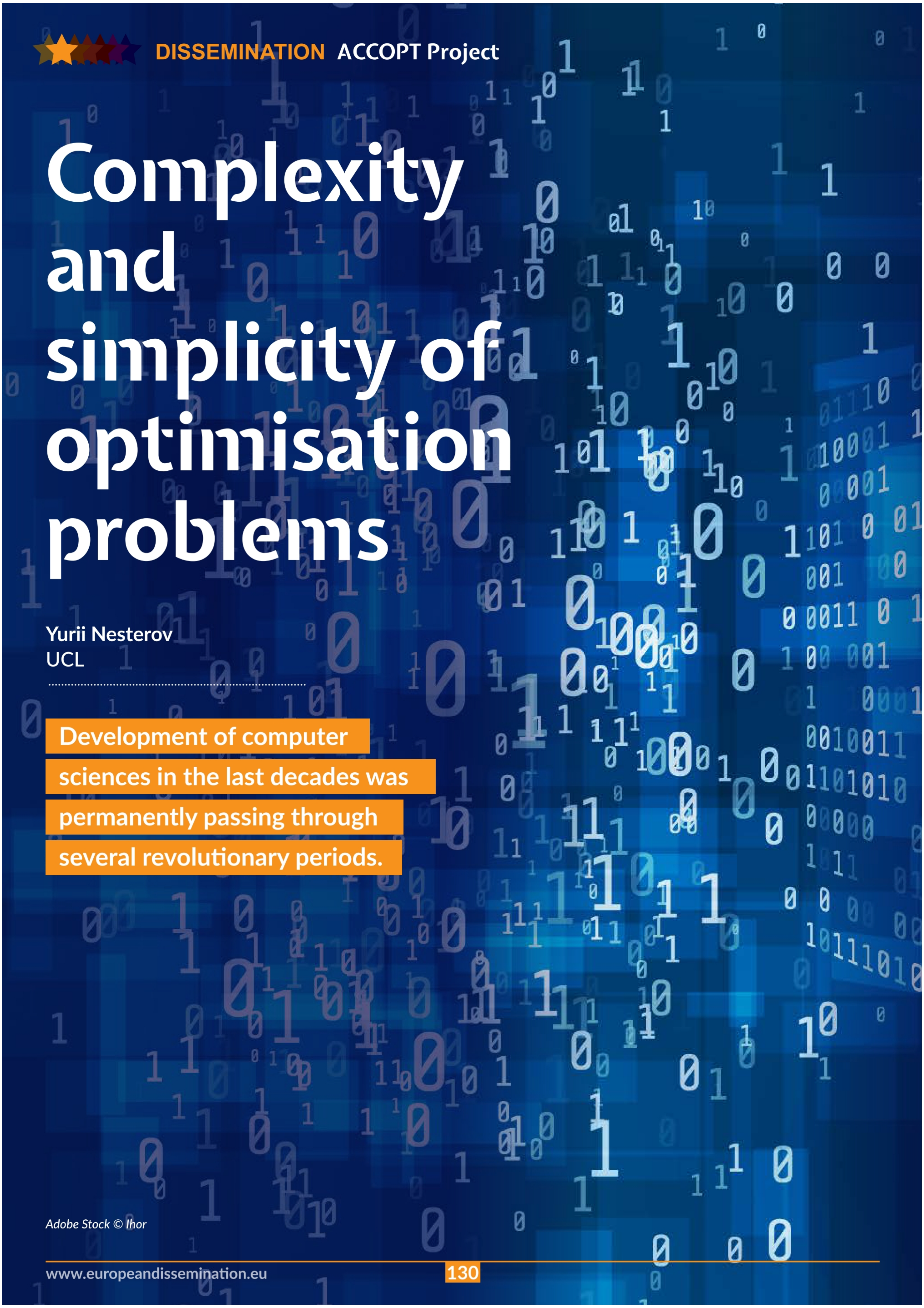
Email:lidam-library@uclouvain.be

https://uclouvain.be/en/research-institutes/lidam/core/core-reprints.html

Complexity and simplicity of optimisation problems

Yurii Nesterov
UCL

Development of computer sciences in the last decades was permanently passing through several revolutionary periods.



Adobe Stock © Ihor

The revolution of personal computers (1980–2000) overlapped with the revolution of Internet and telecoms (1990–2010). Starting from the year 2000, we are participating in the algorithmic revolution, with consequences significantly amplified by the latest revolution of big data (from 2010 till now). As a result, for many leading IT companies today, the main commercial know-how is algorithmic (e.g. Google, Netflix). Thus, we live in a very interesting time, allowing and requiring non-trivial solutions to different well-formalised practical challenges.

Our research group is interested in the mathematical foundations of optimisation algorithms. Algorithms, the backbones of artificial intelligence, need to be developed, properly justified, and efficiently implemented for various practical needs. Our strong orientation on practical purposes introduces very serious time constraints—unusual for mathematicians. Pure mathematics is the only science where the notion of time does not exist at all. The pure mathematician creates immortal theories for immortal (e.g. dead or virtual) things. In our activity, the interesting questions must be answered right now, or in a short period of time. We cannot wait for four hundred years, as was the case with Fermat's great theorem. In this sense, the style of research in computational mathematics is similar to the style of engineering, where the length of events is measured in a natural unit—the length of human life. Projecting this onto our activity, we get a natural scale for the running time of our algorithms.

However, the mathematical aspect of our work—development of theoretical frameworks justified by theorems—remains predominant. It seems that this should be true for computational mathematics in general. Being a very young science, this field has to fix out its own way of accumulating knowledges, the ability which makes the difference between the real sciences and heuristic cookbooks. The algorithms in computational mathematics can be justified only in two possible ways, either by the results of computational experiments or by mathematical theorems. Today, looking at the numerical results published in the respectable scientific journals 40 or 50 years ago, it is easy to come to the only possible conclusion. Indeed, in the modern world, the developments in computing power are so fast that all our numerical achievements become obsolete in a couple of decades. Even if we imagine a very improbable situation of stagnation in this direction, the overall progress of the society will force us to work with new objects, requiring the development of new mathematical models and new optimisation methods. It seems that in this rapidly changing world, only mathematical theories have a chance to survive. For example, for justifying our conclusions on the complexity of some optimisation problems today, we can still use the theorems proved 50 years ago.

The above observations defined the goals, priorities, and tools of our research project Accelerated Convex Optimization (ACCOPT), which is funded by the European Research Council for five years (2018–2023). Let us explain its name in the backward order.

In optimisation problems, we need to find the best possible variant of actions subject (or not) to some constraints (budget, resources, balance, etc.). By convention, the quality of our decision is evaluated by the objective function, which we are going to minimise. Clearly, we can see optimisation formulations everywhere in modern life and even in nature. In engineering, very often these are the optimal design problems. In computer sciences, different optimisation problems arise in image processing, machine learning, neural networks, etc. In economics, we can speak about the optimal allocation of resources, rationality of consumers, and many other things. Even in nature, many processes are explained by minimal principals (e.g. Fermat's principle of propagation of light). According to Leonard Euler, "Nothing takes place in the world whose meaning is not that of some maximum or minimum."

In many situations, we are interested in finding solutions to different optimisation problems. However, like any other very general framework, sometimes optimisation may be misleading. Indeed, in 1970s Nemirovski and Yudin proved that in general, the optimisation problems are unsolvable. Unsolvability here is understood in the 'engineering' sense: for any method, there exists a problem, which needs a computational time bigger than the time of existence of the universe. This sad conclusion explains the word 'convex' in the name of our project. Convex optimisation problems are always solvable within a reasonable computational time. Intuitively, for convex problems, the feasibility of two variants of the solution implies feasibility of any *compromise* solution. This is an opposition to the discrete problems, where we have a yes-or-no situation. Such a choice is clearly difficult both for computers and for human beings (remember the "To be, or not to be" story).

Even for some convex problems, the required computational time can be quite big. It is natural to assume that the complexity of problems grows with their size (number of variables, size of coefficients, etc.). Hence, in the whole class of convex problems, we still have problems with arbitrarily high complexity. How to cut them from our considerations? Here we can again use an allusion to engineering. Indeed, we do not need to solve *all possible* problems in the world. Remember that the main strength or weakness of mathematics is the existence of infinity. We do not need to go up to such a high level of abstraction. For us, it would be great to be capable of solving all problems which we are *able to pose*. This restriction looks quite reasonable, and it relates the required computational time with the time spent on the data collection. As a naive but important example, we can think that the data preparation consists of writing down the coefficients of our model by hand on a paper. Then the preparation time is proportional to the total number of digits in all coefficients. And it would be good if this number could bound somehow the total number of computational resources required by our methods. This interpretation is indeed used in optimisation for justifying performance of one of the most powerful classes of *polynomial-time* methods.

Having a possibility to estimate the worst-case complexity of different algorithms on some problem classes, we can range them and recommend the faster methods for practical applications. In our analysis, an important role belongs to the lower complexity bounds for specific problem classes. If the method has complexity bound proportional to the lower bound, it is called optimal. The important assumption in this framework is the locality of information provided by a unit called *oracle* to the running optimisation schemes at the consequent test points (black box concept).

In the last years, we achieved significant progress in the abilities of gradient-type methods for convex optimisation. It is related to the possibility of looking inside the black box and changing it, making the answer of the oracle more informative. This approach is called the smoothing technique, and it leads to constructing new methods, for which the required number of iterations is proportional to a square root of the number of iterations of the old methods. Since for the latter schemes, the usual number of iterations is of the order of many thousands, the gain in computational time is remarkable.

Another promising direction of research is related to the special methods for huge-scale optimisation problems. One of the main consequences of the big data revolution is the accessibility of the very big data sets. Now it is very easy to collect information for a problem of very big size, where the number of variables is counted by many millions or billions. The usual optimisation methods cannot solve these problems since even the simplest vector operations, like the addition of two vectors, become very expensive. Our main hope in this situation is sparsity. Indeed, huge-scale optimisation problems are usually very sparse. Often, they include matrices of a big size, for which the number of nonzero elements in a row is bounded by a small absolute constant. Special optimisation methods developed for these problems are based on a smart updating strategy of matrix-vector product.

The above research directions are represented in the project by four work packages.

1. **Fast gradient methods in relative scale.** In the standard framework, optimisation methods compute an approximate solution to an optimisation problem with some absolute accuracy. However, in many engineering applications, it is natural to apply relative accuracy. This small change in the meaning requires substantial modification of the optimisation schemes and development of the new principles of acceleration.

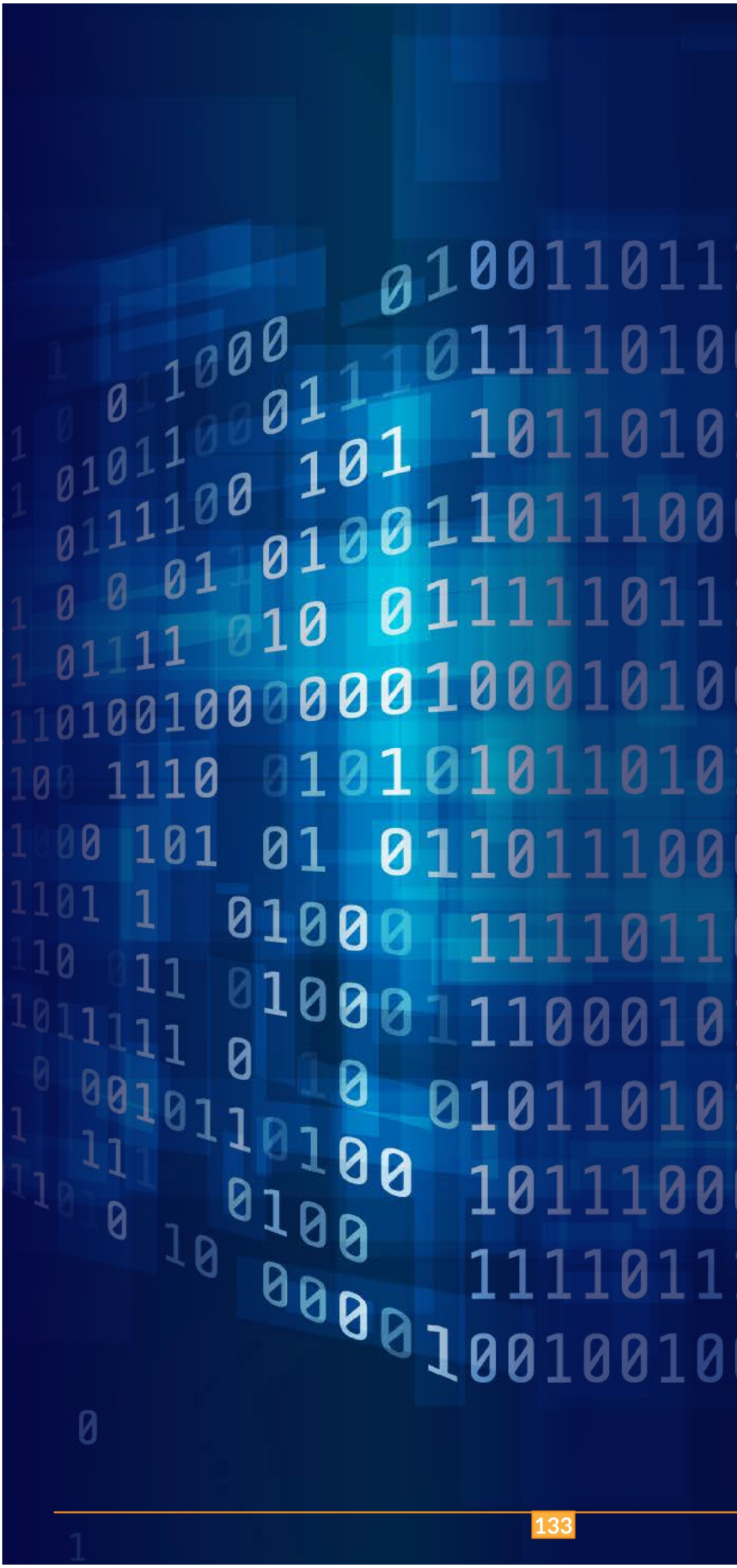
2. **Development of second-order methods for large-scale optimisation problems.** In recent years, we found a possibility for global complexity analysis of the second-order methods. These methods are faster than traditional gradient-type schemes. However, the iterations of these schemes are much more expensive. The goal of this work package is to search for applications where the objective function has sparse Hessians. Using the structure of these objects, we can develop new accelerated schemes for the corresponding large-scale problems.

3. **Huge-scale problems.** For problems of this type, many questions remain open. We need to investigate new possibilities for the development of faster methods. The special structure of these problems can also be taken into account in different ways.

4. **Primal-dual methods for large-scale optimisation problems.** One of the most interesting features of optimisation methods is the newly discovered ability to generate optimal solutions for the dual problem. In convex optimisation, every optimisation problem can be rewritten in an equivalent alternative way. This dual problem is not uniquely defined. But very often it provides us with very interesting information about the model. For example, for models of traffic congestion, the primal variables are the sizes of queues at the bottlenecks, and the dual variables are the equilibrium flows of cars on the roads. For electricity networks, the primal variables are currents in the wires, and dual variables are voltages at the nodes. For economical applications, the primal variables are resources, and the dual variables are the shadow prices, etc. Thus, it is interesting to approximate both primal and dual optimal solutions. This can be done by the modern optimisation methods of different order, generating the accuracy certificates for the initial problem. The main direction of our research is the development of the efficient updating technique, which can work for large-scale problems.

The expected progress in all research directions of our project is related to an intelligent use of the internal structure of optimisation problems. For convex problems, this structure is always visible (to verify convexity, we need to look at the internal structure of the oracle). And it was absolutely ignored by the classical black box theories.

Thus, we understand that methods need help. And by providing it in an appropriate way, we can get much more efficient algorithms than we could imagine.



PROJECT SUMMARY

The ERC Advanced Grant project "Accelerated Convex Optimization (ACCOPT)" is devoted to development of efficient numerical methods for solving modern optimisation problems. New modelling abilities provided by the era of big data, resulted in problems of a very big size, which need creation of new principles for constructing optimisation schemes. All theoretical advances will be confirmed by extensive numerical testing.

PROJECT LEAD PROFILE

Born in 1956, **Professor Nesterov** received his doctorate from Control Problems Institute (Moscow) in 1984. From 1993, he works at UCL (Belgium). His awards include Dantzig Prize (Mathematical Programming Society, 2000), John von-Neumann Theory Prize (INFORM, 2009), Outstanding Paper Award (SIAM, 2014), Euro Gold Medal (European Operations Research Societies, 2016). He has written more than 100 papers and six books.

PROJECT PARTNER

The project is based at the Center of Operations Research and Econometrics and Department of Mathematical Engineering of Catholic University of Louvain (UCL), Belgium. Collaboration partners include groups from Belgium, France, UK, and the USA.

CONTACT DETAILS

Yurii Nesterov

34 voie du Roman Pays, 1348,
Louvain-la-Neuve, Belgium

Yurii.Nesterov@uclouvain.be

FUNDING

This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme under grant agreement no. 788368.