2026 Gauss Prize: Yurii Nesterov

by Allyn Jackson

In 1984, Yurii Nesterov finished his doctoral thesis while working as a researcher at the Soviet Academy of Sciences in Moscow. The thesis presented a new approach to designing certain algorithms for optimization problems and proved mathematically that this approach was the best possible. Though an ingenious and decisive result, it went largely unnoticed. One reason was the Cold War, which restricted communication between mathematicians in the Soviet Union and their counterparts abroad. A more important reason was that the kinds of problems people were trying to solve back in the 1980s were not amenable to Nesterov's method.

But great ideas have staying power. Since in the mid-2000s, Nesterov's thesis work has racked up thousands of citations, as researchers found that an adaptation of his approach—variously called the Fast Gradient Method, Nesterov Accelerated Gradient, and Nesterov Momentum—could greatly improve the performance and efficiency of neural networks. The work of Nesterov thus stands at the heart of many artificial intelligence tools for applications such as image and speech recognition, natural language processing, and recommendation systems.

By the time his name became a household word in neural network research, Nesterov had for decades been a leader in Computational Optimization. In addition to a consummate technical mastery, Nesterov possesses a visionary sense of where potent ideas are to be found. While much of his work is inspired by the needs of science and engineering, his main contributions transcend the particular by unifying disparate classes of problems, providing new frameworks for solutions, and demonstrating through mathematical proof the power and limitations of those frameworks.

At the heart of Nesterov's work is the universality of optimization principles. Seen in everything from a bird's wing shape that optimizes lift in flight, to the formation of water droplets to minimize surface energy, optimization can be thought of as a law of nature. Since the advent of computers in the mid-20th century, the term optimization has also denoted a research area that considers problems for which one seeks an optimal solution subject to constraints. An example might be a farmer aiming to maximize profit by choosing the best mix of crops to plant, given constraints on the land available, the cost of labor, equipment, fertilizer, etc.

Research in optimization is driven by the uncompromising fact that, for many optimization problems that arise in practice, there is no single, general method that works efficiently for all problem instances. Far from hindering research, though, this fact has engendered great creativity and ingenuity as researchers have invented algorithms to find solutions and proved theorems nailing down exactly how fast those algorithms converge to those solutions. Nesterov is one of the greats in this area.

His PhD thesis provides a good example. The starting point is the method of gradient descent, which has been known since the 19th century. The method is akin to the search for an efficient route down a mountain: At each step, choose the direction in which the downhill slope is steepest. Many optimization problems can be represented as a search for a low point in a "landscape" of solutions, and gradient descent guarantees a path to a point lower than the starting point. It will not necessarily reach *the* lowest point; it could wind up in a valley that sits higher than the lowest point in the overall landscape. From experience, we know that this path can be steep and winding.

Nesterov's thesis put a new twist on gradient descent. Instead of calculating the next step based only on information about the previous one, he created a clever algorithm that combined information about all previous steps to generate a kind of average of the gradients (this approach can work only with smooth paths). Other researchers had tried such approaches before, but it was Nesterov who really made it work. Moreover, he proved that no algorithm using only gradient information could run significantly faster than his algorithm.

At the time of his thesis, Russian mathematicians emphasized worst-case analysis of algorithms much more than mathematicians in the West. This difference in emphasis contributed to Nesterov's thesis initially being seen as something of a curiosity. But over the years, worst-case analysis became a potent tool in algorithm development and opened a path for the establishment of the field of algorithmic complexity, which focuses on efficiency of algorithms.

Complexity analysis is also at the heart of another of Nesterov's main achievements, in the area of interior-point methods.

A venerable problem in optimization is linear programming, in which an optimal solution is sought subject to constraints that can be represented as linear spaces—that is, lines or planes or their higher-dimensional counterparts. The linear spaces demarcate a region in which solutions to the problem lie. In the 1940s, the American mathematician George Dantzig invented an algorithm called the simplex method, which turned out to be highly efficient for many linear programming problems.

However, the efficiency was limited: It is easy to construct worst-case problems on which the simplex method becomes wildly inefficient. Russian efforts to find a better algorithm for linear programming culminated in Leonid Khachiyan's 1979 discovery that the so-called ellipsoid method, proposed independently by Nemirovsky and Yudin and by Shor (all from the USSR), bested simplex on the worst cases. Contrary to expectations, though, simplex still performed better than ellipsoid on most problems people needed to solve. In the mid-1980s a researcher in the US, Narendra Karmarkar, developed what is known as interior-point (IP) methods for linear programming. IP methods were even better than the ellipsoid method on the worst-case problems, but again, for many common problems the simplex method continued to perform better in practice.

Clarifying the importance of interior-point research of this period was the 1994 book by Nesterov and his collaborator Arkadi Nemirovski, *Interior-Point Polynomial Algorithms in Convex Programming.* This book had a profound influence by providing a general theoretical setting for understanding and analyzing algorithms not only for linear programming but for the more general class of nonlinear problems known as convex programming. For this latter class, the constraints need not be linear; they can be curves, curved surfaces, or other nonlinear spaces.

Key to Nesterov and Nemirovski's work is their invention of the notion of self-concordant barrier functions. These functions, which represent a generalization of other tools that were in use at the time, do what their name suggests: They provide a special barrier to prevent a search for a solution from wandering outside the region demarcated by the constraints. They also proved that, if one has an optimization problem confined to a certain set describable by a self-concordant barrier function, then there exists an efficient interior-point algorithm for solving the problem.

The book of Nesterov and Nemirovski paved the way for the development of algorithms for solving a variety of problems including what is known as semi-definite programming, which has been widely used in areas such as statistics and control theory. By the time the book appeared in 1994, Nesterov had moved to the Université Catholique de Louvain in Belgium, where today he is a professor emeritus. His influence further expanded with the 2004 publication of *Introductory Lectures on Convex Optimization: A Basic Course*, which provided grounding for generations of students of optimization and became an indispensable reference.

In addition to producing definitive texts laying out broad theories, Nesterov has produced more-focused papers that are small gems. One example is his 2012 paper on coordinate descent methods for large-scale optimization problems. These methods had been in use for decades with little theoretical understanding of why they worked. Nesterov's paper provided a comprehensive theoretical development that opened the way for new uses of these methods.

In recent years, Nesterov has turned his attention to tensor methods, in which one uses not only gradients but higher-order derivatives, which can capture richer details about changes in the geometry of a solution landscape. Requiring huge computing capacity, tensor methods remain theoretical constructs for now but hold the promise of providing future tools of great power and precision.

Nesterov has also thought deeply about the future impact of artificial intelligence on science and society. In the huge electricity requirements of today's AI tools he sees the telltale sign of enormously inefficient training algorithms—and new problems ripe for solution by improved optimization algorithms. As the AI revolution unfolds, the world will continue to need visionaries like Yurii Nesterov.

Biographical information on Yurii Nesterov:
https://en.wikipedia.org/wiki/Yurii_Nesterov